Annals of Operations Research
https://doi.org/10.1007/s10479-024-05969-8

ORIGINAL RESEARCH

Robust assessment of life insurance products

Carole Bernard[1,2] · Corrado De Vecchi[3] · Steven Vanduffel[2]

Received: 28 December 2022 / Accepted: 21 March 2024

Abstract
In this paper we propose a robust assessment for the premium of a standard life insurance contract with respect to the uncertainty on the estimated residual lifetime distribution function. Specifically, we provide a method to derive the range of values that the premium of a given contract can attain when considering all residual lifetime distribution functions that satisfy an L^2 distance constraint to a reference distribution function. Furthermore, we show that the L^2 distance constraint can be used as flexible starting point to include further information regarding future mortality.

Keywords Model risk · Insurance pricing · L^p distance · Longevity risk · Actuarial value

1 Introduction

The actuarial literature related to model risk assessment has experienced rapid growth in recent years. The idea behind this field of research is that any actuarial evaluation is prone to error in that the underlying loss distribution is typically only partially known. There is an extensive literature on finding bounds on a risk measure of a portfolio loss distribution. We refer to Embrechts et al. (2013), Wang et al. (2013), Bignozzi et al. (2015), Puccetti et al. (2016), Bernard et al. (2017), and Liu and Wang (2017) for studies under the assumption that the distribution functions of the risky portfolio components are known but their dependence is

This work was supported by the Research Foundation Flanders (FWO) with the Odysseus project FWOODYS11 and G015320N. Steven Vanduffel is grateful for financial support with funding reference numbers FWO SBO S006721N and FWO WOG W001021N.

✉ Corrado De Vecchi
corrado.de.vecchi@tum.de

Carole Bernard
carole.bernard@grenoble-em.com

Steven Vanduffel
steven.vanduffel@vub.ac.be

1 Department of Accounting, Law and Finance, Grenoble Ecole de Management, Grenoble, France

2 Faculty of Social Sciences and Solvay Business School, Vrije Universiteit Brussel (VUB), Brussels, Belgium

3 Department of Mathematics, Technical University of Munich (TUM), Munich, Germany

Published online: 23 April 2024

either unknown or only partially known, as well as to Barrieu and Scandolo (2015), Cornilly et al. (2018), Li et al. (2018), and Bernard et al. (2023) for various studies of risk bounds under the assumption that moment information of the portfolio loss distribution is available and possibly also some information on its shape. Clearly, a main quantity of actuarial interest that can be affected by model misspecification is the premium to be paid for an insurance contract. The (net) premium of a life-insurance contract depends essentially on two ingredients: the residual lifetime distribution function (actuarial risk) and the discount curve (financial risk). The goal of the present paper is to develop a framework that can help the insurer to deal with model risk that arises from a misspecified residual lifetime distribution function. Recent studies have shown that in a low-interest rate environment longevity risk becomes the major risk-driver of the life-insurance business; see Antolin (2007), Haberman et al. (2011), Rabitti and Borgonovo (2020), and some of the references therein.

Several competing longevity models (or mortality forecast models) have been proposed in the literature. For an overview, see for instance Pitacco et al. (2009). Nonetheless, as is the case for any statistical procedure's output, a projected life table can never be completely trusted. In the life insurance literature, the study of the effect of model misspecification on the price of a contract is usually conducted using a parametric approach: it is assumed that the residual lifetime distribution is within a given family of distributions and one assesses (e.g. via Monte Carlo simulations) the impact of parameter uncertainty on the future (best estimate) lifetime distribution; we refer to Cairns (2000), Olivieri (2001), Olivieri and Pitacco (2011), and Li (2014) who offer discussions of this approach and provide some case studies. While this approach makes it possible to address parameter uncertainty within a given model (family of distributions) it does not account for model uncertainty in that the chosen model may not reflect the real dynamics, that is, the true model is different from the assumed one. One way to partially address this issue is to select a finite set of models and to set a prior probability on each model (Cairns, 2000). However, doing so still exposes the user to model risk, i.e., the risk that the chosen models for the lifetime distributions or their likelihoods are not appropriate. A growing awareness that this risk is relevant in the context of life insurance pricing is well documented in the literature. Recently, regulators are also paying attention to this issue. For instance, regarding the well known Lee-Carter model, EIOPA (2016) states the following: "*It should be noted that the model does neither take into account uncertainty with respect to parameters nor with regard to the model. The future deviations from the best estimate may be larger or smaller because mortality trends may occur which cannot be predicted at present. These include for instance the effects on future mortality of changes in behavioural factors, socio-economic developments and developments in ethics. Which unknown viruses and bacteria may still have an effect on mortality? How will the resistance of antibiotics develop and what medical developments can be expected? All these factors may result in a situation where the future distribution of mortality around the best estimate may differ from the distribution on the basis of historic data calculated in accordance with the model.*"

In order to cope with this issue, we follow a non-parametric approach that can be summarized as follows: a best estimate of the residual lifetime distribution serves as a reference distribution and we provide the price range, i.e. the maximum and the minimum price of any given life insurance contract, when considering all the distribution functions that are somehow close to the reference distribution function (but that are not necessarily coming from the same parametric family). In order to formalize the notion of closeness between probability distributions, we will make use of the L^2 metric between two distribution functions (in a similar spirit as in Pichler (2014)).

There is growing interest in the impact of model risk on insurance pricing. Historically, early results in this field of research can be dated back at least to the seminal article Cantelli (1910). In this article, the author points out that the probability bounds studied for example by P. Tchebichef and I.-J. Bienaymé could be of interest for the development of risk theory with life insurance applications, and not solely from a purely mathematical point of view. As for more recent results, Escobar and Pflug (2020) study the sensitivity and continuity of the distortion premium with respect to the Wasserstein distance. In the context of reinsurance pricing of large catastrophic events, Dietz and Walker (2019) and Dietz and Niehörster (2021) offer analyses that link theory on model risk developed in actuarial science with recent results on decision making under ambiguity derived in the economic literature. In particular, these authors develop a theory that formally motivates premium loadings due to ambiguity. Finally, in the context of life insurance pricing, Pichler (2014) derives upper bounds on a coherent risk measure of a life insurance contract when the lifetime distribution function is known to sit in a Wasserstein ball built around a reference distribution function. Since this latter paper studies a problem closely related to the one we consider here, in Sect. 5 we compare our results to those obtained in this paper.

Section 2 is devoted to the mathematical formulation of the problem at hand and to the modelling of ambiguity using the L^2 metric. We start with a review of some basic life insurance concepts, such as the equivalence principle, and then move to the definition of the premium bounds, whose computation and properties are the main goal of the present analysis.

Section 3 contains our results regarding the properties and computation of the premium bounds obtained when the model risk is described using the L^2 distance. We show that best- and worst-case values for the premium can be explicitly computed using a convex Quadratically Constrained Linear Program (QCLP). By studying the properties of this linear program we prove that premium bounds enjoy desirable properties, such as continuity with respect to the L^2 distance constraint. In some cases, we are able to derive analytic expressions of the feasible probability distribution that respectively maximizes and minimizes the premium.

In Sect. 4, we show that the framework we developed is flexible enough to handle additional constraints other than the L^2 distance. Specifically, we show how one can deal with the important case in which the feasible distribution functions are assumed to be unimodal, and with the case in which interval constraints for the probabilities of death are available. We illustrate that, under these additional constraints, significant improvements of the premium bounds can be obtained.

Section 5 offers a detailed comparison of our results with those obtained in Pichler (2014). In particular, we highlight the main advantages of describing uncertainty via the L^2 metric as compared to the Wasserstein distance used in Pichler (2014).

2 Problem formulation

Let T_x be the residual lifetime of an individual at age x and let K_x be the *curtate remaining lifetime*, defined as the integer part of T_x, $K_x = \lfloor T_x \rfloor$.

It is common to assume that the individual residual lifetime cannot exceed a certain value ω. Hence, $\mathcal{H} = \{0, 1, 2, ..., \omega - x\}$ isx random variable K_x. We denote by $\boldsymbol{q} = \left({}_{0|1}q_x, {}_{1|1}\, q_x, ..., {}_{\omega-x|1}\, q_x\right)^t$ the column vector describing the probability distribution of K_x, namely ${}_{h|1}q_x = P(K_x = h)$, for $h \in \mathcal{H}$. The quantities ${}_{h|1}q_x$ are usually estimated and reported in a life table. Finally, let $g : \mathcal{H} \to \mathbb{R}$ be the *payoff function* of an insurance contract, i.e., the function g associates to each outcome $h \in \mathcal{H}$ that K_x can assume the present value of all future payments the insurer must pay to the beneficiaries.

2.1 Equivalence principle in life insurance

Several pricing principles have been proposed in the actuarial literature. Notable examples are the equivalence principle, the mean-variance pricing principle, the distortion premium principle, and the utility equivalence principle. The interested reader can find an overview in Young (2014). The present analysis focuses mainly on the equivalence principle, which can be defined as follows.

Definition 2.1 *(Equivalence principle)* Given the distribution of K_x and a contract discounted payoff function $g(\cdot)$, the equivalence principle sets the premium π of the insurance contract equal to the present value of future payments, e.g.,

$$\pi = \mathbb{E}(g(K_x)).$$

For further details on the equivalence principle and its application in insurance pricing, we refer to Chapter 4 of Olivieri and Pitacco (2011) and Chapter 6 of Dickson et al. (2009). In this framework, the premium of an insurance contract is set equal to the expected present value of the benefits, also called actuarial value in the life insurance literature. The equivalence pricing principle is conceptually simple and well established in the industry practice. Note that π depends solely on the discounted payoff function and on the probability distribution of K_x, that is π is law-invariant.[1] We report here some concrete examples of discounted payoff functions $g(\cdot)$[2].

- Pure endowment: the amount S will be paid to the beneficiaries after m years if the insured is alive at that time. Since ${}_m p_x = P(K_x \geq m) = \sum_{h=m}^{\omega-x} {}_{h|1}q_x$, the premium

$$\pi = S(1+i)^{-m}\ {}_m p_x = S(1+i)^{-m} \sum_{h=m}^{\omega-x} {}_{h|1}q_x$$

can be written as $\pi = \mathbb{E}\left(g\left(K_x\right)\right)$ with

$$g(h) = \begin{cases} 0 & \text{if } h < m \\ S(1+i)^{-m} & \text{if } h \geq m. \end{cases}$$

- Endowment insurance: the amount C_h will be paid at time h to the beneficiaries if the insured dies between time $h-1$ and h, for $h \leq m$, where m denotes the policy term. The amount S will be paid after m years to the beneficiaries if the insured is alive at that time. The premium of this contract

$$\pi = \sum_{h=1}^{m} C_h (1+i)^{-h}\ {}_{h-1|1}q_x + S(1+i)^{-m} \sum_{h=m}^{\omega-x} {}_{h|1}q_x,$$

[1] A premium π is law-invariant if $X =_d Y$ implies $\pi(X) = \pi(Y)$, in which $=_d$ describes equality in distribution.

[2] For notational convenience, the examples of discounted payoff functions reported in the sequel are written considering the case of a constant annual discounting interest rate i, corresponding to a flat yield curve. This leads to a discount function in the form $v(h) = (1+i)^{-h}$, for all $h \in \mathcal{H}$. Nonetheless, all our results remain valid for any possible shape of the yield curve adopted to compute the present value of future cash flows.

can be written as $\pi = \mathbb{E}\left(g\left(K_x\right)\right)$ with

$$g(h) = \begin{cases} C_{h+1}(1+i)^{-(h+1)} & \text{if } h < m \\ S(1+i)^{-m} & \text{if } h \geq m. \end{cases}$$

- Life annuities: the amount b_h is paid to the beneficiaries at each time $h = 1, 2, ..., \omega - x$ as long as the insured is alive. Note that there exist many payment structures for life annuities, such as constant ($b_h = b$ for $h = 1, 2, ..., \omega - x$), arithmetically increasing ($b_h = b_1(1 + (h-1)\alpha)$), or geometrically increasing ($b_h = b_1(1+\alpha)^{h-1}$). In any case, we have

$$\pi = \sum_{h=1}^{\omega-x} \sum_{j=1}^{h} b_j (1+i)^{-j} \ {}_{h|1}q_x.$$

This can be written as $\pi = \mathbb{E}\left(g\left(K_x\right)\right)$ with

$$g(h) = \begin{cases} 0 & \text{if } h = 0 \\ \sum_{j=1}^{h} b_j (1+i)^{-j} & \text{if } h = 1, 2, ..., \omega - x. \end{cases}$$

Observe that, in the setting we consider, the premium of a life insurance contract depends solely on the distribution of K_x. However, longevity trends have proven to be quite unpredictable, i.e., K_x is subject to distributional uncertainty, and several competing methodologies have been proposed to estimate its probability distribution.

2.2 Premium bounds

Any law invariant pricing principle is affected by possible errors made in the evaluation of the probability distribution of interest. In order to cope with the possible uncertainty regarding the residual lifetime distribution, we consider the case in which such distribution is not completely specified. Following Escobar and Pflug (2020), we denote with $\mathfrak{F}$ a general ambiguity set, i.e., a collection of probability distributions that are compatible with the available information.

Given a discounted payoff function g and an ambiguity set $\mathfrak{F}$, we will study the upper- and lower-bound for the premium computed according to the equivalence principle. These bounds will be denoted as $\overline{\pi}_g^{\mathfrak{F}}$ and $\underline{\pi}_g^{\mathfrak{F}}$ and are defined as

$$\overline{\pi}_g^{\mathfrak{F}} = \sup\{\mathbb{E}(g(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}. \tag{1}$$

$$\underline{\pi}_g^{\mathfrak{F}} = \inf\{\mathbb{E}(g(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}. \tag{2}$$

The interval $[\underline{\pi}_g^{\mathfrak{F}}, \overline{\pi}_g^{\mathfrak{F}}]$ will be used to measure the impact of model misspecification (ambiguity) on the price of the insurance contract identified by the payoff function g. As insurance companies tend to be conservative in their evaluations, it seems natural that in the presence of the distributional uncertainty described by $\mathfrak{F}$, the upper-bound $\overline{\pi}_g^{\mathfrak{F}}$ could be instrumental in setting the commercial premium. In this regard, the upper bound, $\overline{\pi}_g^{\mathfrak{F}}$, satisfies the following basic properties that are considered desirable for a premium principle in the life insurance context, and this holds true regardless of the specific form of the ambiguity set $\mathfrak{F}$.

Proposition 2.2 *For any ambiguity set $\mathfrak{F}$, given two payoff functions g_1 and g_2, the following properties hold:*

1. *Monotonicity: if $g_1(\cdot) \leq g_2(\cdot)$, then $\overline{\pi}_{g_1}^{\mathfrak{F}} \leq \overline{\pi}_{g_2}^{\mathfrak{F}}$.*

2. *Translational invariance: if* $g_1(\cdot) = c + g_2(\cdot)$ *with* $c \in \mathbb{R}$, *then* $\overline{\pi}_{g_1}^{\mathfrak{F}} = c + \overline{\pi}_{g_2}^{\mathfrak{F}}$.
3. *Positive homogeneity: if* $g_1(\cdot) = \lambda g_2(\cdot)$ *with* $\lambda \geq 0$, *then* $\overline{\pi}_{g_1}^{\mathfrak{F}} = \lambda \overline{\pi}_{g_2}^{\mathfrak{F}}$.
4. *Convexity: if* $g(\cdot) = \alpha g_1(\cdot) + (1-\alpha) g_2(\cdot)$ *with* $\alpha \in [0, 1]$, *then* $\overline{\pi}_{g}^{\mathfrak{F}} \leq \alpha \overline{\pi}_{g_1}^{\mathfrak{F}} + (1-\alpha) \overline{\pi}_{g_2}^{\mathfrak{F}}$.

The proof is straightforward and is thus relegated to Appendix A.1. Proposition 2.2 partially extends the results of Theorem 16 in Pichler (2014). On the one hand, Proposition 2.2 is more general than Theorem 16 in Pichler (2014) in that our result is valid for any arbitrary ambiguity set, whereas Pichler (2014) considers the special case of an ambiguity set based on a Wasserstein ball. On the other hand, we consider only the expectation of $g(K_x)$, whereas Theorem 16 of Pichler (2014) considers a general concave distortion risk measure and thus is indeed more general from this point of view.

Hereafter, given $\tilde{F} \in \mathfrak{F}$ and a payoff function $g(\cdot)$, we denote with $\pi_{\tilde{F}}$ the premium computed with the distribution function $\tilde{F}$, i.e., $\pi_{\tilde{F}} = \mathbb{E}(g(K_x))$ with $K_x \sim \tilde{F}$.

In the remainder of this analysis, we will consider the case in which the insurer is able to specify a distribution function $F \in \mathfrak{F}$ that is considered as benchmark or best-estimate distribution. To account for expenses, desired profit margin and model uncertainty the insurance company will add loadings to $\overline{\pi}_{g}^{\mathfrak{F}}$, and in this regard the quantity $\overline{\pi}_{g}^{\mathfrak{F}} - \pi_F$ is to be interpreted as a safety loading.

A main advantage of determining the safety loading using the above mentioned approach with respect to, e.g., modifying the best-estimate distribution using the Wang transform (Wang, 2002), is that it allows to take into account explicitly the model risk faced by the insurer.

2.3 Modelling distributional uncertainty

In order to compute the premium bounds (1) and (2), we need to provide a precise definition of the ambiguity set $\mathfrak{F}$. In this section, we propose to describe ambiguity (uncertainty) using the L^2 distance between distribution functions. For a general introduction to the L^p metric and its financial applications see Rachev et al. (2008). For applications in actuarial science, see for instance López-Díaz et al. (2012) and Yang et al. (2014). In particular, we assume that the distribution of K_x belongs to a subset of an L^2-ball built around a target distribution denoted with F. F represents the current best estimate for the df of K_x, although it is subject to model misspecification. We denote as $\boldsymbol{f} = (f_0, f_1, ..., f_{\omega-x})^t$ its reference probability distribution, i.e. $f_h = P(K_x = h)$ for $h = 0, 1, 2, ..., \omega - x$ under the model $K_x \sim F$. The L^2-ball of radius $\sqrt{\varepsilon}$ and center F is defined as

$$\mathcal{M}_\varepsilon(F) = \left\{ \tilde{F} \,\middle|\, d(\tilde{F}, F) \leq \sqrt{\varepsilon} \right\}, \tag{3}$$

in which $d(\tilde{F}, F)$ is the L^2 distance between the dfs $\tilde{F}$ and F, i.e.,

$$d(\tilde{F}, F) = \sqrt{\int_{-\infty}^{+\infty} \left(\tilde{F}(t) - F(t)\right)^2 dt}. \tag{4}$$

Sometimes we will refer to the L^2-ball as L^2 ambiguity set. Since K_x is a discrete random variable taking values $\{0, 1, 2, ..., \omega - x\}$, any candidate df of K_x is a step function that can have jumps only at the values $\{0, 1, 2, ..., \omega - x\}$, satisfies $\tilde{F}(t) = 0$ for $t < 0$ and $\tilde{F}(t) = 1$, for $t \geq \omega - x$. Thus, the distributional uncertainty regarding K_x can be described by considering all distributions $\tilde{F}$ belonging to $\mathcal{M}_\varepsilon(F)$ that are piecewise constant and with jumps at the points $\{0, 1, 2, ..., \omega - x\}$. Although this set is a subset of $\mathcal{M}_\varepsilon(F)$, in what follows we use the same notation for it as a slight abuse of notation.

If $\tilde{F}$ and F are both piecewise constant, then the function $t \mapsto \left(\tilde{F}(t) - F(t)\right)^2$ is also a step function with jumps at the points $\{0, 1, 2, ..., \omega - x\}$. Therefore, when we restrict ourselves to these distributions, the L^2 distance expression $d(\tilde{F}, F)$ can be expressed as

$$d(\tilde{F}, F) = \sqrt{\sum_{h=0}^{\omega-x} (\tilde{F}_h - F_h)^2}, \tag{5}$$

where $\tilde{F}_h$ and F_h are the constant values taken by $\tilde{F}$ and F on the interval $[h, h+1)$, for $h = 0, 1, ..., \omega - x$. A distribution function $\tilde{F} \in \mathcal{M}_\varepsilon(F)$ is uniquely determined by its probability distribution $\boldsymbol{q} = \left({}_{0|1}q_x, {}_{1|1}\,q_x, ..., {}_{\omega-x|1}\,q_x\right)^t$. Therefore, we sometimes write $d(\boldsymbol{q}, F)$ to denote the L^2 distance between F and the distribution function $\tilde{F}$ such that $\tilde{F}(h) = \sum_{j=0}^{h} {}_{j|1}q_x$. Going forward, we focus on the case in which $\mathfrak{F} = \mathcal{M}_\varepsilon(F)$, as described above. As this ambiguity set depends on the parameter ε, we denote the worst- and best-case prices as $\overline{\pi_\varepsilon}$ and $\underline{\pi_\varepsilon}$, respectively. Thus, the first problem we aim to study in this analysis can be formulated as follows:

$$\begin{array}{rll} \overline{\pi_\varepsilon} = & \max & \mathbb{E}\left(g\left(K_x\right)\right) \\ & \text{subject to} & K_x \sim \tilde{F}, \\ & & \tilde{F} \in \mathcal{M}_\varepsilon(F). \end{array} \tag{6}$$

$$\begin{array}{rll} \underline{\pi_\varepsilon} = & \min & \mathbb{E}\left(g\left(K_x\right)\right) \\ & \text{subject to} & K_x \sim \tilde{F}, \\ & & \tilde{F} \in \mathcal{M}_\varepsilon(F). \end{array} \tag{7}$$

and we also study the distributions attaining the bounds in (6) and (7).

3 Computing premium bounds

In this section we focus on the computation and properties of $\overline{\pi}_\varepsilon$ and $\underline{\pi}_\varepsilon$ in (6) and (7). Moreover, we show how to compute the distributions that achieve the bounds. As a by-product of the analysis conducted here, we highlight the convenience of using the L^2 distance to describe distributional uncertainty. As for notation, in what follows, given a vector $\boldsymbol{a} = (a_1, a_2, ..., a_n) \in \mathbb{R}^n$ with $a_j > 0$ for $j = 1, 2, ..., n$, we write $\boldsymbol{a} > 0$.

3.1 Problem reformulation

Note that the L^2 ambiguity set can be equivalently expressed in terms of vectors $({}_{0|1}q_x, ..., {}_{\omega-x|1}\,q_x)$, i.e., in terms of probability distributions such that $\sum_{j=0}^{h} {}_{j|1}q_x = \tilde{F}_h$. Thus, Problems (6) and (7) can be rewritten as

$$\begin{aligned}
\min_{\boldsymbol{q}} \quad & \langle \boldsymbol{y}, \boldsymbol{q} \rangle = \sum_{h=0}^{\omega-x} y_h \ {}_{h|1}q_x \\
\text{subject to} \quad & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 \leq \varepsilon, \\
& \sum_{h=0}^{\omega-x} {}_{h|1}q_x = 1, \\
& {}_{h|1}q_x \geq 0, h = 0, ..., \omega - x,
\end{aligned} \tag{8}$$

where the vector $\boldsymbol{y} = (y_0, y_1, ..., y_{\omega-x})$ can be identified according to the discounted payoff function g considered. Specifically, Problem (6) corresponds to the case in which $y_h = -g(h)$, and Problem (7) corresponds to the case $y_h = g(h)$. In the language of operational research, Problem (8) can be classified as a Quadratically Constrained Linear Program, a class of problems for which a general closed form solution is not presently available. The following results state the well-posedness and the convexity of our problems.

Proposition 3.1 *Given $\varepsilon > 0$, Problem (8) is well-posed and its feasible region is a non-empty, compact, and convex set.*

The proof is given in Appendix A.2.

Proposition 3.2 *Problems (6) and (7) are well-posed. Moreover, for any $c \in \left[\underline{\pi}_{\varepsilon}, \overline{\pi}_{\varepsilon} \right]$ there exists $\tilde{F} \in \mathcal{M}_{\varepsilon}(F)$ such that $\pi_{\tilde{F}} = c$.*

The proof follows immediately from Proposition 3.1. Since in Problem (8) we are looking for the minimum of a convex function over a convex set, Problem (8) can be classified as a convex problem.

Remark 3.1 Numerical solutions to Problem (8) [and therefore to Problems (6) and (7)] and its optimizing distribution can be easily obtained. We remind the reader that for a convex problem, any local minimum is a global minimum and there exist many efficient algorithms to obtain numerical solutions for this class of optimization problems. See Boyd and Vandenberghe (2004).

As for any procedure based on an optimization problem, it is important that small perturbations of one constraint (in our case, for example, the radius for the L^2-ball) do not significantly affect the solutions. See for example Escobar and Pflug (2020) for a detailed study of continuity of distortion risk measures with respect to the Wasserstein distance. In the following proposition, we ensure the continuity of premium bounds with respect to the L^2 distance constraint.

Proposition 3.3 (Continuity of premium bounds) *Given $\varepsilon > 0$, let $\overline{\pi_{\varepsilon}}$ and $\underline{\pi_{\varepsilon}}$ be defined as in (6) and (7). Then, the mappings*

$$\varepsilon \mapsto \overline{\pi_{\varepsilon}} \quad \textit{and} \quad \varepsilon \mapsto \underline{\pi_{\varepsilon}} \tag{9}$$

are concave and convex, respectively, and thus continuous.

The proof of Proposition 3.3 is given in Appendix A.3.

Let us make the following two observations on the solutions to Problem (8). First, when ε is too big, the L^2 distance constraint in Problem (8) may be redundant and in this case the optimizing distribution function is a degenerate distribution, i.e., a distribution concentrating all probability mass in one point, which is clearly not acceptable for applications in the life insurance context. Hence, in Proposition 3.4 we derive a sufficient condition on ε that prevents this situation. Second, we focus on certain cases in which it is possible to derive an explicit analytic expression for the optimizing distribution of Problem (8), and this is done in Theorem 3.6.

Proposition 3.4 *Let $\boldsymbol{f} > 0$. Then, an optimizing distribution of Problem (8) is a degenerate probability distribution if and only if*

$$\varepsilon \geq \min_{h \in \mathcal{H}^{\boldsymbol{y}}_{min}} \sum_{j=0}^{h-1} F_j^2 + \sum_{j=h}^{\omega-x} (F_j - 1)^2, \tag{10}$$

with $\mathcal{H}^{\boldsymbol{y}}_{min} = \{h \in \mathcal{H} \mid y_h = y_{min}\}$, where $y_{min} = \min\{y_0, y_1, ..., y_{\omega-x}\}$.

The proof is given in Appendix A.4. Proposition 3.4 provides a sufficient and necessary condition on ε such that the probability distributions attaining the bounds in (6) and (7) are not degenerate. We underline that asking that no degenerate distributions belong to the feasible set $\mathcal{M}_\varepsilon(F)$ is a strictly stronger condition than the one in the statement of Proposition 3.4. In what follows, we present a case in which the solution is not degenerate and we can derive an explicit formula. This can be useful for example to check that the numerical solutions to Problem (8) are accurate in these cases. In order to derive this analytical expression, we need some conditions regarding the reference distribution F (or, equivalently, on the reference probability distribution $\boldsymbol{f}$) and the feasible distributions functions of Problem (8). However, we do not need any assumption on $\boldsymbol{y}$ or, equivalently, on the discounted payoff function g. First, we need a lemma.

Lemma 3.5 *Let $\boldsymbol{f} > 0$. Then, there exists ε such that any feasible probability distribution of Problem (8) satisfies $\boldsymbol{q} > 0$, i.e., ${}_{h|1}q_x > 0$ for all $h \in \mathcal{H}$. In particular, $\boldsymbol{q} > 0$ holds for any $\varepsilon \in \left(0,\ \min\left\{\frac{f_h^2}{2} \,\middle|\, h = 0, 1, ..., \omega - x\right\}\right)$.*

The proof is given in Appendix A.4. Lemma 3.5 gives us a sufficient condition on ε such that the assumptions in the next theorem are satisfied. In particular, if $\boldsymbol{f} > 0$ (i.e. $f_h > 0$ for $h = 0, 1, ..., \omega - x$), one can simply choose $\varepsilon < \min\left\{\frac{f_h^2}{2} \,\middle|\, h = 0, 1, ..., \omega - x\right\}$ and the assumptions of Theorem 3.6 are satisfied.

Theorem 3.6 *Let $\boldsymbol{f} > 0$ and $\varepsilon > 0$ such that any feasible probability distribution satisfies $\boldsymbol{q} > 0$. Then, the optimizing distribution of Problem (8) is unique and given by the vector $\boldsymbol{q}^*$ obtained as*

$$\begin{aligned}
{}_{0|1}q_x^* &= f_0 + \frac{y_1 - y_0}{2\lambda^*}, \\
{}_{h|1}q_x^* &= f_h + \frac{y_{h-1} - 2y_h + y_{h+1}}{2\lambda^*} \quad \textit{for } h = 1, 2, ..., \omega - x - 1, \\
{}_{\omega-x|1}q_x^* &= f_{\omega-x} + \frac{y_{\omega-x-1} - y_{\omega-x}}{2\lambda^*},
\end{aligned} \tag{11}$$

where $\lambda^* = \sqrt{\frac{\sum_{h=0}^{\omega-x-1} (y_{h+1} - y_h)^2}{4\varepsilon}}$. Furthermore, $\boldsymbol{q}^*$ satisfies $d(\boldsymbol{q}^*, F) = \sqrt{\varepsilon}$.

The proof is given in Appendix A.5. Theorem 3.6 provides an explicit formula for the minimizing distribution of Problem (8). Furthermore, under the assumptions of Theorem 3.6, i.e.,

Table 1 Parameters used in the numerical examples

i	x	ω
0.025	65	120

ε is small enough, the distributions attaining the bounds in Problems (6) and (7) are unique, admit a closed-form representation and have the maximal L^2 distance admissible from the reference probability distribution.

To give an example, consider an endowment insurance in which the amount C_h will be paid at time h to the beneficiaries if the insured dies between time $h-1$ and h, where $h \leq m$ and m denotes the policy term. The amount S will be paid at time m to the beneficiaries if the insured is alive at that time. The discounted payoff function of this contract writes as

$$g(h) = \begin{cases} C_{h+1}(1+i)^{-(h+1)}, & \text{if } h < m \\ S(1+i)^{-m}, & \text{if } h \geq m. \end{cases}$$

Under the assumption of Theorem 3.6, we can find the probability distribution solving Problem (6) by using formula (11) with $y_h = -g(h)$, for $h \in \mathcal{H}$. To conclude this section, we add a remark on the choice of ε.

Remark 3.2 The value of ε reflects the level of ambiguity regarding the curtate residual lifetime distribution. One practical method that can be used to choose a specific value of ε is the following. Consider the case in which the reference distribution F is estimated but there is also a finite set of alternative candidate distributions that are deemed reasonable. In this situation, one can compute the L^2 distance between the reference distribution and each other candidate and set $\sqrt{\varepsilon}$ equal to the maximal L^2 distance observed. This procedure ensures that all other distributions deemed reasonable belong to the ambiguity set $\mathcal{M}_\varepsilon(F)$.

3.2 Numerical illustration

In this section we illustrate how the results obtained in Sect. 3 can be used to obtain a robust assessment of annuities. In doing so, we also highlight some of the limitations of the framework developed so far. In the following examples, the reference probability distribution F of K_x is taken as the empirical distributions arising from the Italian population table (dd. 25/06/2022), as available on the website of the Italian National Institute of Statistics (ISTAT). Unless otherwise specified, the parameters adopted in the numerical examples are taken from Table 1. The numerical optimizations were conducted using the standard interior-point algorithm implemented using the function *fmincon* available in the software Matlab.[3]

The product considered here is a standard (paid in advance) whole life annuity, paying 1 euro per year as long as the insured is alive. Assuming a constant interest rate i, the corresponding discounted payoff function $g(h)$ is given as

$$g(h) = \begin{cases} 0, & \text{if } h = 0 \\ a_{\overline{h}|}, & \text{if } h = 1, 2, \ldots, \omega - x, \end{cases} \tag{12}$$

where $a_{\overline{h}|} = \sum_{j=1}^{h} (1+i)^{-j}$, and the premium writes as $\pi = \mathbb{E}\left(g(K_x)\right) = \sum_{h=1}^{\omega-x} a_{\overline{h}|} \; {}_{h|1}q_x$.

[3] In order to guarantee accuracy, the input values for “MaxFunctionEvaluations” and “MaxIterations” were both set equal to 10000. All the other hyperparameters coincide with the Matlab default options for the interior-point algorithm.

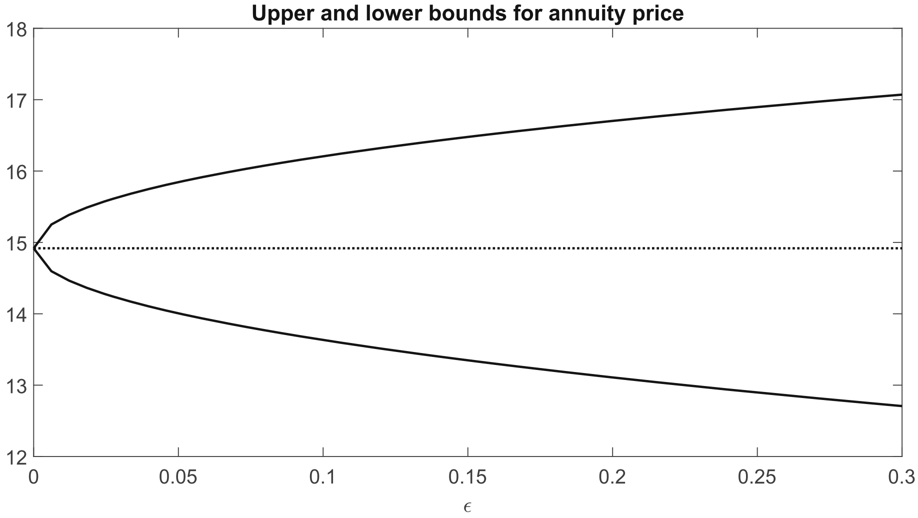

Fig. 1 Premium bounds with respect to ε, for ε varying between $\varepsilon = 0$ and $\varepsilon = 0.3$. The dashed line describes the premium level computed using the reference distribution. The black lines represent the upper and lower bounds corresponding to each value of ε. The annuity discounted payoff function is given in (12). The other parameters are taken from Table 1

The premium upper and lower bounds are displayed in Fig. 1. Here we find a numerical confirmation of what was anticipated in Proposition 3.1, which states that the premium upper bound is concave with respect to ε while the lower bound is convex.

Figure 2 shows for various levels of ε the probability distributions for K_x attaining the premium upper bound (Panel A) and lower bound (Panel B). Clearly, the worst-case distributions displayed in Panel A of Fig. 2 are obtained by moving as much probability mass as possible from early to late ages, in particular to the maximal attainable age. The reason behind this shape is that we are maximizing the expected value of an increasing discounted payoff function taking its highest value in $\omega - x$ (corresponding to the case in which the insured reaches the maximal attainable age ω). By contrast, in Panel B of Fig. 2 one can observe that the best-case distributions are attained by moving as much probability mass as possible to early ages, and specifically to the earliest age considered, the reason being that the distributions displayed in Panel B are now the minimizers of the expected value of an increasing discounted payoff function.

The distributions reported in Fig. 2 appear to be somewhat odd. This is due to the fact that we are adopting a non-parametric approach: we describe uncertainty using a probability metric and hence we do not force the solution to belong to a certain parametric model. The use of a non-parametric approach is justified when the actuary is willing to consider the possibility that future mortality could exhibit some features that standard parametric models are not able to take into account. The distributions displayed in Fig. 2 correspond to the worst- and best-case scenarios in this setup. Nevertheless, while the graphs in Fig. 2 reflect extreme cases, some interesting interpretations are still possible. By fixing ω, the actuary is essentially fixing the maximal age (biological bound) she believes an insured can reach. Panel A is then suggesting that this biological bound will not be overcome, but, as worst-case scenario, this maximal age could be reached by a higher share of the portfolio population, for example due to medical progress.

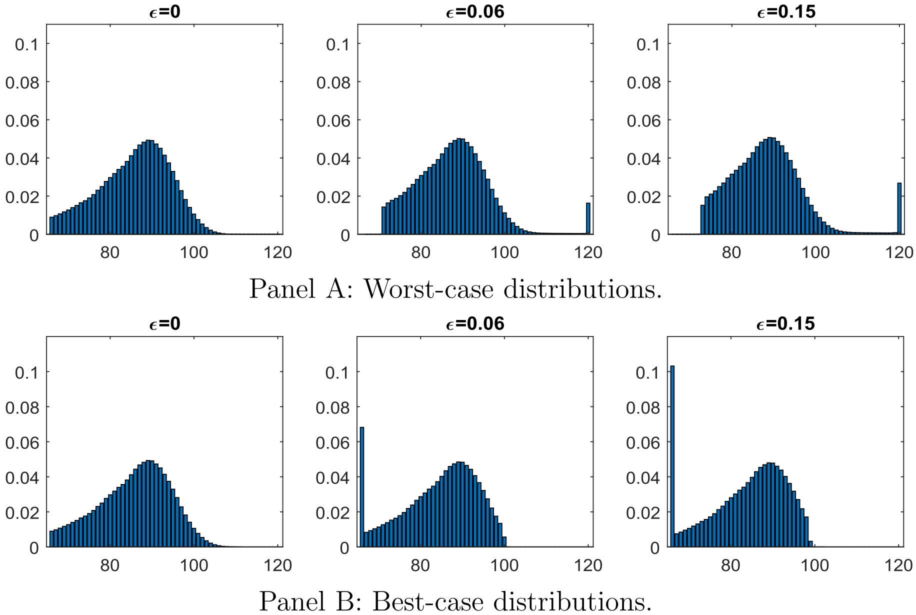

Fig. 2 Worst- and best-case probability distributions for K_x under L^2 distance constraints. The annuity discounted payoff function is given in (12). The other parameters are taken from Table 1

Nonetheless, in Sect. 4 we illustrate how additional constraints on the set of feasible distributions can be useful in those situations in which the actuary does not deem realistic to have a high probability mass concentrated at very early or late ages.

4 Improving the bounds

The worst- and best-case distributions that we display in Fig. 2 exhibit a somewhat particular shape that may be seen as unrealistic. We show how our setting is flexible enough to incorporate additional constraints that appear reasonable, leading to tighter bounds and extreme distributions that are smoother, without impairing the bounds' numerical tractability. This is done by considering additional constraints on the ambiguity set. Specifically, we discuss two additional constraints: a mode preserving constraint and an interval forecast constraint.

4.1 Unimodality and interval forecast constraints

In a life insurance context, the mode of the residual life time distribution is sometimes referred as the Lexis point (Pitacco et al., 2009). A unimodality constraint is not new in the literature related to risk bounds. Li et al. (2018) and Bernard et al. (2020) derive bounds on VaR under a unimodality constraint. However, to the best our knowledge, our study is the first to consider an additional unimodality constraint when ambiguity is described using a probability metric. First, let us introduce the definition of unimodality that we consider in this analysis.

Definition 4.1 (Keilson and Gerber, 1971) Given a discrete random variable X taking value on $\mathcal{H}$, we say that its distribution is unimodal if there exists at least one $h_m \in \mathcal{H}$ (called the mode) such that $P(X = h)$ is non-decreasing for $h = 0, 1, 2..., h_m$ and $P(X = h)$ is non-increasing for $h = h_m, h_m + 1, ..., \omega - x$.

For example, a random variable with a discrete uniform distribution on $\mathcal{H}$ is unimodal.

Second, we also want to introduce constraints in the form

$$|{}_{h|1}q_x - f_h| \leq \delta_h, \text{ for } h = 0, 1, ..., h^* \text{ with } h^* \leq \omega - x. \tag{13}$$

Adding constraints as in (13) allows us to possibly include information on future mortality coming from an interval forecast of the insured's probabilities of death. Alternatively, the radius δ_h can be seen as a quantity that describes the level uncertainty regarding the point forecast described by the reference probability distribution $\boldsymbol{f}$. Note that δ_h depends on h, and we are thus able to account for the (plausible) situation in which the forecasted annual probabilities of death at a later time can be more uncertain than the survival probabilities in the near future.

Consider now the case in which one is interested in finding the bounds for the premium of an insurance contract having a payoff function g, considering all distributions of K_x that satisfy an L^2 distance constraint from a reference distribution F, some interval constraints in the form of (13), and that additionally are unimodal with mode equal to h_m. The probability distributions attaining these bounds can be found by solving the following problem:

$$\begin{aligned}
\min_{\boldsymbol{q}} \quad & \langle \boldsymbol{y}, \boldsymbol{q} \rangle = \sum_{h=0}^{\omega-x} y_h \; {}_{h|1}q_x \\
\text{subject to} \quad & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 \leq \varepsilon, \\
& \sum_{h=0}^{\omega-x} {}_{h|1}q_x = 1, \\
& {}_{h|1}q_x \geq 0, \;\; h = 0, ..., \omega - x, \\
& {}_{h|1}q_x \geq {}_{h+1|1}\, q_x, \;\; h = h_m, ..., \omega - x - 1, \\
& {}_{h|1}q_x \geq {}_{h-1|1}\, q_x, \;\; h = 2, ..., h_m, \\
& |{}_{h|1}q_x - f_h| \leq \delta_h, \;\; h = 0, 1, ..., h^*,
\end{aligned} \tag{14}$$

in which the vector $\boldsymbol{y} = (y_0, y_1, ..., y_{\omega-x})$ can be defined as $y_h = -g(h)$ and $y_h = g(h)$ for $h \in \mathcal{H}$ in order to derive the distributions that attain the upper and lower bound for $\mathbb{E}(g(K_x))$, respectively. Note that $h^* + x$ can be seen as the maximum age for which the insurer is confident in fixing an interval constraint for probabilities of death.

The following propositions ensure that if the reference distribution is unimodal, the additional constraints that we are discussing do not affect the numerical tractability of the problem at hand, and the continuity of the premium bounds with respect to the L^2 distance constraint.

Proposition 4.2 (Unimodality and interval constraints) *Let the reference distribution be unimodal with modal value h_m and $\varepsilon > 0$. Then, Problem (14) is well-posed and its feasible region is a non-empty, compact, and convex set.*

The proof of Proposition 4.2 follows from an argument that is similar to the proof of Proposition 3.1 and thus is omitted.

Proposition 4.3 (Continuity of premium bounds with additional constraints) *Let the reference distribution be unimodal with modal value h_m and $\varepsilon > 0$, and let $\overline{\pi}$ and $\underline{\pi}$ be the respective upper and lower bound for the premium computed using the solutions of Problem (14). Then, the mappings*

$$\varepsilon \mapsto \overline{\pi} \quad \textit{and} \quad \varepsilon \mapsto \underline{\pi} \tag{15}$$

are concave and convex, respectively, and thus continuous.

The proof of Proposition 4.3 is given in Appendix A.7.

Remark 4.1 Further constraints could be added to Problem (14) without impairing its numerical tractability, according to the specific information available to the insurer. One could consider, e.g., one or more constraints on the moments of K_x, in the form

$$\mathbb{E}\left(K_x^k\right) \in [\overline{\mu_k}, \underline{\mu_k}], \text{ in which } k \in \mathbb{N} \text{ and } \overline{\mu_k}, \underline{\mu_k} \in \mathbb{R}. \tag{16}$$

Note that for $k = 1$, the constraint in (16) corresponds to a constraint on the life expectancy of the policyholder. One could easily show that the numerical tractability of Problem (14) is preserved if we add constraints in the form of (16), but we omit it the proof for brevity.

4.2 Numerical illustration

We now illustrate numerically the impact of the additional interval and unimodality constraints on the premium bounds and on their attaining distributions, with respect to the case in which the sole L^2 distance radius is fixed.

To allow comparability, we consider the same product, benchmark distribution function and numerical optimization procedure as in Sect. 3.2. As for the product, a (whole life) standard annuity with discounted payoff function given in (12). As for the reference distribution F, the empirical distribution of K_x arising from the whole Italian population. The other parameters are given in Table 1.

The interval constraints in Problem (14) will be fixed as $\delta_h = f_h \alpha_h$, with $\alpha_h \leq \alpha_{h+1}$ for $h = 0, 1, ..., h^*$, leading to

$$|_{h|1}q_x - f_h| \leq \delta_h \iff f_h(1-\alpha_h) \leq_{h|1} q_x \leq f_h(1+\alpha_h).$$

Then, α_h can be seen as a relative increasing rate of error. Specifically, we consider $\alpha_h = \frac{h}{\omega - x}$, for $h = 1, 2, ..., h^*$, with $h^* + x = 100$. This specification of α_h is consistent with the fact that uncertainty regarding forecasted probabilities is higher for late ages.

Figure 3 points out an interesting aspect regarding the solutions to Problems (8) and (14). If one compares Figs. 2 and 3, it is clear that the interval and unimodality constraints can have an impact on the shape of the distribution attaining the premium bounds. In particular, the constraint on the modal value prevents the worst-case (resp. best-case) probability distribution from having a high probability mass concentrated at maximal (resp. minimal) attainable age, a feature that we observed in Fig. 2. Thus, these additional constraints can be useful in those situations in which probability distributions in the form represented in Fig. 2 are not considered acceptable as worst- or best-case scenarios, and one wants to obtain a stressed but still reasonable shape.

The additional constraints considered in Problem (14) can have an impact also on the premium bounds, as illustrated in Fig. 4.

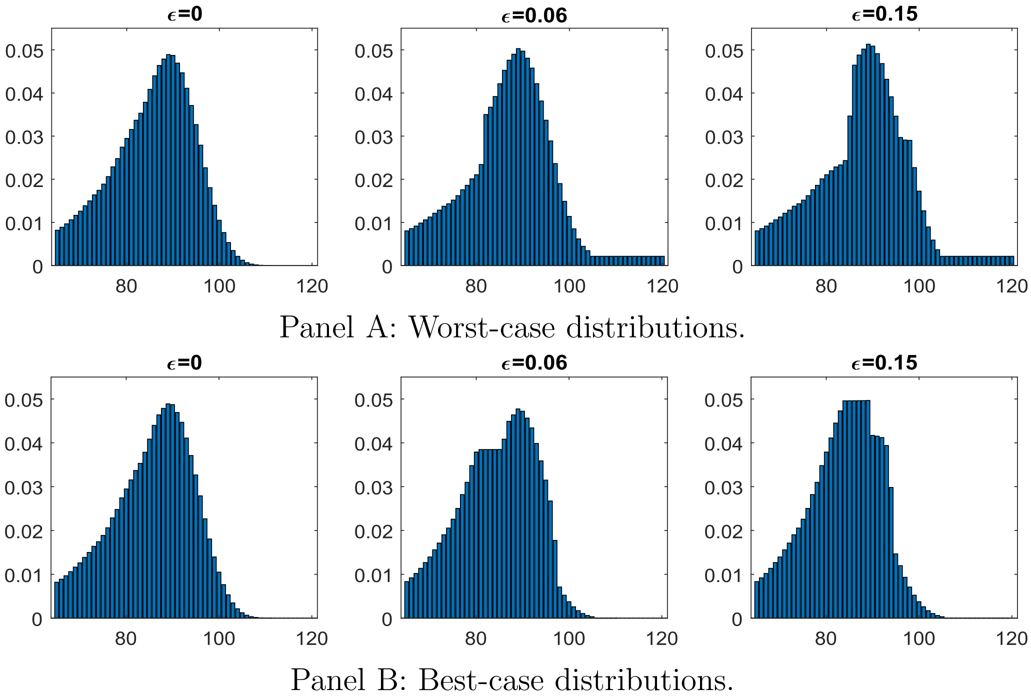

Fig. 3 Worst- and best-case probability distributions for K_x under L^2 distance, interval forecast and unimodality constraints with modal value at age 85. The annuity discounted payoff function is given in (12). The other parameters are taken from Table 1

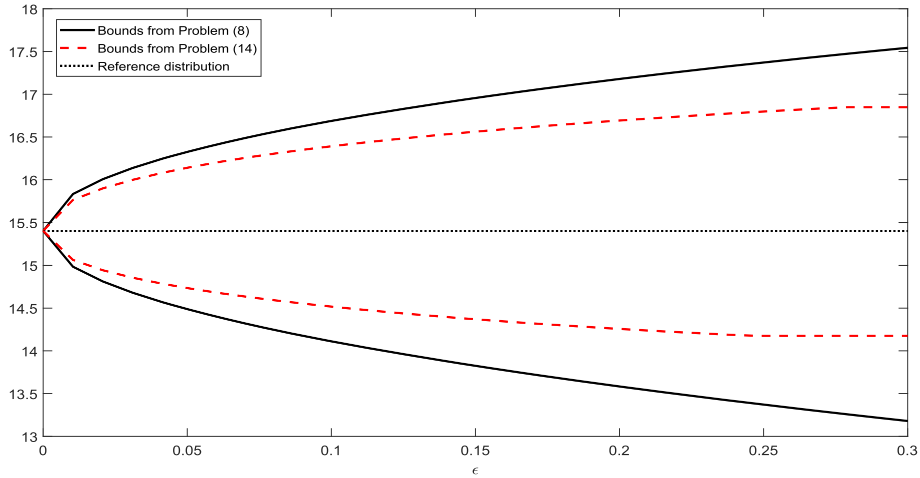

Fig. 4 Premium bounds with respect to ε, for ε varying between $\varepsilon = 0$ and $\varepsilon = 0.3$. The dashed line describes the premium level computed using the reference distribution. The black lines represent the upper and lower bounds corresponding to each value of ε obtained with the sole L^2 distance. The red lines represent the upper and lower bounds corresponding to each value of ε, under L^2 distance, interval forecast and unimodality constraints with modal value at age 85. The annuity discounted payoff function is given in (12). The other parameters are taken from Table 1

4.3 Further applications

The aim of this section is to underline that the results obtained to compute the bounds on the premium can also be used to compute the bounds of other functions of the residual lifetime K_x. Here are a few examples.

- *Stop-loss premium bounds*. So far, we have considered premiums computed using the equivalence principle (Definition 2.1). Another well-established premium principle in the actuarial literature is the stop-loss premium. For an introduction and properties of the stop-loss premium and the related stop-loss order, see for example Kaas (1993). If $d \in \mathbb{R}$ and $g(K_x)$ is the random payoff of the contract, then we denote with π_d the stop-loss premium with threshold d,

$$\pi_d = \mathbb{E}\left((g(K_x) - d)_+\right), \tag{17}$$

  where $(g(K_x) - d)_+ = \max\left(g(K_x) - d, 0\right)$. Clearly, π_d can be seen as the expected value of $t(K_x)$ where $t(\cdot) = \max\left(g(\cdot) - d, 0\right)$ is a deterministic function of K_x. Thus, the bounds of π_d can be computed and studied using the results obtained in Sect. 3.
- *Robust expected discounted utility*. Consider an individual who aims to buy an annuity but has to decide which payment structure is the best. Let $\boldsymbol{b} = (b_0, b_1, \ldots, b_{\omega-x})$ be the vector that identifies the payment structure of an annuity, i.e., for $h = 1, 2, \ldots, \omega - x$, b_h is the amount that is paid at policy anniversary h, while b_0 can be set equal to $-\pi$ where π is the premium that the insurance company charges for this annuity. The standard tool adopted in the economic theory of inter-temporal choices to compare consumption plans (such as annuities) is the expected discounted utility. See for example the seminal paper of Yaari (1965) and the stream of literature that followed. Since here we are comparing insurance contracts that offer payments only at policy anniversary h for $h \in \{0, 1, 2, \ldots, \omega - x\}$,

the value of the discounted utility at time h for an annuity with payment structure $\boldsymbol{b} = (b_0, b_1, ..., b_{\omega-x})$ can be written as

$$D_{\boldsymbol{b}}(h) = \sum_{j=0}^{h} \gamma(j)u(b_j), \tag{18}$$

where $u(\cdot)$ is the utility function describing the individual preferences and $\gamma(\cdot)$ is the subjective discount function. Thus, the expected discounted utility of an annuity with payment structure $\boldsymbol{b}$ becomes

$$\mathbb{E}(D_{\boldsymbol{b}}(K_x)) = \mathbb{E}\left(\sum_{j=0}^{K_x} \gamma(j)u(b_j)\right) = \sum_{h=0}^{\omega-x} {}_{h|1}q_x \sum_{j=0}^{h} \gamma(j)u(b_j). \tag{19}$$

Observe that the function $D_{\boldsymbol{b}}(\cdot)$ is merely a deterministic function of K_x. Hence, if the ambiguity regarding the distribution of K_x is described using L^2-balls, and possible additional constraints, all the results obtained in this paper can be used to derive the bounds of the expected discounted utility in (19), regardless of the characteristics of the subjective discount function $\gamma(\cdot)$ and of the utility function $u(\cdot)$. A well-established result in decision theory under ambiguity is that an ambiguity averse decision maker should choose the option that maximizes her expected utility under the worst-case scenario. This was proven in a decision theoretic setting in the seminal paper of Gilboa and Schmeidler (1989). See for example d'Albis and Thibault (2012) for a discussion of this approach in the context of the optimal annuitization problem. Being able to compute upper and lower bounds for the expected discounted utility can help an ambiguity averse decision maker to choose which insurance contract is optimal for her.
Specifically, given an ambiguity set $\mathfrak{F}$ and a set $\mathcal{S}$ of possible payment structures, an ambiguity averse decision maker with a utility function $u(\cdot)$ and a subjective discount function $\gamma(\cdot)$ faces the following problem:

$$\max_{\boldsymbol{b}\in\mathcal{S}} \min_{K_x\sim\tilde{F},\ \tilde{F}\in\mathfrak{F}} \mathbb{E}(D_{\boldsymbol{b}}(K_x)). \tag{20}$$

If the ambiguity set $\mathfrak{F}$ is described using an L^2-ball built around a reference distribution and if the set $\mathcal{S}$ of considered payment structures is finite, then our results can be used to compute $\min_{K_x\sim\tilde{F},\tilde{F}\in\mathcal{M}} \mathbb{E}(D_{\boldsymbol{b}}(K_x))$ for all $\boldsymbol{b} \in \mathcal{S}$ and the solution of (20) can then be found.

Our approach can be implemented to study worst- and best-case scenarios of any law-invariant functional that can be expressed as the expected value of a function of K_x. Hence, an interesting venue for future research is to investigate if our results can be extended to beyond this setting, for instance to the case of coherent risk measures.

5 Comparison with Pichler (2014)

Given a coherent risk measure ϱ, Pichler (2014) studies the upper bound of the values that $\varrho(g(K_x))$[4] can take when there is ambiguity regarding the distribution of K_x. In the present paper we focus on the upper and lower bounds on the expectation of $g(K_x)$, meaning that

[4] Note Pichler (2014) denotes the discounted payoff function using the symbol L, while we denote the discounted payoff function using g.

we consider the case in which $\varrho(\cdot) = \mathbb{E}(\cdot)$, and thus there is apparently an overlap between our analysis and that of Pichler (2014). However, this is not the case. The aim of this section is to point out some methodological differences and to list the advantages of our set-up.

Pichler (2014) describes the ambiguity around a reference distribution using the Wasserstein distance, whereas we consider the L^2 distance defined in (4). For any given $p \geq 0$, the Wasserstein distance of order p between two distributions on the real line writes as

$$d_{W_p}(\widetilde{F}, F) = \left(\int_0^1 \left| \widetilde{F}^{-1}(\alpha) - F^{-1}(\alpha) \right|^p d\alpha \right)^{\frac{1}{p}}. \tag{21}$$

For $p = 1$, $d_{W_1}(\widetilde{F}, F)$ coincides with the L^1 distance between distribution functions, but correspondence between Wasserstein distances of order p and L^p distances it is not true in general for $p \neq 1$. Thus, even if one fixes the same reference distribution F and the same radius, the premium bounds obtained using the Wasserstein distance or using the L^2 distance do not coincide in general.

A priori, it is not clear which metric one should use to describe distributional uncertainty in the context of life insurance pricing. See for example Chapter 3 in Rachev et al. (2008) for a discussion of the characteristics of various probability metrics with financial applications. The main advantages of our approach are the following.

First, our approach makes the problem much more tractable from the numerical and mathematical points of view. From a mathematical point of view, Pichler (2014) obtains most of his results under the assumption of Hölder continuity of the discounted payoff function [see Definition 19 in Pichler (2014)], and the bounds obtained in Theorems 20 and 22 require the computation of the Hölder constant. Observe that the assumption of Hölder continuity needs to be verified case by case, and this can be difficult if one has a non-constant term structure or a contract with non-constant payments. In the present paper we do not make any assumption on the discounted payoff function $g(\cdot)$, but we are still able to derive premium bounds and analytic expressions of their attaining distributions (Theorem 3.6), and to prove that the premium bounds are continuous with respect to the parameter ε (Proposition 3.3). This feature allows us, for example, to extend the applicability of the results obtained in Sect. 3 to other contexts, as in Sect. 4.3.

From a numerical point of view, we show that with the L^2 metric the bounds on the price can be reformulated as an easy-to-solve convex problem, while Pichler (2014) states that computing the bounds numerically using the Wasserstein distance requires solving an involved bilinear optimization.

Second, our set-up is flexible enough to include further constraints in the ambiguity set, without impairing the bounds' numerical tractability. This features are illustrated in Sect. 4 (Proposition 4.2, Proposition 4.3 and Remark 4.1). We show that this flexibility can be important to obtain a realistic description of model risk in the life insurance context that is able to include, for example, a unimodality preserving constraint and constraints compatible with an interval forecast of the deferred probabilities of death. Note that this possibility of including further constraints on top of the Wasserstein distance was not discussed in Pichler (2014). Finally, we underline that the possible need for additional constraints, on top of the one expressed using a probability metric, arises from the analysis of the worst-case distribution's shape we conduct in Sect. 3.2, and this sort of analysis is missing in Pichler (2014).

6 Conclusions

In this paper we develop new and flexible tools that insurance companies can use to assess the impact of model risk on the premium of standard life-insurance products. Our analysis focuses on those situations in which uncertainty regarding future mortality trends is present and the premium is determined using the equivalence principle. We study the upper and lower bound the premium can attain under the condition that the residual lifetime is required to remain close to a reference distribution where closeness is measured using the L^2 metric. In particular, the use of this metric makes it possible to reformulate the premium bounds' problem as a Quadratically Constrained Linear Program, which is numerically tractable (convex) and easy-to-implement. We further study the properties of this linear program and show that in some cases explicit formulas for the premium bounds and their attaining distribution functions can be obtained. The use of the sole L^2 distance may lead to worst-case distributions that display a somewhat odd shape. In order to overcome this issue, we illustrate how an L^2 distance constraint can be a flexible starting point to include further possibly available information regarding future mortality, such as interval forecast and unimodality constraints. The possibility of including such constraints in our setting allows to obtain a more realistic but still robust assessment for the premium of a life insurance product.

Funding Open Access funding enabled and organized by Projekt DEAL.

A Proofs

A.1 Proof of Proposition 2.2

Proof Given an ambiguity set $\mathfrak{F}$,

1. Let $\tilde{F} \in \mathfrak{F}$, $g_1(\cdot) \leq g_2(\cdot)$ and $K_x \sim \tilde{F}$. Then, $g_2(K_x)$ is first order stochastically larger than $g_1(K_x)$, that is $P(g_1(K_x) \leq z) \geq P(g_2(K_x) \leq z)$ for all $z \in \mathbb{R}$. This implies $\mathbb{E}(g_1(K_x)) \leq \mathbb{E}(g_2(K_x))$ for all $\tilde{F} \in \mathfrak{F}$, which brings us to $\overline{\pi}_{g_1}^{\mathfrak{F}} \leq \overline{\pi}_{g_2}^{\mathfrak{F}}$.
2. If $g_1(\cdot) = c + g_2(\cdot)$ with $c \in \mathbb{R}$, then
$$\begin{aligned}\overline{\pi}_{g_1}^{\mathfrak{F}} &= \sup\{\mathbb{E}(g_1(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} = \sup\{c + \mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}\\ &= c + \sup\{\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} = c + \overline{\pi}_{g_2}^{\mathfrak{F}}.\end{aligned}$$
3. if $g_1(\cdot) = \lambda g_2(\cdot)$ with $\lambda \geq 0$, then
$$\begin{aligned}\overline{\pi}_{g_1}^{\mathfrak{F}} &= \sup\{\mathbb{E}(g_1(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} = \sup\{\lambda\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}\\ &= \lambda\overline{\pi}_{g_2}^{\mathfrak{F}}.\end{aligned}$$
4. if $g(\cdot) = \alpha g_1(\cdot) + (1-\alpha)g_2(\cdot)$ with $\alpha \in [0, 1]$ then,
$$\overline{\pi}_{g}^{\mathfrak{F}} = \sup\{\alpha\mathbb{E}(g_1(K_x)) + (1-\alpha)\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}$$

$$\le \alpha \sup\{\mathbb{E}(g_1(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\} + (1-\alpha) \sup\{\mathbb{E}(g_2(K_x)) : K_x \sim \tilde{F},\ \tilde{F} \in \mathfrak{F}\}$$
$$= \alpha \overline{\pi}_{g_1}^{\mathfrak{F}} + (1-\alpha)\overline{\pi}_{g_2}^{\mathfrak{F}}.$$

□

A.2 Proof of Proposition 3.1

Proof Let us denote with $\mathcal{D}_\varepsilon$ the feasible region of Problem (8), i.e.

$$\mathcal{D}_\varepsilon = \mathcal{D}_1 \cap \mathcal{D}_2 \cap \mathcal{D}_3, \tag{22}$$

where

$$\mathcal{D}_1 = \left\{ \boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \middle| \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 \le \varepsilon \right\}, \tag{23}$$

$$\mathcal{D}_2 = \left\{ \boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \middle| \boldsymbol{e}^t \boldsymbol{q} = 1 \right\} \text{ with } \boldsymbol{e} = (1, 1, ..., 1), \tag{24}$$

$$\mathcal{D}_3 = \left\{ \boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \middle| \boldsymbol{q} \ge 0 \right\}. \tag{25}$$

Observe that $\mathcal{D}_2 \cap \mathcal{D}_3$ is clearly a convex and closed set. The set $\mathcal{D}_1$ is defined as the ε-sublevel set of the function $d^2 : \boldsymbol{q} \to \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2$, which is the sum of $\omega - x + 1$ functions defined by

$$d_h : \mathbb{R}^{h+1} \to \mathbb{R}, \text{ with } d_h({}_{0|1}q_x, ..., {}_{h|1}\, q_x) = \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 .$$

For any given h, the Hessian matrix of d_h is a $h+1$ by $h+1$ matrix whose elements are all equal to 2. It is thus a positive semidefinite matrix. Hence, all functions d_h are convex, which means that d^2 is also a convex function of $\boldsymbol{q}$. The continuity and convexity of d^2 imply that its ε-sublevel set $\mathcal{D}_1$ is closed and convex. To conclude, the feasible set $\mathcal{D}_\varepsilon$ is obtained as the intersection of two closed and convex sets and hence it is closed and convex. Since $\mathcal{D}_\varepsilon$ is a subset of $[0, 1]^{\omega-x+1}$, $\mathcal{D}_\varepsilon$ is also bounded and therefore compact.

Finally, $\mathcal{D}_\varepsilon$ cannot be empty since the vector describing the probability distribution induced by the reference df F always belongs to $\mathcal{D}_\varepsilon$.

Second, we prove that Problem (8) admits at least one solution. Note that in Problem (8) we are looking for the minimum of the function $\boldsymbol{q} \mapsto \langle \boldsymbol{y}, \boldsymbol{q} \rangle$ that is linear with respect to $\boldsymbol{q}$. Therefore, the image of $\mathcal{D}_\varepsilon$ through this function is a closed interval and hence there exists at least one element in $\mathcal{D}_\varepsilon$ that minimizes $\langle \boldsymbol{y}, \boldsymbol{q} \rangle$. □

A.3 Proof of Proposition 3.3

Proof To prove the statement, it is sufficient to show that the mapping $\varepsilon \mapsto \langle \boldsymbol{y}, \boldsymbol{q}_\varepsilon^* \rangle$ is convex, where $\boldsymbol{q}_\varepsilon^*$ is an optimizing distribution for Problem (8). Fix a reference distribution F, and let D_ε be defined as in (22). From Proposition 3.1, D_ε is convex and compact for any $\varepsilon > 0$. Observe that

$$\langle \boldsymbol{y}, \boldsymbol{q}_\varepsilon^* \rangle = \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q} \rangle \,\middle|\, \boldsymbol{q} \in \mathcal{D}_\varepsilon \right\}.$$

Fix now $\lambda \in (0,1)$, $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$. Let us define the following set,

$$\mathcal{D}^{\lambda}_{\varepsilon_1,\varepsilon_2} = \left\{ \boldsymbol{q}_\lambda \;\middle|\; \boldsymbol{q}_\lambda = \lambda \boldsymbol{q}_1 + (1-\lambda)\boldsymbol{q}_2,\ \boldsymbol{q}_1 \in \mathcal{D}_{\varepsilon_1},\ \boldsymbol{q}_2 \in \mathcal{D}_{\varepsilon_2} \right\}.$$

In Appendix A.2 we showed that the function $d^2 : \boldsymbol{q} \rightarrow \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2$ is convex, and thus we have

$$d^2(\boldsymbol{q}_\lambda, F) \leq \lambda d^2(\boldsymbol{q}_1, F) + (1-\lambda) d^2(\boldsymbol{q}_2, F) = \lambda\varepsilon_1 + (1-\lambda)\varepsilon_2.$$

It is then clear that $\mathcal{D}^{\lambda}_{\varepsilon_1,\varepsilon_2} \subseteq \mathcal{D}_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2}$. Finally, we have

$$\begin{aligned}
\langle \boldsymbol{y}, \boldsymbol{q}^*_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2} \rangle &= \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q} \rangle \;\middle|\; \boldsymbol{q} \in \mathcal{D}_{\lambda\varepsilon_1+(1-\lambda)\varepsilon_2} \right\} \leq \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q}_\lambda \rangle \;\middle|\; \boldsymbol{q}_\lambda \in \mathcal{D}^{\lambda}_{\varepsilon_1,\varepsilon_2} \right\} \\
&= \lambda \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q}_1 \rangle \;\middle|\; \boldsymbol{q}_1 \in \mathcal{D}_{\varepsilon_1} \right\} + (1-\lambda) \min \left\{ \langle \boldsymbol{y}, \boldsymbol{q}_2 \rangle \;\middle|\; \boldsymbol{q}_2 \in \mathcal{D}_{\varepsilon_2} \right\} \\
&= \lambda \langle \boldsymbol{y}, \boldsymbol{q}^*_{\varepsilon_1} \rangle + (1-\lambda) \langle \boldsymbol{y}, \boldsymbol{q}^*_{\varepsilon_2} \rangle.
\end{aligned}$$

□

A.4 Proof of Proposition 3.4

Proof Consider the following optimization problem:

$$\begin{aligned}
\min_{\boldsymbol{q}} \quad & \sum_{h=0}^{\omega-x} y_h \; {}_{h|1}q_x \\
\text{subject to} \quad & \sum_{j=0}^{\omega-x} {}_{j|1}q_x = 1, \\
& {}_{h|1}q_x \geq 0, \forall h = 0, ..., \omega - x.
\end{aligned} \tag{26}$$

Note that the solution of Problem (26) coincides with y_{min}, trivially. Furthermore, a degenerate probability distribution attains the equality $y_{min} = \sum_{h=0}^{\omega-x} y_h \; {}_{h|1}q_x$ if and only if ${}_{h^*|1}q_x = 1$ for one $h^* \in \mathcal{H}^{\boldsymbol{y}}_{min}$ and ${}_{h^*|1}q_x = 0$ for $h \neq h^*$. Thus, the number of degenerate probability distributions solving Problem (26) coincides with the number of elements in $\mathcal{H}^{\boldsymbol{y}}_{min}$. Let us now denote with $\boldsymbol{q}_{h^*}$ a degenerated probability distribution such that ${}_{h^*|1}q_x = 1$ and ${}_{h|1}q_x = 0$ for $h \neq h^*$. The square of the L^2 distance between the df of $\boldsymbol{q}_{h^*}$ and the reference distribution F is given by

$$d(\boldsymbol{q}_{h^*}, F)^2 = \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\omega-x} (1 - F_h)^2.$$

It is then clear that if $\varepsilon \geq \min_{h \in \mathcal{H}^{\boldsymbol{y}}_{min}} \sum_{j=0}^{h-1} F_j^2 + \sum_{j=h}^{\omega-x} (F_j - 1)^2$, then there exists $h^* \in \mathcal{H}^{\boldsymbol{y}}_{min}$ such that $\boldsymbol{q}_{h^*}$ belongs to the feasible set of Problem (8), and thus Problem (8) admits degenerate solutions.

We shall now prove that if $\varepsilon < \min_{h \in \mathcal{H}^{\boldsymbol{y}}_{min}} \sum_{j=0}^{h-1} F_j^2 + \sum_{j=h}^{\omega-x} (F_j - 1)^2$, a solution of Problem (8) cannot have all probability mass concentrated in one point. Consider h^* in $\mathcal{H} \backslash \mathcal{H}^{\boldsymbol{y}}_{min}$, and consider the distribution $\boldsymbol{q}_{h^*}$ such that ${}_{h^*|1}q_x = 1$ and ${}_{h|1}q_x = 0$ for $h \neq h^*$. Let $y^* = \langle \boldsymbol{y}, \boldsymbol{q}_{h^*} \rangle$. Assume that $\boldsymbol{q}_{\tilde{h}^*}$ is a distribution attaining the minimum of Problem (8), and thus $d(\boldsymbol{q}_{h^*}, F) \leq \varepsilon$. Fix now $\tilde{h} \in \mathcal{H}^{\boldsymbol{y}}_{min}$.

First, we focus on the case where $\tilde{h} > h^*$. Consider the distribution $\tilde{\boldsymbol{q}}$ such that

$$\begin{aligned} {}_{h|1}\tilde{q}_x &= 1-\delta, \text{ for } h = h^*, \\ {}_{h|1}\tilde{q}_x &= \delta, \text{ for } h = \tilde{h}, \\ {}_{h|1}\tilde{q}_x &= 0, \text{ elsewhere,} \end{aligned}$$

with $0 < \delta < 1$. Since $\tilde{h} \in \mathcal{H}^{\boldsymbol{y}}_{min}$, it is clear that $\langle \boldsymbol{y}, \tilde{\boldsymbol{q}} \rangle < \langle \boldsymbol{y}, \boldsymbol{q}_{h^*} \rangle$. Fix $\delta = \frac{1-F_{\omega-x-1}}{2}$, so that $0 < 1-\delta-F_h < 1-F_h$ holds for any $h \leq \omega - x - 1$. Since $\boldsymbol{f} > 0$ and δ is strictly positive, we have

$$\begin{aligned} d(\tilde{\boldsymbol{q}}, F)^2 &= \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\tilde{h}-1} (1-\delta-F_h)^2 + \sum_{h=\tilde{h}}^{\omega-x} (1-F_h)^2 \\ &< \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\omega-x} (1-F_h)^2 = d(\boldsymbol{q}_{h^*}, F)^2 \leq \varepsilon. \end{aligned}$$

Thus, $\tilde{\boldsymbol{q}}$ belongs to the feasible set of Problem (8) and leads to a strictly lower value of the objective function. This shows that the degenerate distribution $\boldsymbol{q}_{h^*}$ cannot be a solution of Problem (8).

Let us now consider the case $\tilde{h} < h^*$. Again, we define the distribution $\tilde{\boldsymbol{q}}$ such that

$$\begin{aligned} {}_{h|1}\tilde{q}_x &= 1-\delta, \text{ for } h = h^*, \\ {}_{h|1}\tilde{q}_x &= \delta, \text{ for } h = \tilde{h}, \\ {}_{h|1}\tilde{q}_x &= 0, \text{ elsewhere.} \end{aligned}$$

with $0 < \delta < 1$. Since $\tilde{h} \in \mathcal{H}^{\boldsymbol{y}}_{min}$, it is clear that $\langle \boldsymbol{y}, \tilde{\boldsymbol{q}} \rangle < \langle \boldsymbol{y}, \boldsymbol{q}_{h^*} \rangle$. Fix $\delta = \frac{F_{\tilde{h}}}{2}$. Since $\boldsymbol{f} > 0$ and δ is strictly positive, we have

$$0 < F_h - \delta < F_h \implies \left(\sum_{j=0}^{h} {}_{h|1}\tilde{q}_x - F_h \right)^2 = (F_h - \delta)^2 < F_h^2, \text{ for } \tilde{h} \leq h < h^* - 1.$$

Thus,

$$\begin{aligned} d(\tilde{\boldsymbol{q}}, F)^2 &= \sum_{h=0}^{\tilde{h}-1} F_h^2 + \sum_{h=\tilde{h}}^{h^*-1} (F_h-\delta)^2 + \sum_{h=\tilde{h}}^{\omega-x} (1-F_h)^2 \\ &< \sum_{h=0}^{h^*-1} F_h^2 + \sum_{h=h^*}^{\omega-x} (1-F_h)^2 = d(\boldsymbol{q}_{h^*}, F)^2 \leq \varepsilon. \end{aligned}$$

Thus, $\tilde{\boldsymbol{q}}$ belongs to the feasible set of Problem (8) and leads to a strictly lower value of the objective function. This shows that a degenerate distribution $\boldsymbol{q}_{h^*}$ cannot be a solution of Problem (8). □

A.5 Proof of Lemma 3.5

Proof Let $\boldsymbol{f} > 0$ and $\boldsymbol{q}$ be a probability distribution such that there exists $k \in \mathcal{H}$ for which ${}_{h|1}q_x = 0$. Then, $d(\boldsymbol{q}, F)^2 \geq \varepsilon^*$ with

$$\begin{aligned} \varepsilon^* = \min_{\boldsymbol{z}} \quad & \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} z_j - F_h \right)^2 \\ \text{subject to} \quad & \boldsymbol{z} \in \mathbb{R}^{\omega-x+1}, \\ & z_k = 0. \end{aligned} \tag{27}$$

Assume $0 < k < \omega - x$. In order to solve (27), we adopt the standard Lagrangian multiplier method with a Lagrangian function defined as

$$\mathcal{L}(\boldsymbol{z}, \lambda) = \sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} z_j - F_h \right)^2 - \lambda \left(z_k - 0 \right). \tag{28}$$

After differentiating with respect to z_h, we get the following system of equations:

$$\begin{cases} \sum_{h=i}^{\omega-x} \sum_{j=0}^{h} z_j = \sum_{h=i}^{\omega-x} F_h, \text{ for } i \neq k, \\ \sum_{h=k}^{\omega-x} \sum_{j=0}^{h} z_j = \sum_{h=k}^{\omega-x} F_h - \frac{\lambda}{2}. \end{cases} \tag{29}$$

The system in (29) is a system of linear equations. It can thus be solved using a standard substitution method, which leads to the following solution:

$$\begin{cases} z_h = f_h, \text{ for } h \neq k-1, k, k+1, \\ z_h = f_h - \frac{\lambda}{2}, \text{ for } h = k-1, k+1, \\ z_k = f_k + \lambda. \end{cases} \tag{30}$$

To satisfy the constraint in (27), we get $\lambda = -f_k$. At this point we can easily compute the solution of (27), which is $\varepsilon^* = \frac{f_k^2}{2}$. Thus, if the value of ε in Problem (8) is lower than $\frac{f_k^2}{2}$, no feasible distribution can satisfy ${}_{k|1}q_x = 0$. Using a similar argument, one can show that if $k = 0$ or $k = \omega - x$, the solutions of (27) are $\varepsilon^* = \frac{f_{\omega-x}^2}{2}$ and $\varepsilon^* = f_0^2$, respectively. □

A.6 Proof of Theorem 3.6

Proof Recall that from Proposition 3.1 we know that Problem (8) admits at least one optimizing distribution. First, we show that if $\boldsymbol{f} > 0$ and ε is such that any feasible probability distribution satisfies $\boldsymbol{q} > 0$, then an optimizing distribution $\boldsymbol{q}^*$ of Problem (8) must satisfy $d(\boldsymbol{q}^*, F)^2 = \varepsilon$. Let $\boldsymbol{q}$ be a feasible distribution such that $d(\boldsymbol{q}, F)^2 < \varepsilon$. Fix h_{min} in $\mathcal{H}_{min}^{\boldsymbol{y}}$, $\tilde{h}$ in $\mathcal{H} \backslash \mathcal{H}_{min}^{\boldsymbol{y}}$, and let us define $\boldsymbol{q}^*$ as follows:

$$\begin{aligned} {}_{h_{min}|1}q_x^* &= {}_{h_{min}|1}\, q_x + \delta, \\ {}_{\tilde{h}|1}q_x^* &= {}_{\tilde{h}|1}\, q_x - \delta, \\ {}_{h|1}q_x^* &= {}_{h|1}\, q_x, \quad \text{for } h \neq h_{min}, \tilde{h}. \end{aligned}$$

Observe that for any $\delta > 0$, we have $\langle \boldsymbol{y}, \boldsymbol{q}^* \rangle < \langle \boldsymbol{y}, \boldsymbol{q} \rangle$. Clearly, $d(\boldsymbol{q}^*, F)^2$ is a continuous function of δ and for $\delta = 0$ we have $d(\boldsymbol{q}^*, F)^2 = d(\boldsymbol{q}, F)^2 < \varepsilon$. Thanks to the continuity of

$d(\boldsymbol{q}^*, F)^2$ w.r.t. δ, we know that there exists $\delta > 0$ such that $d(\boldsymbol{q}^*, F)^2 \leq \varepsilon$. This implies that $\boldsymbol{q}$ can not be an optimizing distribution for Problem (8). Thus, any optimizing distribution must have the maximal L^2 distance from the reference distribution.

Second, we look for the optimizing distributions of Problem (8) by means of KKT conditions. This is justified by the fact that Problem (8) has differentiable objective and constraint functions. Slater's conditions are satisfied since the reference probability distribution $\boldsymbol{f}$ is strictly feasible in that $\boldsymbol{f}$ has strictly positive probabilities by assumption and $d(\boldsymbol{f}, F)^2 = 0 < \varepsilon$. Thus, strong duality holds. As a consequence, any optimizing distribution of Problem (8) must satisfy the KKT conditions. Moreover, from Proposition 3.1 we know that Problem (8) is convex and therefore we conclude that a feasible point is an optimizing distribution of Problem (8) if and only if it satisfies the KKT conditions. For a detailed explanation of Slaters's condition, strong duality, and KKT conditions we refer to Boyd and Vandenberghe (2004), sections 5.2.3 and 5.5.3, respectively. The KKT conditions of Problem (8) write as follows:

1. $d(\boldsymbol{q}^*, F)^2 - \varepsilon \leq 0$,
2. $\sum_{h=0}^{\omega-x} {}_{h|1}q_x^* = 1$,
3. ${}_{h|1}q_x^* \geq 0$, for $h = 0, 1, ..., \omega - x$,
4. $\lambda^* \geq 0$,
5. $\nu_h^* \geq 0$, for $h = 0, 1, ..., \omega - x$,
6. $\lambda^* \left(d(\boldsymbol{q}^*, F)^2 - \varepsilon\right) = 0$,
7. ${}_{h|1}q_x^* \ \nu_h^* = 0$, for $h = 0, 1, ..., \omega - x$,
8. $\nabla_{\boldsymbol{q}^*}\mathcal{L}(\boldsymbol{q}^*, \lambda^*, \mu^*, \boldsymbol{\nu}^*) = \mathbf{0}$,

where $\mathcal{L}(\boldsymbol{q}, \lambda, \mu, \boldsymbol{\nu})$ is the Lagrangian function corresponding to Problem (8),

$$\mathcal{L}(\boldsymbol{q}, \lambda, \mu, \boldsymbol{\nu}) = \sum_{h=0}^{\omega-x} y_h \ {}_{h|1}q_x + \lambda \left(\sum_{h=0}^{\omega-x} \left(\sum_{j=0}^{h} {}_{j|1}q_x - F_h \right)^2 - \varepsilon \right) + \mu \left(\sum_{h=0}^{\omega-x} {}_{h|1}q_x^* - 1 \right) + \sum_{h=0}^{\omega-x} {}_{h|1}q_x \ \nu_h. \tag{31}$$

and the KKT condition 8 is obtained by equating to 0 the following partial derivatives,

$$\begin{aligned}\frac{\partial \mathcal{L}(\boldsymbol{q}, \lambda, \mu, \boldsymbol{\nu})}{\partial_{k|1}q_x} &= y_k + 2\lambda \left(\sum_{h=i}^{\omega-x} \left(\sum_{j=0}^{h} {}_{k|1}q_x - F_h \right) \right) + \mu + \nu_k \\ &= y_k + 2\lambda \left(\sum_{h=i}^{\omega-x} \sum_{j=0}^{h} {}_{j|1}q_x \right) + 2\lambda \sum_{h=i}^{\omega-x} F_h + \mu + \nu_k.\end{aligned}$$

Hence, for any given λ, μ and $\boldsymbol{\nu}$ we can write the system of equations $\nabla_{\boldsymbol{q}}\mathcal{L}(\boldsymbol{q}, \lambda, \mu, \boldsymbol{\nu}) = \mathbf{0}$ in the following compact form

$$\boldsymbol{y} + 2\lambda \left(M\boldsymbol{q} - \boldsymbol{F}_{cum} \right) + \mu + \boldsymbol{\nu} = \mathbf{0} \tag{32}$$

where $\boldsymbol{F}_{cum}$ is the column vector such that $\boldsymbol{F}_{cum}(k) = \sum_{h=k}^{\omega-x} F_h$, for $k = 0, 1, ..., \omega - x$ and M is a $\omega - x + 1 \times \omega - x + 1$ symmetric matrix whose components are $M(k, j) = \omega - x + 2 - \max(k, j)$, for $k, j = 1, 2, ..., \omega - x + 1$. Using the standard Gauss Algorithm, one can easily check that M has full rank, and hence it is an invertible matrix and we deduce that the system of equations in (32) has a unique solution in the form

$$\boldsymbol{q}^* = M^{-1}\left(\boldsymbol{F}_{cum} - \frac{\boldsymbol{y} + \mu + \boldsymbol{\nu}}{2\lambda}\right). \tag{33}$$

After some calculations, we find

$$\begin{aligned}
{}_{0|1}q_x^* &= f_0 + \frac{y_1 - y_0 + \nu_1 - \nu_0}{2\lambda^*},\\
{}_{h|1}q_x^* &= f_h + \frac{y_{h-1} - 2y_h + y_{h+1} + \nu_{h-1} - 2\nu_h + \nu_{h+1}}{2\lambda^*} \quad \text{for } h = 1, 2, ..., \omega - x - 1,\\
{}_{\omega-x|1}q_x^* &= f_{\omega-x} + \frac{y_{\omega-x-1} - 2y_{\omega-x} + \nu_{\omega-x-1} - 2\nu_{\omega-x} - \mu}{2\lambda^*}.
\end{aligned}$$

At this stage, we just need to find the values of λ, μ, and $\boldsymbol{\nu}$ that satisfy KKT conditions 1-7.

First, $\boldsymbol{\nu}$: since we are considering a case in which any feasible point must satisfy $\boldsymbol{q} > 0$, the constraints in the form ${}_{h|1}q_x \geq 0$ cannot be active at the optimal point, and therefore in our problem KKT conditions 3, 5, and 7 are satisfied if and only if $\boldsymbol{\nu} = \boldsymbol{0}$.

Second, μ: observe that $\sum_{h=0}^{\omega-x} {}_{h|1}q_x^* = \sum_{h=0}^{\omega-x} f_h + \frac{-y_n-\mu}{2\lambda}$. Hence, to satisfy the KKT condition 2 we need

$$\sum_{h=0}^{\omega-x} {}_{h|1}q_x^* = 1 \iff \mu^* = -y_n.$$

Third, λ: we have already shown that any optimizing distribution of Problem (8) must satisfy $d(\boldsymbol{q}^*, F)^2 - \varepsilon = 0$. This constraint is therefore active at the solution, and we just need to find the corresponding $\lambda > 0$. Using the previous results on $\boldsymbol{\nu}^*$ and μ^* we have

$$\begin{aligned}
d(\boldsymbol{q}^*, F)^2 = \varepsilon \iff & \sum_{h=0}^{\omega-x}\left(\sum_{j=0}^{h} {}_{j|1}q_x^* - F_h\right)^2 = \varepsilon\\
\iff & \sum_{h=0}^{\omega-x}\left(\sum_{j=0}^{h}\left(f_j + \frac{c_j}{2\lambda}\right) - F_h\right)^2 = \varepsilon\\
\iff & \sum_{h=0}^{\omega-x}\left(\sum_{j=0}^{h}\frac{c_j}{2\lambda}\right)^2 = \varepsilon
\end{aligned}$$

where $\sum_{j=0}^{h} c_j = y_{h+1} - y_h$ for $h = 0, 1, ..., \omega - x - 1$ and $\sum_{j=0}^{\omega-x} c_j = 0$. We conclude that the only λ that satisfies both condition 1 and condition 4 is $\lambda^* = \sqrt{\frac{\sum_{h=0}^{\omega-x-1}(y_{h+1}-y_h)^2}{4\varepsilon}}$, which completes the proof. □

A.7 Proof of Proposition 4.3

Proof Let us denote with $\mathcal{D}_\varepsilon^*$ the feasible region of Problem (14), i.e.

$$\mathcal{D}_\varepsilon^* = \mathcal{D}_1 \cap \mathcal{D}_2 \cap \mathcal{D}_3 \cap \mathcal{D}_4 \cap \mathcal{D}_5 \tag{34}$$

where $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are defined in (23), (24) and (25), respectively, and

$$\mathcal{D}_4 = \Big\{ \boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \Big| {}_{h|1}q_x \geq {}_{h+1|1}\, q_x,\ h = h_m, ..., \omega - x - 1, \text{ and } \\ {}_{h|1}q_x \geq {}_{h-1|1}\, q_x,\ h = 2, ..., h_m, \Big\}, \tag{35}$$

$$\mathcal{D}_5 = \Big\{ \boldsymbol{q} \in \mathbb{R}^{\omega-x+1} \Big| |{}_{h|1}q_x - f_h| \leq \delta_h,\ h = 0, 1, ..., h^* \Big\}, \tag{36}$$

The rest of proof follows by Proposition 14 and a proof that is similar to the arguments in Appendix A.3. □

References

Antolin, P. (2007). Longevity risk and private pensions. *OECD Working papers on insurance and private pensions*.

Barrieu, P., & Scandolo, G. (2015). Assessing financial model risk. *European Journal of Operational Research, 242*(2), 546–556.

Bernard, C., Rüschendorf, L., Vanduffel, S., & Wang, R. (2017). Risk bounds for factor models. *Finance and Stochastics, 21*(3), 631–659.

Bernard, C., Kazzi, R., & Vanduffel, S. (2020). Range value-at-risk bounds for unimodal distributions under partial information. *Insurance: Mathematics and Economics, 94*, 9–24.

Bernard, C., Pesenti, S., & Vanduffel, S. (2023). Robust distortion risk measures. *Mathematical Finance*.

Bignozzi, V., Puccetti, G., & Rüschendorf, L. (2015). Reducing model risk via positive and negative dependence assumptions. *Insurance: Mathematics and Economics, 61*, 17–26.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Cairns, A. J. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics, 27*(3), 313–330.

Cantelli, F. P. (1910). Intorno a un teorema fondamentale della teoria del rischio. *Bollettino Attuari Italiano, 24*, 1–23.

Cornilly, D., Rüschendorf, L., & Vanduffel, S. (2018). Upper bounds for strictly concave distortion risk measures on moment spaces. *Insurance: Mathematics and Economics, 82*, 141–151.

d'Albis, H., & Thibault, E. (2012). Optimal annuitization, uncertain survival probabilities, and maxmin preferences. *Economics Letters, 115*(2), 296–299.

Dickson, D. C., Hardy, M. R., & Waters, H. R. (2009). *Actuarial mathematics for life contingent risks*. Cambridge University Press.

Dietz, S., & Niehörster, F. (2021). Pricing ambiguity in catastrophe risk insurance. *The Geneva Risk and Insurance Review, 46*, 112–132.

Dietz, S., & Walker, O. (2019). Ambiguity and insurance: Capital requirements and premiums. *Journal of Risk and Insurance, 86*(1), 213–235.

EIOPA. Discussion paper on the review of specific items in the Solvency II delegated regulation. (2016).

Embrechts, P., Puccetti, G., & Rüschendorf, L. (2013). Model uncertainty and var aggregation. *Journal of Banking & Finance, 37*(8), 2750–2764.

Escobar, D., & Pflug, G. C. (2020). The distortion principle for insurance pricing: properties, identification and robustness. *Annals of Operations Research, 292*(3), 771–794.

Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics, 18*(2), 141–153.

Haberman, S., Khalaf-Allah, M., & Verrall, R. (2011). Entropy, longevity and the cost of annuities. *Insurance: Mathematics and Economics, 48*(2), 197–204.

Kaas, R. (1993). How to (and how not to) compute stop-loss premiums in practice. *Insurance: Mathematics and Economics, 13*(3), 241–254.

Keilson, J., & Gerber, H. (1971). Some results for discrete unimodality. *Journal of the American Statistical Association, 66*(334), 386–389.

Li, J. (2014). A quantitative comparison of simulation strategies for mortality projection. *Annals of Actuarial Science, 8*(2), 281–297.

Li, L., Shao, H., Wang, R., & Yang, J. (2018). Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics, 09*(1), 146–158.

Liu, H., & Wang, R. (2017). Collective risk models with dependence uncertainty. *ASTIN Bulletin: The Journal of the IAA, 47*(2), 361–389.
López-Díaz, M., Sordo, M. A., & Suárez-Llorens, A. (2012). On the lp-metric between a probability distribution and its distortion. *Insurance: Mathematics and Economics, 51*(2), 257–264.
Olivieri, A. (2001). Uncertainty in mortality projections: an actuarial perspective. *Insurance: Mathematics and Economics, 29*(2), 231–245.
Olivieri, A., & Pitacco, E. (2011). *Introduction to insurance mathematics: Technical and financial features of risk transfers*. Springer.
Pichler, A. (2014). Insurance pricing under ambiguity. *European Actuarial Journal, 4*, 335–364.
Pitacco, E., Denuit, M., Haberman, S., & Olivieri, A. (2009). *Modeling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press.
Puccetti, G., Rüschendorf, L., & Manko, D. (2016). Var bounds for joint portfolios with dependence constraints. *Dependence Modeling, 4*(1), 000010151520160021.
Rabitti, G., & Borgonovo, E. (2020). Is mortality or interest rate the most important risk in annuity models? a comparison of sensitivity analysis methods. *Insurance: Mathematics and Economics, 95*, 48–58.
Rachev, S., Stoyanov, S., & Fabozzi, F. (2008). *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization: The Ideal Risk, Uncertainty, and Performance Measu res*. Wiley.
Wang, R., Peng, L., & Yang, J. (2013). Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. *Finance and Stochastics, 17*, 395–417.
Wang, S. S. (2002). A universal framework for pricing financial and insurance risks. *ASTIN Bulletin, 32*(2), 213–234.
Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies, 32*(2), 137–150.
Yang, J., Zhuang, W., & Hu, T. (2014). Lp-metric under the location-independent risk ordering of random variables. *Insurance: Mathematics and Economics, 59*, 321–324.
Young, V. (2014). *Premium Principles*. Wiley StatsRef: Statistics Reference Online.