Insurance Pricing under Ambiguity

Alois Pichler [a,b,*]

[a] *University of Vienna, Austria. Department of Statistics and Operations Research*
[b] *Actuary. Member of the Austrian Actuarial Association*

Abstract

Starting from the equivalence principle various pricing methods have been considered in insurance, each aiming at finding a fair price and reflecting the intrinsic risk of an insurance contract at an appropriate level. In this paper we consider first risk measures, in particular the Conditional Tail Expectation in order to establish a risk averse price for an insurance contract. We elaborate explicit formulas for important types of life insurance contracts and observe that distorted probability distributions play an essential role.

In order to determine a risk averse premium the actuary is moreover interested in computing the premium under related probability distributions. We elaborate that a suitable concept to qualify related probability distributions is provided by the Wasserstein distance. For this distance the actuary often is able to control and quantify the distance to the real probability distribution in an explicit way. We elaborate the pricing issues in this context and give reliable, sharp bounds for a risk averse price, which are easily available for the actuary.

Keywords: Risk Measures, Continuity properties, Wasserstein and Kantorovich Distance, Distortion Function
2010 MSC: 62P05, 60B05, 62E17

1. Introduction

How should one price an insurance contract? An important and fair answer to this question is provided by the equivalence principle, that is to say the fair lump sum premium for an insurance contract is just the expected value of its possible losses. Besides the fact that other loadings have to be added to the premium (for, say, administration and acquisition of the contract), the realized loss of a concrete, individual portfolio will differ from the premiums, the receipts and expenditures of the entire company naturally differ.

Although many other financial products (other than insurance contracts) are priced by the same method as the equivalence principle, the variance of potential outcomes was investigated significantly earlier in insurance than in banking: Hattendorff presented his result on the variance of the realized loss already in 1868 (cf. [Hat68], [GLS03] on Hattendorff's Theorem). This result naturally generalizes to the loss distribution of an entire portfolio, and insurance companies absorb individual losses, as long as they do not exceed a specified amount: The central limit theorem justifies describing the loss distribution of an insurance company approximately.

To account for individual insurance contracts with high individual exposure – or for reinsurance contracts – alternative techniques or methods have been elaborated in the literature to establish an adjusted price, a risk-adjusted, risk-averse or risk-based price. Examples include utility functions or appropriate transforms, cf. [Den89] and [Wan00]. Other methods overvalue potential losses, for example by employing risk measures, which have been introduced in [ADEH99] for general financial contracts. The most prominent among those risk measure is the Conditional Tail Expectation (CTE) [1]

*Corresponding Author
URL: `http://isor.univie.ac.at/` (Alois Pichler)
[1]The

(cf. [KRS09] for its statistical properties), which is now part of the US and Canada insurance industry regulations.

The loss distribution is an essential ingredient for all of these computations. However, the loss distribution is not known precisely enough in many situation of actuarial relevance, available are typically a few empirical observations or the associated empirical distribution up to some precision. For this reason it is essential for an actuary to understand the impact of the loss distribution on the premium and on the reserve for both, pricing and reserving thoroughly and carefully enough, that is to say with actuarial due diligence.

In this paper we study the impact of probability distributions on prices and risk functions. We shall recall the notion of risk measures in Section 2 and provide explicit evaluations for important types of life insurance contracts in Section 3. Section 5 introduces the Wasserstein distance to measure the distance of probability distributions, and the following section addresses pricing issues (risk measures) under ambiguity. Concrete examples throughout the paper and a numerical discussion at its end illustrate the results.

2. Risk Measures

We consider $\mathbb{R}-$valued random variables L, in the context of insurance a random variable is typically associated with loss. Risk measures, which have been introduced [ADEH99] and discussed in a flood of scientific papers since then, assign a real number to any potential loss L. In the following we consider the loss functions L on spaces such as $L \in \mathscr{L} = \mathscr{L}_p(\Omega, \Sigma, P)$ for $1 \le p < \infty$ or $\mathscr{L} = \mathscr{L}_\infty(\Omega, \Sigma, P)$, although there are some topological differences. This does not matter in the present context of insurance, as any loss distribution of practical importance satisfies $L \in \mathscr{L}_\infty \subset \mathscr{L}_p \subset \mathscr{L}_1$; however, we shall address the topological differences at the appropriate place in the paper.

Definition 1 (Risk measures)**.** A risk measure $\rho\colon \mathscr{L} \to \mathbb{R} \cup \{\infty\}$ is a function assigning a real number (or ∞) $\rho(L)$ to a $\mathbb{R}-$valued random variable $L \in \mathscr{L}$. A risk measure has the following properties (i) – (iv):

(i) Monotonicity: If $L_1 \le L_2$, then $\rho(L_1) \le \rho(L_2)$ [2] $(L_1, L_2 \in \mathscr{L})$;
(ii) Convexity: $\rho(\lambda L_1 + (1-\lambda)L_0) \le \lambda\rho(L_1) + (1-\lambda)\rho(L_0)$ whenever $0 \le \lambda \le 1$ $(L_0, L_1 \in \mathscr{L})$;
(iii) Translation Equivariance: If $y \in \mathbb{R}$, then $\rho(L+y) = \rho(L) + y$; [3]
(iv) Positive Homogeneity: For $\lambda > 0$, $\rho(\lambda L) = \lambda \cdot \rho(L)$ $(L \in \mathscr{L})$.
(v) Version Independence: A risk functional ρ is *version independent* [4], if $\rho(L_1) = \rho(L_2)$ whenever L_1 and L_2 share the same law, that is $P(L_1 \le y) = P(L_2 \le y)$ for all $y \in \mathbb{R}$.

In this definition we allow ρ to evaluate to ∞. However, throughout this paper we shall assume that ρ is *proper*, that is there is at least one $L \in \mathscr{L}$ for which $\rho(L) < \infty$.

- *Conditional Tail Expectation (CTE)*

is sometimes also called

- *Average Value-at-Risk, or conditional Value-at-Risk,*
- *expected shortfall,*
- *tail value-at-risk* or newly
- *super-quantile.*

[2] For L_1 and L_2 random variables we write $L_1 \le L_2$ whenever $L_1(k) \le L_2(k)$ for all samples k.
[3] The random variable $L + y$ is $L + y \cdot \mathbb{1}$ where $\mathbb{1}$ is the constant random variable, $\mathbb{1}(k) = 1$.
[4] sometimes also *law invariant* or *distribution based.*

2.1. The Conditional Tail Expectation, CTE_α

The most prominent example of a risk measure is the Conditional Tail Expectation, CTE. The α-level CTE ($0 \le \alpha < 1$) is given via the equivalent representations

$$\begin{aligned}\text{CTE}_\alpha(L) :&= \frac{1}{1-\alpha}\int_\alpha^1 F_L^{-1}(p)\,\mathrm{d}p && (1)\\ &= \inf_{q\in\mathbb{R}}\ q + \frac{1}{1-\alpha}\mathbb{E}(L-q)_+ && (2)\\ &= \sup\left\{\mathbb{E}\,LZ\colon\ 0\le Z\le(1-\alpha)^{-1},\ \mathbb{E}Z=1\right\} && (3)\\ &= \sup\left\{\mathbb{E}_Q L\colon\ \frac{\mathrm{d}Q}{\mathrm{d}P}\le\frac{1}{1-\alpha}\right\} && (4)\end{aligned}$$

where $F_L^{-1}(\alpha) = q_\alpha(L) = \inf\{q\colon P(L\le q)\ge\alpha\}$ is the left-continuous $\alpha-$quantile (sometimes also the *Value-at-Risk* or *lower inverse* cdf at level α).

In case that $F_L(q_\alpha(L)) = \alpha$ (which is certainly the case if F_L is continuous at q_α) the additional formula

$$\text{CTE}_\alpha(L) = \mathbb{E}[L|\,L\ge q_\alpha(L)] \tag{5}$$

is often useful.

The equivalence of these different representations (1) – (5) can be found, e.g., in [RU00] and [PR07].

The function $\alpha\mapsto(1-\alpha)\cdot\text{CTE}_\alpha(L)$ – by (1) – is concave. It is moreover linear between neighboring values of $\alpha\in\{F_L(x) : x\in\mathbb{R}\}$, which is an essential observation when evaluating $\text{CTE}_\alpha(L)$ by means of (5) for discrete distributions.

The CTE at level 1 is defined as

$$\text{CTE}_1(L) := \lim_{\alpha\to1}\text{CTE}_\alpha(L) = \operatorname{ess\,sup} L.$$

2.2. Interpretation in insurance

In the context of insurance L is a random variable associated with loss. It follows from the definition that

$$\text{CTE}_0(L) = \mathbb{E}\,L \le \text{CTE}_\alpha(L)\le\text{CTE}_1(L) = \operatorname{ess\,sup} L,$$

which shows that $\text{CTE}_0(L)$ is the price of the insurance contract according the equivalence principle. $\text{CTE}_\alpha(L)$ is a higher lump sum premium for the insurance contract, incorporating a certain loading by judging the insurance contract to be risky for the company: According to (1) outcomes with higher losses ($\{L>q_\alpha\}$) are over-weighted with $\frac{1}{1-\alpha}>1$, whereas outcomes with smaller losses ($\{L<q_\alpha\}$) are not even considered, they are ignored by CTE_α. Naturally the contract's price cannot exceed $\text{CTE}_1(L)$, as this is the maximum potential loss anyhow and no downside risk can be associated with the premium $\text{CTE}_1(L)$ for the insurance company.

3. Applications in life insurance

An insurance contract often has a monotone payoff function: if the insured lives longer, then she/ he will receive more cash (monotone increasing) – or less (monotone decreasing) – depending on the contract. This monotonicity property allows to compute the Conditional Tail Expectation for some important life insurance contracts in an explicit way, the results often come along with a helpful interpretation, as changing the measure results in a simple modification of the life table in an appropriate way. Moreover the results allow evaluating the CTE by employing simple and standard actuarial tools involving just standard present values. For this we list the results for four elementary insurance types.

The payoff L, representing the insurer's loss, depends on the lifetime K. As usual in actuarial science we assume the lifetime K integer valued and write $L_k := L(k)$ when convenient in this situation. We employ the international, standard actuarial notation (basically following [Ger97]), such as ${}_kp_x = P(K \geq k)$ and assume throughout the paper that

$$P(K < \infty | K \geq x) = \sum_{k=0} {}_kp_x\, q_{x+k} = 1 \qquad (x \in \mathbb{N}). \tag{6}$$

3.1. Increasing payoff functions

For an increasing payoff function L it holds that $\{L \leq L(k)\} = \{K \leq k\}$ by monotonicity, and hence $P(L \leq L(k)) = P(K \leq k) = {}_kq_x$ for any fixed $k \in \{0, 1 \dots\}$: The quantile for $\alpha = {}_kq_x$ is $F_L^{-1}({}_kq_x) = L(k)$.

(i) Pure endowment
The random variable for the payoff of a pure endowment is $L(K) := v^n \cdot \mathbb{1}_{\{K \geq n\}}$, which is an increasing function in K (a step function). The lump sum for the endowment insurance, according the equivalence principle, is $\mathbb{E}L = {}_nE_x$. Obviously $F_L^{-1}(p) = \begin{cases} 0 & \text{if } p < {}_nq_x \\ v^n & \text{if } p \geq {}_nq_x \end{cases}$, such that the Conditional Tail Expectation evaluates to

$$\mathrm{CTE}_\alpha(L) = \begin{cases} \frac{{}_np_x}{1-\alpha} v^n & \text{if } \alpha \leq {}_nq_x \\ v^n & \text{if } \alpha \geq {}_nq_x \end{cases}$$

by (1). It follows immediately for $\alpha = {}_kq_x$ that

$$\mathrm{CTE}_{{}_kq_x}(L) = \begin{cases} \frac{{}_np_x}{{}_kp_x} v^n & \text{if } k \leq n \\ v^n & \text{if } k \geq n \end{cases} = \begin{cases} v^k \cdot {}_{n-k}E_{x+k} & \text{if } k \leq n \\ v^n & \text{if } k \geq n, \end{cases} \tag{7}$$

which is the price for an endowment with sum insured due given the insured survives n years, or dies within the first k years. The CTE_α of a general level α can be found by appropriately interpolating (7).

(ii) Life Annuities
The payoff for an whole life annuity-due is $L = \ddot{a}_{\overline{K+1}|} = \sum_{j=0}^K v^j$, its present value is $\mathbb{E}\, \ddot{a}_{\overline{K+1}|} = \ddot{a}_x$. The payoff function $k \mapsto \ddot{a}_{\overline{k+1}|}$ is increasing, as for the pure endowment. Hence, by (5),

$$\begin{aligned} \mathrm{CTE}_{{}_kq_x}(L) &= \mathbb{E}\left[L \middle| L \geq \ddot{a}_{\overline{k+1}|}\right] = \mathbb{E}\left[\ddot{a}_{\overline{K+1}|} \middle| K \geq k\right] \\ &= \frac{1}{{}_kp_x} \sum_{j=k} {}_jp_x\, q_{x+j}\, \ddot{a}_{\overline{j+1}|} = \frac{1}{{}_kp_x} \sum_{j=0} {}_{j+k}p_x\, q_{x+k+j}\, \ddot{a}_{\overline{k+j+1}|} \\ &= \sum_{j=0} {}_jp_{x+k}\, q_{x+k+j} \left(\ddot{a}_{\overline{k}|} + v^k\, \ddot{a}_{\overline{j+1}|}\right) = \ddot{a}_{\overline{k}|} + v^k\, \ddot{a}_{x+k}. \end{aligned} \tag{8}$$

This is the premium for an annuity with modified mortality, the insured is assumed *immortal* in the first k years. By the elementary relation $A_x = 1 - d \cdot \ddot{a}_x$ this is equivalent to the expression

$$\mathrm{CTE}_{{}_kq_x}(L) = \ddot{a}_{\overline{k}|} + v^k\, \ddot{a}_{x+k} = \ddot{a}_{x+k} + A_{x+k}\, \ddot{a}_{\overline{k}|}.$$

For the more general $n-$year temporary life annuity-due with present value $\ddot{a}_{x:\overline{n}|} = \mathbb{E}\, \ddot{a}_{\overline{\min\{n,K+1\}}|}$ the CTE evaluates to

$$\mathrm{CTE}_{{}_kq_x}(L) = \begin{cases} \ddot{a}_{\overline{k}|} + v^k\, \ddot{a}_{x+k:\overline{n-k}|} & \text{if } k \leq n \\ \ddot{a}_{\overline{n}|} & \text{if } k \geq n. \end{cases}$$

After these illustrating examples it is evident that the general pattern for increasing payoff function L is

$$\mathrm{CTE}_{\,_kq_x}(L) = \mathbb{E}\left[L \mid L \geq L(k)\right] = \tilde{\mathbb{E}}L,$$

where $\tilde{\mathbb{E}}L$ is the usual present value with respect to the measure which assumes that the insured is immortal within the first k years: the modified life table $\tilde{q}$ to compute $\tilde{\mathbb{E}}L$ as in (7) and (8) is

$$\tilde{q} = (\underbrace{0,\, 0, \dots 0}_{k \text{ times}},\, q_{x+k},\, q_{x+k+1}, \dots). \tag{9}$$

For this modified life table it should be noted that

$$\tilde{P}(K < \infty) = \sum_{k=0} {}_k\tilde{p}_x\, \tilde{q}_{x+k} = 1 \tag{10}$$

by (6), which means that any outcome will die with probability 1 under the modified measure as well.

3.2. Decreasing payoff functions

If the payoff function is decreasing then $\{L \leq L(k)\} = \{K \geq k\}$ by monotonicity, and consequently $P(L \leq L(k)) = P(K \geq k) = {}_kp_x$ for any fixed $k \in \{0, 1 \dots\}$. The quantile for $\alpha = {}_kp_x$ thus is given by $F_L^{-1}({}_kp_x) = L(k)$.

(i) Life insurance
The payoff for a whole life insurance contract is $L := v^{K+1}$, its present value is $\mathbb{E}L = A_x$. The payoff $k \mapsto v^{k+1}$ is *de*creasing. Then the quantile for $\alpha = {}_kp_x$ is $q_\alpha(L) = v^{k+1}$, such that

$$\mathrm{CTE}_{_kp_x}(L) = \mathbb{E}\left[L \mid L \geq v^{k+1}\right] = \mathbb{E}\left[v^{K+1} \mid K \leq k\right] = \frac{{}_kA_x}{{}_kq_x} \tag{11}$$

by employing (5).
For the more general term life insurance ($L = v^{K+1} \cdot \mathbb{1}_{\{K \leq n\}}$) one deduces the closed form

$$\mathrm{CTE}_{_kp_x}(L) = \begin{cases} \frac{{}_nA_x}{{}_kq_x} & \text{if } k \geq n, \\ \frac{{}_kA_x}{{}_kq_x} & \text{if } k \leq n \end{cases} \tag{12}$$

for the Conditional Tail Expectation.

(ii) Endowment
Endowment insurance combines life insurance and a pure endowment by use of the decreasing payoff $L = v^{\min\{K+1, n\}}$. Its CTE generalizes (11) and (12) by

$$\mathrm{CTE}_{_kp_x}(L) = \begin{cases} \frac{{}_nA_x + {}_nE_x \cdot {}_{k-n}q_{x+n}}{{}_kq_x} & \text{if } k \geq n, \\ \frac{{}_kA_x}{{}_kq_x} & \text{if } k \leq n. \end{cases}$$

To describe the general pattern for a decreasing payoff function L consider the modified life table

$$\tilde{q} = (q_x,\, q_{x+1}, \dots q_{x+k-1},\, 0,\, 0, \dots), \tag{13}$$

for which $\tilde{P}(K < \infty) = {}_kq_x \leq 1$, which is in notable contrast to (10). It follows that $\tilde{P} = \sum_k \frac{{}_k\tilde{p}_x\, \tilde{q}_{x+k}}{{}_k\tilde{q}_x} \delta_k$ is a probability measure for which

$$\mathrm{CTE}_{_kp_x}(L) = \tilde{\mathbb{E}}L.$$

3.3. General insurance contracts

The methods developed in the latter Sections 3.1 and 3.2 can be applied to general insurance contracts, although closed forms are more involved or simply not available in many situations. Just consider an endowment insurance with regular premium Π. Its payoff

$$L^{\Pi} = v^n \cdot \mathbb{1}_{\{K \geq n\}} - \Pi \cdot \sum_{j=0}^{\min\{K,n-1\}} v^j$$

is neither de-, nor increasing. However, the $\mathrm{CTE}_\alpha\left(L^{\Pi}\right)$ can be easily computed by appropriately ordering the potential payoffs and applying (5):

Denote $P = \sum_k {}_kp_x\, q_{x+k} \cdot \delta_k =: \sum_k p_k \delta_k$ the probability measure and $\cdot \mapsto (\cdot)$ the permutation such that

$$L_{(1)} \geq L_{(2)} \geq L_{(3)} \geq \dots .$$

Define

$$Z_{(k)} := \begin{cases} \frac{1}{1-\alpha} & \text{if } p_{(1)} + \dots p_{(k)} \leq 1-\alpha, \\ 0 & \text{if } p_{(1)} + \dots p_{(k-1)} \geq 1-\alpha \end{cases} = \begin{cases} \frac{1}{1-\alpha} & \text{if } P\left(L \geq L_{(k)}\right) \leq 1-\alpha, \\ 0 & \text{if } P\left(L \geq L_{(k-1)}\right) \geq 1-\alpha \end{cases} \tag{14}$$

with the supplementary requirement that $\mathbb{E}\, Z = \sum_k p_k\, Z_k = 1$ for the remaining index. Then, by (3),

$$\mathrm{CTE}_\alpha(L) = \mathbb{E}\, LZ = \sum_k {}_kp_x\, q_{x+k}\, L_k\, Z_k. \tag{15}$$

Equivalently, by employing the modified measure

$$\tilde{P} := \sum_k {}_kp_x\, q_{x+k}\, Z_k \cdot \delta_k$$

it holds – in accordance with (4) – that

$$\mathrm{CTE}_\alpha(L) = \tilde{\mathbb{E}}\, L.$$

The measure $\tilde{P}$ can be incorporated in a life table $\tilde{q}$ again such that

$$\tilde{P} = \sum_k {}_kp_x\, q_{x+k}\, Z_k \cdot \delta_k = \sum_k {}_k\tilde{p}_x\, \tilde{q}_{x+k} \cdot \delta_k,$$

but in this general situation the pattern is not as striking as in (9) and (13).

3.4. Pricing and adapted reserving strategies

Risk averse pricing

A lump sum premium Π_α can be chosen such that

$$\mathrm{CTE}_\alpha\left(L^{\Pi_\alpha}\right) = 0, \tag{16}$$

where $L^{\Pi_\alpha} := L - \Pi_\alpha$ is the loss function with incorporated premium payment at the very beginning of the contract. It is evident from the axiom on translation equivariance in Definition 1 that $\Pi_\alpha = \mathrm{CTE}_\alpha(L)$.

For regularly paid premiums the annual premium Π_α as well can be chosen such that

$$\mathrm{CTE}_\alpha\left(L^{\Pi_\alpha}\right) = 0, \tag{17}$$

where $L^{\Pi_\alpha} = L - \Pi_\alpha \cdot \ddot{a}_{\overline{\min\{K,n-1\}}|}$. It follows then that $0 = \mathrm{CTE}_\alpha\left(L^{\Pi_\alpha}\right) \geq \mathbb{E}L^{\Pi_\alpha} = \mathbb{E}L - \Pi_\alpha \ddot{a}_{x:\overline{n}|}$, i.e.

$$\Pi_\alpha \geq \frac{\mathbb{E}L}{\ddot{a}_{x:\overline{n}|}} = \Pi_0,$$

which demonstrates that the annual premium Π_α according (17) is higher than the annual premium Π_0 proposed by the equivalence principle.

The equations (16) and (17) represent a risk-averse pricing strategy, a generalization of the equivalence principle, which is discussed in many places of actuarial literature.

The reserves corresponding to risk averse premiums

It is tempting to define the balance sheet reserves (net premium reserve) after Δ years in line with the pricing formulae (16) and (17) as $V_\Delta = \mathrm{CTE}_\alpha \left(L_\Delta^\Pi \middle| K \geq \Delta \right)$, where the measure is restricted to the future lifetime $\{K \geq \Delta\}$ and the payoff L_Δ^Π is adjusted to account for the future cash flow. This is just in line with the net premium reserve for the equivalence principle. But this is misleading, the setting violates important time-consistency constraints ([Sha12]), which may cause undesirable jumps in the company's annual balance sheets, and consecutive profit and loss accounts. The proper adjustment has to accept an annually adjusted level α_Δ. The reserves, adjusted for time consistency and in line with (16) and (17), are

$$V_\Delta := \mathrm{CTE}_{\alpha_\Delta} \left(L_\Delta^\Pi \middle| K \geq \Delta \right) = \tilde{\mathbb{E}} \left(L_\Delta^\Pi \middle| K \geq \Delta \right),$$

where $\tilde{P}$ is the modified measure.

For a comprehensive discussion and justifications of these settings, in particular for a time consistent choice of the sequence α_Δ we refer to [PP12b] and [PP12a].

4. Further risk functionals

It is obvious that averaging risk functionals, for example

$$\rho_\lambda (L) := (1-\lambda)\, \mathrm{CTE}_{\alpha_0} (L) + \lambda\, \mathrm{CTE}_{\alpha_1} (L),$$

(where λ, α_0, $\alpha_1 \in [0,1]$) are risk functional again, and so is

$$\rho_{\mathcal{G}} (L) := \sup_{g \in \mathcal{G}} \rho_g (L),$$

where ρ_g is a risk measure for every $g \in \mathcal{G}$.

Kusuoka's theorem states that every risk functional can be obtained by combining the latter two operations, provided it is version independent.

Theorem 1 (Kusuoka)**.** *Any law invariant, positively homogeneous and lower semi-continuous risk functional ρ on $\mathscr{L}$ obeys the representation*

$$\rho (L) = \sup_{\mu \in \mathscr{M}} \int_0^1 \mathrm{CTE}_\alpha (L)\, \mu (\mathrm{d}\alpha),$$

where $\mathscr{M}$ is a set of probability measures on $[0,1]$.

Moreover

$$\rho (L) = \sup_{h \in \mathscr{H}} \int_0^1 F_L^{-1} (\alpha)\, h (\alpha)\, \mathrm{d}\alpha, \tag{18}$$

where $\mathscr{H}$ is a set of positive, continuous, bounded and increasing functions h satisfying $\int_0^1 h(\alpha)\, \mathrm{d}\alpha = 1$.

For positive loss functions $P(L \geq 0) = 1$ the representation

$$\rho (L) = \sup_{h \in \mathscr{H}} \int_0^\infty H_h (F_L (q))\, \mathrm{d}q,$$

holds in addition, where $H_h(\alpha) = \int_\alpha^1 h(p)\, \mathrm{d}p$.

Proof. As for the proof cf. [Kus01] and [Pfl06]. The latter statement follows from (18) by means of integration by parts and substitution. □

Kusuoka's representation demonstrates that the Conditional Tail Expectation is indeed central, as all other risk measures can be expressed by employing the CTE at different levels α. This observation fosters the central role of the CTE. Moreover it is inherited from the Conditional Tail Expectation that

$$\mathbb{E}\,L \leq \rho\,(L) \leq \operatorname{ess\,sup} L,$$

which makes $\rho\,(L)$ a candidate for a more involved insurance price: As above, the price according the equivalence principle, $\mathbb{E}\,L$, is a lower bound for the risk-adjusted price. Moreover any of these risk measures ρ overestimate the impact of expensive losses in relation, and in contrast to comparably cheap losses.

Example. A comprehensive list of Kusuoka representations for important risk functionals is provided in [PR07]. A compelling example is the absolute semi-deviation risk measure for some fixed $c \in [0, 1]$,

$$\rho_c\,(L) := \mathbb{E}\left[L + c \cdot (L - \mathbb{E}L)_+\right],$$

assigning an additional loading of c to any loss L exceeding the price $\mathbb{E}L$, computed according the equivalence principle. Its Kusuoka representation is (cf. [Sha12, SDR09])

$$\rho_c\,(L) = \sup_{0 \leq \lambda \leq 1} (1 - c\,\lambda)\,\mathbb{E}L + c\,\lambda \cdot \mathrm{CTE}_{1-\lambda}\,(L)\,.$$

5. Ambiguity

In the previous sections we have proposed $\rho\,(L)$, in particular $\mathrm{CTE}_\alpha\,(L)$ as a risk averse premium for an insurance contract. It was an important observation, which is already intrinsic in (4), that the probability measure can be exchanged in order to evaluate the Conditional Tail Expectation of a loss distribution L.

The following point on L's distribution function has to be made as well: To compute $\mathbb{E}\,L$, $\mathrm{CTE}_\alpha\,(L)$ or $\rho\,(L)$ the distribution function $F_L\,(x) = P\,(L \leq x)$ has to be known precisely: In order to price a contract the actuary has to be sure about the loss distribution F_L.

In a real world situation this is, however, not often the case. The loss distribution may result from empirical observations, may have errors in measurement or may be derived from other observations related to the risk of interest. In extreme situations the loss distribution may even be derived by scientific arguments without any observations. In order to price a contract the responsible actuary thus will compute the prices for other possible loss distributions $\tilde{F}_L$, say, as well and compare the resulting premiums.

The actuary is interested in the question: What is the price $\mathrm{CTE}_\alpha\,(L)$ for the given loss function L, but subject to a different probability distribution? To investigate this question a notion of distance of probability distributions is advisable. Various distances have been proposed in the mathematical literature, [Rac91, RSF11] provide an impressive overview and aggregation (applications can be found in [PP11] and in [LDSSL12]). In the present actuarial context the Wasserstein distance is appropriate, as it obeys a sequence of properties one would intuitively request from a distance of probability measures having applications in insurance in mind.

Definition 2 (Wasserstein distance)**.** Let (Ξ, Σ, P) and $\left(\tilde{\Xi}, \tilde{\Sigma}, \tilde{P}\right)$ be probability spaces and $d\colon \Xi \times \tilde{\Xi} \to \mathbb{R}$ a distance function. The Wasserstein $r-$distance ($r \geq 1$) is

$$\mathsf{d}_r\left(P, \tilde{P}\right)^r := \inf_\pi \int_{\Xi \times \tilde{\Xi}} d\left(\xi, \tilde{\xi}\right)^r \pi\left(\mathrm{d}\xi, \mathrm{d}\tilde{\xi}\right),$$

where the infimum is among all probability measures π which have marginals P and $\tilde{P}$, that is $\pi\left(A \times \tilde{\Xi}\right) = P\,(A)$ and $\pi\,(\Xi \times B) = \tilde{P}\,(B)$ whenever $A \in \Sigma$ and $B \in \tilde{\Sigma}$.

Mathematical and topological details for the Wasserstein distance can be found in [Villani Vil03]. In the Russian literature (cf. [Ver06]) the Wasserstein distance is usually called Kantorovich distance.

We adopt the concept with actuarial applications in mind.

Remark 1. If $P=\sum_i p_i\delta_{\xi_i}$ and $\tilde{P}=\sum_j \tilde{p}_j\delta_{\tilde{\xi}_j}$ are discrete measures, then the Wasserstein distance can be computed by the linear program (LP)

$$\begin{array}{rl} \text{minimize} & \\ (\text{in } \pi) & \sum_{i,j} d_{i,j}^r \pi_{i,j} = \mathbb{E}_\pi d^r \\ \text{subject to} & \sum_j \pi_{i,j} = p_i, \\ & \sum_i \pi_{i,j} = \tilde{p}_j, \\ & \pi_{i,j} \geq 0, \end{array} \tag{19}$$

where $d_{i,j}$ is the matrix with entries $d_{i,j}=d\left(\xi_i,\tilde{\xi}_j\right)$. As any linear program it has a dual program, which can be stated as

$$\begin{array}{rl} \text{maximize} & \\ (\text{in } \lambda,\mu) & \sum_i p_i\lambda_i + \sum_j \tilde{p}_j\mu_j = \mathbb{E}_P\lambda + \mathbb{E}_{\tilde{P}}\mu \\ \text{subject to} & \lambda_i + \mu_j \leq d_{i,j}^r. \end{array} \tag{20}$$

In all situations relevant for insurance we have that $\Xi=\tilde{\Xi}$ is a vector space, and d is a norm, that is $d\left(\xi,\tilde{\xi}\right)=\left\|\xi-\tilde{\xi}\right\|$. In this situation the following holds true.

Lemma 1. *Let $\Xi=\tilde{\Xi}$ be a vector space and the distance d a norm. Then*

$$\|e_P-e_{\tilde{P}}\| \leq \mathsf{d}_r\left(P,\tilde{P}\right), \tag{21}$$

where $e_P:=\int xP\left(\mathrm{d}x\right)$ is the barycenter under P.

Proof. By Jensen's inequality,

$$\begin{aligned} \|e_P-e_{\tilde{P}}\| &= \left\|\int xP\left(\mathrm{d}x\right)-\int y\tilde{P}\left(\mathrm{d}y\right)\right\| \\ &= \left\|\int x-y\;\pi\left(\mathrm{d}x,\mathrm{d}y\right)\right\| \leq \int \|x-y\|\,\pi\left(\mathrm{d}x,\mathrm{d}y\right), \end{aligned}$$

whenever π has the right marginals according to Definition 2. Taking the infimum with respect to all possible measures π reveals (21) for $r=1$. The general case deduces from

$$\mathsf{d}_r\left(P,\tilde{P}\right) \leq \mathsf{d}_{r'}\left(P,\tilde{P}\right) \qquad \left(r\leq r'\right),$$

which is a consequence of Hölder's inequality. □

Interpretation of the Wasserstein distance in life insurance

In life insurance the probability measure $P=\sum_k {}_kp_x\, q_{x+k}\cdot\delta_k$ is usually given via a life table. To compare life tables it is natural to employ the function $d\left(k,\tilde{k}\right)=\left|k-\tilde{k}\right|$, which measures the difference in age of two sample-individuals. For this choice (21), i.e.

$$|e_P-e_{\tilde{P}}| \leq \mathsf{d}_r\left(P,\tilde{P}\right),$$

relates the average life expectancy under P and $\tilde{P}$, as $e_P=\sum_k k\cdot {}_kp_x q_{x+k}$ is the average life expectancy (under P): The average life expectancy under different measures will not differ more than their Wasserstein distance.

The following example further supports the interpretation that $\mathsf{d}_r\left(P,\tilde{P}\right)$ reflects the average difference of life expectancies of different generations.

Example. The life table of a different generation is in practical situations often approximated by an age-shift of, say, Δ years (age-adjusted life expectation). The shifted probability measure is $\tilde{P}=\sum_k {}_kp_x\, q_{x+k}\cdot\delta_{k+\Delta}$. Then the Wasserstein distance recovers the age-shift,

$$\mathsf{d}_r\left(P,\tilde{P}\right)=\Delta, \tag{22}$$

when employing the distance function $d\left(k,\tilde{k}\right)=\left|k-\tilde{k}\right|$ representing the natural difference in age.

To accept (22) choose $\pi_{k,\tilde{k}}:=\begin{cases}p_k & \text{if } k-\tilde{k}=\Delta\\ 0 & \text{else}\end{cases}$ and observe that π is feasible for (19) with the objective $\mathbb{E}_\pi d^r=\sum_{k,\tilde{k}}\pi_{k,\tilde{k}}\left|k-\tilde{k}\right|^r=\sum_k p_k\Delta^r=\Delta^r$, so $\mathsf{d}_r\left(P,\tilde{P}\right)\leq\Delta$.

On the other side choose $\lambda_k:=k\cdot r\Delta^{r-1}$ and $\mu_{\tilde{k}}:=-\left(r-1\right)\Delta^r-\tilde{k}\cdot r\Delta^{r-1}$. Then $\lambda_k+\mu_{\tilde{k}}=\Delta^r+r\Delta^{r-1}\left(k-\tilde{k}-\Delta\right)\leq\left|k-\tilde{k}\right|^r$ by convexity of $y\mapsto|y|^r$. The dual variables λ and μ thus are feasible for the dual program (20). The objective for the dual is $\mathbb{E}_P\lambda+\mathbb{E}_{\tilde{P}}\mu=\sum_k {}_k p_x\, q_{x+k} kr\Delta^{r-1}+\sum_k {}_k p_x\, q_{x+k}\left(\Delta^r\left(1-r\right)-\left(k+\Delta\right)r\Delta^{r-1}\right)=\Delta^r$, which shows that $\mathsf{d}_r\left(P,\tilde{P}\right)\geq\Delta$: the duality gap thus vanishes and $\mathsf{d}_r\left(P,\tilde{P}\right)=\Delta$.

In the context of insurance the loss function L is always $\mathbb{R}-$valued, such that the image measure (pushforward measure) $P^L:=P\circ L^{-1}$ is a measure on $\mathbb{R}$. The following lemma to compute the Wasserstein distance for measures on the real line facilitates evaluating the Wasserstein distance in many situations of practical actuarial importance.

Lemma 2. *For measures P and $\tilde{P}$ on $\mathbb{R}$ it holds that*

$$\mathsf{d}_r\left(P,\tilde{P}\right)^r=\int_0^1\left|F^{-1}\left(\alpha\right)-\tilde{F}^{-1}\left(\alpha\right)\right|^r\mathrm{d}\alpha,$$

where $F\left(x\right)=\int_{-\infty}^x P\left(\mathrm{d}p\right)$, $\tilde{F}\left(x\right)=\int_{-\infty}^x\tilde{P}\left(\mathrm{d}p\right)$ *and* $F^{-1}\left(\alpha\right)=\inf\left\{q\colon F\left(q\right)\geq\alpha\right\}$ *(*$\tilde{F}^{-1}\left(\alpha\right)=\inf\left\{q\colon\tilde{F}\left(q\right)\geq\alpha\right\}$*, resp.) are the quantiles.*

Proof. The proof is contained in [Vil03, Theorem 2.18 and the following remark] or [AGS05] in a more general context. □

This initial and preparing discussion of the Wasserstein distance on probability measures enables a comprehensive treatment of actuarial pricing under distorted probability measures, where the potential distance is a model parameter which can be interpreted and estimated by the actuary. The next section addresses a thorough mathematical treatment of this approach.

6. Ambiguity risk measures

We have proposed in the previous Section 2 $\rho\left(L\right)$ as a price of an individual insurance contract, where ρ is a version independent risk measure. We are ready now to address the question of ambiguity: The loss function L is known precisely and described in the insurance contract. But F_L, its distribution, is not known entirely, but only up to some gap which is specified in terms of the Wasserstein distance.

Definition 3. Let $K\geq 0$ and $\rho\left(L\right)=\sup_{h\in\mathscr{H}}\int F_L^{-1}\left(\alpha\right)h\left(\alpha\right)\mathrm{d}\alpha$ be a version independent risk measure. The $K-$*ambiguity risk measure* is

$$\rho_K\left(L\right):=\sup_{\mathsf{d}_r\left(P,Q\right)\leq K}\rho_Q\left(L\right),$$

where $\rho_Q\left(L\right)=\sup_{h\in\mathscr{H}}\int_0^1 F_{Q;L}^{-1}\left(\alpha\right)h\left(\alpha\right)\mathrm{d}\alpha$ and $F_{Q;L}^{-1}\left(\alpha\right)=\inf\left\{q\colon Q\left(L\leq q\right)\geq\alpha\right\}$.

We have the following theorem.

Theorem 2. *For any $K\geq 0$, ρ_K is a version independent risk measure.*

Proof. Let $L_1\leq L_2$, then $Q\left(L_1\leq q\right)\geq Q\left(L_2\leq q\right)$, and $F_{Q;L_1}^{-1}\left(\alpha\right)\leq F_{Q;L_2}^{-1}\left(\alpha\right)$ for every measure Q, from which the monotonicity property follows.

As for convexity notice that every measure $\rho_{h;Q}(L)=\int_0^1 F_{Q;L}^{-1}(\alpha)\,h(\alpha)\,\mathrm{d}\alpha$ is convex for every Q and $h\in\mathscr{H}$. Thus

$$\begin{aligned}\rho_K\left((1-\lambda)\,L_0+\lambda\,L_1\right)&=\sup_{\mathsf{d}_r(P,Q)\le K}\ \sup_{h\in\mathscr{H}}\rho_{h;Q}\left((1-\lambda)\,L_0+\lambda\,L_1\right)\\&\le\sup_{\mathsf{d}_r(P,Q)\le K}\ \sup_{h\in\mathscr{H}}(1-\lambda)\,\rho_{h;Q}(L_0)+\lambda\rho_{h;Q}(L_1)\\&\le(1-\lambda)\sup_{\mathsf{d}_r(P,Q)\le K}\ \sup_{h\in\mathscr{H}}\rho_{h;Q}(L_0)+\lambda\sup_{\mathsf{d}_r(P,Q)\le K}\ \sup_{h\in\mathscr{H}}\rho_{h;Q}(L_1)\\&=(1-\lambda)\,\rho_K(L_0)+\lambda\,\rho_K(L_1)\,.\end{aligned}$$

Translation equivariance follows from $F_{Q;L+c}^{-1}=c+F_{Q;L}^{-1}$, and positive homogeneity follows from $F_{Q;\lambda\cdot L}^{-1}(\alpha)=\lambda\cdot F_{Q;L}^{-1}(\alpha)$ whenever $\lambda>0$.

It remains to show that ρ_K is version independent. But this is evident by the equivalent expression

$$\rho_K(L)=\sup\left\{\int_0^1\tilde{F}^{-1}(\alpha)\,h(\alpha)\,\mathrm{d}\alpha\ \middle|\ \begin{array}{c}h\in\mathscr{H},\ \tilde{F}^{-1}\text{ an increasing and l.s.c. quantile with}\\ \int_0^1\left|F_L^{-1}(\alpha)-\tilde{F}^{-1}(\alpha)\right|^r\mathrm{d}\alpha\le K^r\end{array}\right\},\tag{23}$$

which holds by means of Lemma 2. □

The $K-$ambiguity measure ρ_K depends on r, the parameter of the Wasserstein distance. To investigate the impact of this parameter we shall write r explicitly as $\rho_{K;r}$ in the following Lemma. It turns out that the highest deviation is to be expected for the parameter $r=1$.

Lemma 3 (The role of the parameter r)**.** *It holds that*

$$\rho_{K;r}(L)\ge\rho_{K;r'}(L)\qquad(r\le r')\,.$$

Proof. Recall that $\mathsf{d}_r(P,Q)\le\mathsf{d}_{r'}(P,Q)$ whenever $r\le r'$, hence

$$\rho_{K;r'}(L)=\sup_{\mathsf{d}_{r'}(P,Q)\le K}\rho_Q(L)\le\sup_{\mathsf{d}_r(P,Q)\le K}\rho_Q(L)=\rho_{K;r}(L)\,,$$

which is the assertion. □

The following theorem particularly holds for the Conditional Tail Expectation.

Theorem 3. *If ρ is generated by a single function h, that is $\rho(L)=\int_0^1 F_L^{-1}(\alpha)\,h(\alpha)\,\mathrm{d}\alpha$, then*

$$K\mapsto\rho_K(L)$$

is concave.

Proof. Recall from (23) that

$$\rho_{(1-\lambda)K_0+\lambda K_1}(L)=\sup\left\{\int_0^1\tilde{F}^{-1}(\alpha)\,h(\alpha)\,\mathrm{d}\alpha\colon\ \left\|F_L^{-1}-\tilde{F}^{-1}\right\|_r\le(1-\lambda)\,K_0+\lambda K_1\right\}$$

where we assume here and in what follows implicitly that $\tilde{F}^{-1}$ needs to be a quantile function.

Let two quantiles satisfying $\left\|F_L^{-1}-\tilde{F}_0^{-1}\right\|_r\le K_0$ and $\left\|F_L^{-1}-\tilde{F}_1^{-1}\right\|_r\le K_1$ be chosen and define

$$\tilde{F}_\lambda^{-1}:=(1-\lambda)\,\tilde{F}_0^{-1}+\lambda\tilde{F}_1^{-1}.$$

Then

$$\left\|F_L^{-1}-\tilde{F}_\lambda^{-1}\right\|_r\le(1-\lambda)\,K_0+\lambda K_1,$$

such that

$$\begin{aligned}\rho_{(1-\lambda)K_0+\lambda K_1}(L) \geq & \sup\left\{\int_0^1 \tilde{F}_\lambda^{-1}(\alpha)\, h(\alpha)\, \mathrm{d}\alpha :\ \left\|F_L^{-1}-\tilde{F}_0^{-1}\right\|_r \leq K_0,\ \left\|F_L^{-1}-\tilde{F}_1^{-1}\right\|_r \leq K_1\right\}. \\ = & \sup\left\{(1-\lambda)\int_0^1 \tilde{F}_0^{-1}(\alpha)\, h(\alpha)\, \mathrm{d}\alpha :\ \left\|F_L^{-1}-\tilde{F}_0^{-1}\right\|_r \leq K_0\right\} \\ & + \sup\left\{\lambda\int_0^1 \tilde{F}_1^{-1}(\alpha)\, h(\alpha)\, \mathrm{d}\alpha :\ \left\|F_L^{-1}-\tilde{F}_1^{-1}\right\|_r \leq K_1\right\}.\end{aligned}$$

Now the assertion

$$\rho_{(1-\lambda)K_0+\lambda K_1}(L) \geq (1-\lambda)\,\rho_{K_0}(L) + \lambda\rho_{K_1}(L)$$

is immediate. □

It turns out that the smoothness of the loss function is of importance when evaluating risk measures with different probability distributions. For this recall the following definition.

Definition 4. A function L is Hölder continuous with exponent $0 \leq \beta$ if there is constant $C \geq 0$ such that $|L(x) - L(y)| \leq C \cdot d(x,y)^\beta$. The smallest of these constant is denoted $H_\beta(L)$, such that

$$|L(x) - L(y)| \leq H_\beta(L) \cdot d(x,y)^\beta .$$

For $\beta = 1$ the notion of Hölder continuity coincides with Lipschitz continuity and $H_1(L)$ is L's Lipschitz constant.

In addition it should be noted that $\beta > 1$ makes sense whenever the space is discrete (any $\beta-$continuous function is constant on $[0,1]$ whenever $\beta > 1$).

Example. The Hölder constant for an annuity ($L(k) = \sum_{j=0}^{k+1} v^k$) is $H_\beta(L) = v < 1$, and for a pure life insurance with loss function $L = v^{k+1}$ the Hölder constant is $H_\beta(L) = v - v^2 = v\,d$.

Theorem 4 (Continuity of risk measures with respect to changing the measure)**.** *Let P and $\tilde{P}$ be probability measures and L Hölder-continuous with exponent $\beta < r$ and constant $H_\beta(L)$. Then*

$$\left|\rho_P(L) - \rho_{\tilde{P}}(L)\right| \leq \|h\|_{r'_\beta} \cdot H_\beta(L) \cdot \mathsf{d}_r\left(P, \tilde{P}\right)^\beta$$

where $r'_\beta \geq \frac{r}{r-\beta}$.

Proof. First let $r'_\beta = \frac{r}{r-\beta}$ and define $r_\beta := \frac{r}{\beta}$, such that $\frac{1}{r_\beta} + \frac{1}{r'_\beta} = 1$. Then, by Hölder's inequality,

$$\begin{aligned}\rho_P(L) - \rho_{\tilde{P}}(L) &= \int_0^1 F_L^{-1}(p)\, h(p) - \tilde{F}_L^{-1}(p)\, h(p)\, \mathrm{d}p \\ &\leq \left(\int_0^1 \left|F_L^{-1}(p) - \tilde{F}_L^{-1}(p)\right|^{r_\beta} \mathrm{d}p\right)^{\frac{1}{r_\beta}} \cdot \left(\int_0^1 h(p)^{r'_\beta}\, \mathrm{d}p\right)^{\frac{1}{r'_\beta}}.\end{aligned}$$

Note now that $P^L = P \circ L^{-1}$ and $\tilde{P}^L = \tilde{P} \circ L^{-1}$ are measures on the real line $\mathbb{R}$ for which, by Lemma 2,

$$\mathsf{d}_{r_\beta}\left(P^L, \tilde{P}^L\right)^{r_\beta} = \int_0^1 \left|F_L^{-1}(p) - \tilde{F}_L^{-1}(p)\right|^{r_\beta} \mathrm{d}p.$$

Moreover

$$\begin{aligned}\mathsf{d}_{r_\beta}\left(P^L, \tilde{P}^L\right)^{r_\beta} &= \int |x-y|^{r_\beta}\, \pi\left(L^{-1}(\mathrm{d}x), L^{-1}(\mathrm{d}y)\right) \\ &= \int |L(x) - L(y)|^{r_\beta}\, \pi(\mathrm{d}x, \mathrm{d}y) \leq \int \left(H_\beta(L) \cdot d(x,y)^\beta\right)^{r_\beta} \pi(\mathrm{d}x, \mathrm{d}y) \\ &= H_\beta(L)^{r_\beta} \cdot \int d(x,y)^r\, \pi(\mathrm{d}x, \mathrm{d}y),\end{aligned}$$

and thus $\mathsf{d}_{r_\beta}\left(P^L,\tilde{P}^L\right)^{r_\beta}\le H_\beta\left(L\right)^{r_\beta}\cdot\mathsf{d}_r\left(P,\tilde{P}\right)^r$ when passing to the infimum. Hence

$$\rho_P\left(L\right)-\rho_{\tilde{P}}\left(L\right)\le\left\Vert h\right\Vert_{r'_\beta}\cdot H_\beta\left(L\right)\cdot\mathsf{d}_r\left(P,\tilde{P}\right)^\beta.$$

The same inequality can be derived with the roles of P and $\tilde{P}$ exchanged, from which the assertion follows for $r'_\beta=\frac{r}{r-\beta}$.

The general situation $r'_\beta\ge\frac{r}{r-\beta}$ derives from the fact that $\left\Vert h\right\Vert_r\le\left\Vert h\right\Vert_{r'}$ whenever $r\le r'$, completing the proof. □

The latter theorem allows the following helpful estimate for the $K-$ambiguity risk measure.

Corollary 1. *Let P and $\tilde{P}$ be probability measures and L Hölder-continuous with exponent $\beta<r$ and constant $H_\beta\left(L\right)$. Then*

$$\rho\left(L\right)\le\rho_K\left(L\right)\le\rho\left(L\right)+K^\beta\cdot\left\Vert h\right\Vert_{r'_\beta}\cdot H_\beta\left(L\right)$$

where $r'_\beta\ge\frac{r}{r-\beta}$.

For the Conditional Tail Expectation the spectral function is $h\left(\cdot\right)=\frac{1}{1-\alpha}\mathbb{1}_{[\alpha,1]}\left(\cdot\right)$, for which $\left\Vert h\right\Vert_{r'_\beta}$ can be given explicitly by

$$\left\Vert h\right\Vert_{r'_\beta}=\left(\int_\alpha^1\left(\frac{1}{1-\alpha}\right)^{r'_\beta}\mathrm{d}\lambda\right)^{1/r'_\beta}=\left(1-\alpha\right)^{\frac{1-r'_\beta}{r'_\beta}}=\frac{1}{\left(1-\alpha\right)^{\beta/r}}$$

(cf. [PW09] for an initial statement in this direction).

The Conditional Tail Expectation allows the following improvement.

Theorem 5. *For the Conditional Tail Expectation,*

$$\mathrm{CTE}_{\alpha;K}\left(L\right)-\mathrm{CTE}_\alpha\left(L\right)\le\frac{K^r}{1-\alpha}\cdot H_r\left(\max\left\{L,\,q_\alpha\right\}\right),$$

where $\max\left\{L,\,q_\alpha\right\}$ is the function $x\mapsto\max\left\{L\left(x\right),\,q_\alpha\right\}$ and $q_\alpha=F_L^{-1}\left(\alpha\right)$ is the $\alpha-$quantile under P.

Proof. Let $\tilde{P}$ be chosen such that $\mathsf{d}_r\left(P,\tilde{P}\right)\le K$. Recall (2) for the Conditional Tail Expectation, thus

$$\begin{aligned}\mathrm{CTE}_{\alpha;\tilde{P}}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)&=\inf_{\tilde{q}}\,\tilde{q}+\frac{1}{1-\alpha}\mathbb{E}_{\tilde{P}}\left(L-\tilde{q}\right)_+-\inf_q\,q+\frac{1}{1-\alpha}\mathbb{E}_P\left(L-q\right)_+\\&=\inf_{\tilde{q}}\,\tilde{q}+\frac{1}{1-\alpha}\mathbb{E}_{\tilde{P}}\left(L-\tilde{q}\right)_+-q_\alpha-\frac{1}{1-\alpha}\mathbb{E}_P\left(L-q_\alpha\right)_+\end{aligned}$$

where we have used that the infimum in (2) is attained at the quantile $q_\alpha\left(L\right)$. Hence

$$\begin{aligned}\mathrm{CTE}_{\alpha;\tilde{P}}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)&\le q_\alpha+\frac{1}{1-\alpha}\mathbb{E}_P\left(L-q_\alpha\right)_+-q_\alpha-\frac{1}{1-\alpha}\mathbb{E}_{\tilde{P}}\left(L-q_\alpha\right)_+\\&=\frac{1}{1-\alpha}\left(\mathbb{E}_P\left(L-q_\alpha\right)_+-\mathbb{E}_{\tilde{P}}\left(L-q_\alpha\right)_+\right).\end{aligned}$$

By Hölder continuity

$$\left(L\left(y\right)-q_\alpha\right)_+-\left(L\left(x\right)-q_\alpha\right)_+\le H_r\left(\left(L-q_\alpha\right)_+\right)\cdot d\left(x,y\right)^r$$

and in view of the duality relation (cf. (20)) it follows that

$$\mathbb{E}_{\tilde{P}}\left(L-q_\alpha\right)_+-\mathbb{E}_P\left(L-q_\alpha\right)_+\le H_r\left(\left(L-q_\alpha\right)_+\right)\cdot\mathsf{d}_r\left(P,\tilde{P}\right)^r,$$

such that

$$\mathrm{CTE}_{\alpha;\tilde{P}}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)\le\frac{1}{1-\alpha}H_r\left(\left(L-q_\alpha\right)_+\right)\cdot\mathsf{d}_r\left(P,\tilde{P}\right)^r.$$

Finally note that $H_r\left(\left(L-q_\alpha\right)_+\right)=H_r\left(\max\left\{L,q_\alpha\right\}\right)$, which completes the proof. □

The next theorem finally elaborates that the bound in Theorem 5 is the best possible bound, at least for discrete measures and for small perturbations K.

Theorem 6. *For a discrete probability measure $P=\sum p_x\delta_x$ with $p_x>0$ an a discrete set X there is $\tilde{K}>0$ such that*

$$\mathrm{CTE}_{\alpha;K}\left(L\right)=\mathrm{CTE}_{\alpha;P}\left(L\right)+\frac{K^r}{1-\alpha}H_r\left(\max\left\{L,\,q_\alpha\right\}\right)$$

for every $K\leq\tilde{K}$.

Proof. Choose $(x_0,x_1)\in\operatorname{argmax}_{x_0\neq x_1}\frac{\max\{L(x_0),q_\alpha\}-\max\{L(x_1),q_\alpha\}}{d(x_0,x_1)^r}$ and observe that

$$L\left(x_0\right)-L\left(x_1\right)\geq d\left(x_0,x_1\right)^r\cdot H_r\left(\max\left\{L,\,q_\alpha\right\}\right).$$

Define the measures

$$\tilde{P}\left(y\right):=\begin{cases}P\left(x_0\right)+\Delta & \text{if } y=x_0,\\ P\left(x_1\right)-\Delta & \text{if } y=x_1,\\ P\left(y\right) & \text{else}\end{cases}$$

$(\Delta\leq P\left(x_1\right))$ and

$$\pi_{x,y}:=\begin{cases}\Delta & \text{if } x=x_1 \text{ and } y=x_0,\\ P\left(x_1\right)-\Delta & \text{if } x=x_1 \text{ and } y=x_1,\\ P\left(x\right) & \text{if } x_0\neq x=y,\\ 0 & \text{else}\end{cases}$$

such that $P\left(x\right)=\sum_y\pi_{x,y}$ and $\tilde{P}\left(y\right)=\sum_x\pi_{x,y}$. Moreover

$$\begin{aligned}\mathsf{d}_r\left(P,\tilde{P}\right)^r &\leq \sum_{x,y}\pi_{x,y}d\left(x,y\right)^r=\Delta\cdot d\left(x_1,x_0\right)^r \\ &\leq P\left(x_1\right)\cdot d\left(x_1,x_0\right)^r\end{aligned}\tag{24}$$

by the choice of Δ.

Next, let the dual variable Z ($\mathbb{E}Z=1$, $0\leq Z\leq\frac{1}{1-\alpha}$) be chosen such that $\mathrm{CTE}_\alpha\left(L\right)=\mathbb{E}\,LZ$. Without loss of generality one may assume $P\left(Z=\frac{1}{1-\alpha}\right)=\alpha$, as otherwise the atom $\left\{0<Z<\frac{1}{1-\alpha}\right\}$ can be split into two atoms with the desired property.

As $L\left(x_0\right)\geq L\left(x_1\right)\geq q_\alpha$ it follows hence that $Z\left(x_0\right)=Z\left(x_1\right)=\frac{1}{1-\alpha}$, and consequently $\tilde{\mathbb{E}}Z=1$. Next,

$$\begin{aligned}\mathrm{CTE}_{\alpha;\tilde{P}}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)=\mathbb{E}_{\tilde{P}}LZ-\mathbb{E}_PLZ&=\frac{\Delta}{1-\alpha}\left(L\left(x_0\right)-L\left(x_1\right)\right)\\ &=\frac{H_r\left(\max\left\{L,\,q_\alpha\right\}\right)}{1-\alpha}\cdot\Delta\cdot d\left(x_0,x_1\right)^r.\end{aligned}$$

In combination with (24) it follows that

$$\mathrm{CTE}_{\alpha;\tilde{P}}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)\geq\frac{H_r\left(\max\left\{L,\,q_\alpha\right\}\right)}{1-\alpha}\cdot\mathsf{d}_r\left(P,\tilde{P}\right)^r,$$

that is

$$\mathrm{CTE}_{\alpha;K}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)\geq\frac{H_r\left(\max\left\{L,\,q_\alpha\right\}\right)}{1-\alpha}\cdot K^r,$$

provided that $K\leq P\left(x_1\right)\cdot d\left(x_1,x_0\right)^r=:\tilde{K}$. By Theorem 5 hence

$$\mathrm{CTE}_{\alpha;K}\left(L\right)-\mathrm{CTE}_{\alpha;P}\left(L\right)=\frac{H_r\left(\max\left\{L,\,q_\alpha\right\}\right)}{1-\alpha}\cdot K^r.$$

□

As an immediate application we mention the particular case $r = 1$, as it can be used to give immediate upper bounds for the price of insurance, allowing distortions in the Wasserstein distance up to a value of K.

The ingredients of the following corollary are often easily available for the actuary and provide a quick bound for the lump sum premium.

Corollary 2. *It holds that*

$$\sup_{\mathsf{d}_1\left(P,\tilde{P}\right)\leq K} \mathrm{CTE}_{\alpha;K}\left(L\right) \leq \mathrm{CTE}_{\alpha;P}\left(L\right) + \frac{K}{1-\alpha} H_1\left(\max\left\{L,\, q_\alpha\right\}\right),$$

with equality for K in a neighborhood of 0 and $H_1\left(\max\left\{L,\, q_\alpha\right\}\right)$ denoting the Lipschitz constant. In particular we have

$$\sup_{\mathsf{d}_1\left(P,\tilde{P}\right)\leq K} \tilde{\mathbb{E}}\, L \leq \mathbb{E}\, L + K \cdot H_1\left(\max\left\{L,\, q_\alpha\right\}\right)$$

for the price according the equivalence principle.

7. Numerical analysis and treatment

An explicit formula for the CTE_α is available just for selected types of insurance contracts, and a more general procedure to compute CTE_α is provided by (15). The CTE can be computed equally well by the linear program (LP)

$$\begin{array}{rl} \underset{(\text{in } Z)}{\text{maximize}} & \sum_k p_k L_k Z_k \\ \text{subject to} & \sum_k p_k Z_k = 1, \\ & 0 \leq Z_k,\, Z_k \leq \frac{1}{1-\alpha}, \end{array}$$

where we have employed the measures

$$P := \sum_k {}_k p_x\, q_{x+k} \cdot \delta_k = \sum_k p_k \cdot \delta_k \qquad \left(p_k = P\left(K = k\right) = {}_k p_x\, q_{x+k}\right)$$

as above.

To illustrate the results we have computed the Conditional Tail Expectation for important types of insurance in line with 3.1 and 3.2, they are depicted in Figure 1.

Equally well can the risk measure ρ_h be naturally evaluated by

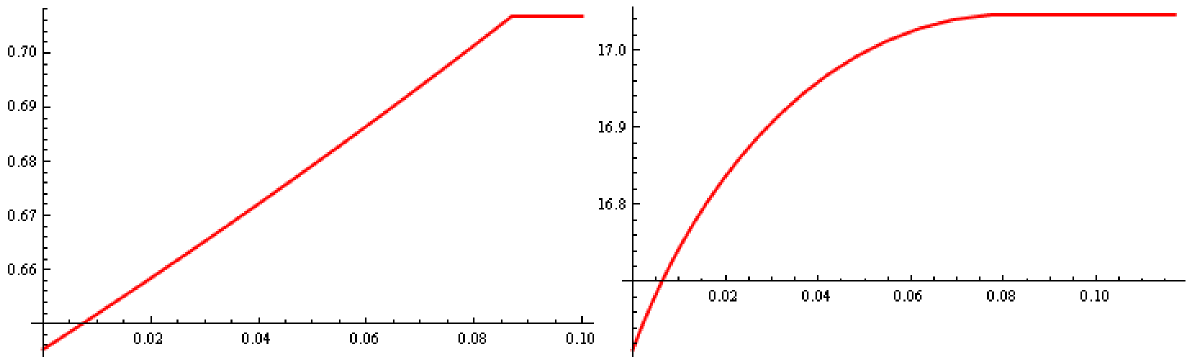
$$\rho_h\left(L\right) = \int F_L^{-1}\left(\alpha\right) h\left(\alpha\right) \mathrm{d}\alpha = \mathbb{E}\, LZ = \sum_k p_k L_k Z_k,$$

where the optimal dual variable Z is provided by

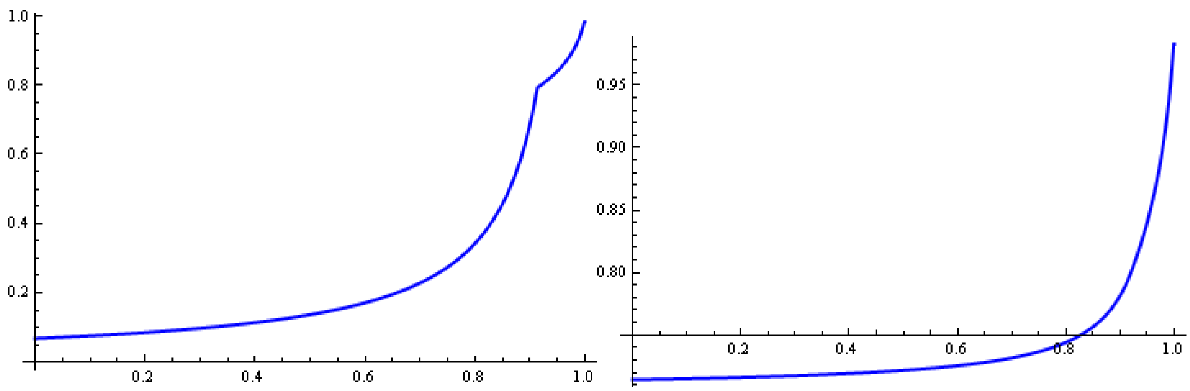
$$Z_{(k)} := \frac{1}{p_{(k)}} \int_{1-p_{(1)}-\dots p_{(k)}}^{1-p_{(1)}-\dots p_{(k-1)}} h\left(p\right) \mathrm{d}p.$$

It is significantly more difficult to evaluate the $K-$ambiguity risk measure – even for $\mathrm{CTE}_{\alpha;K}\left(L\right)$, the $K-$ambiguity Conditional Tail Expectation. This problem can be stated as

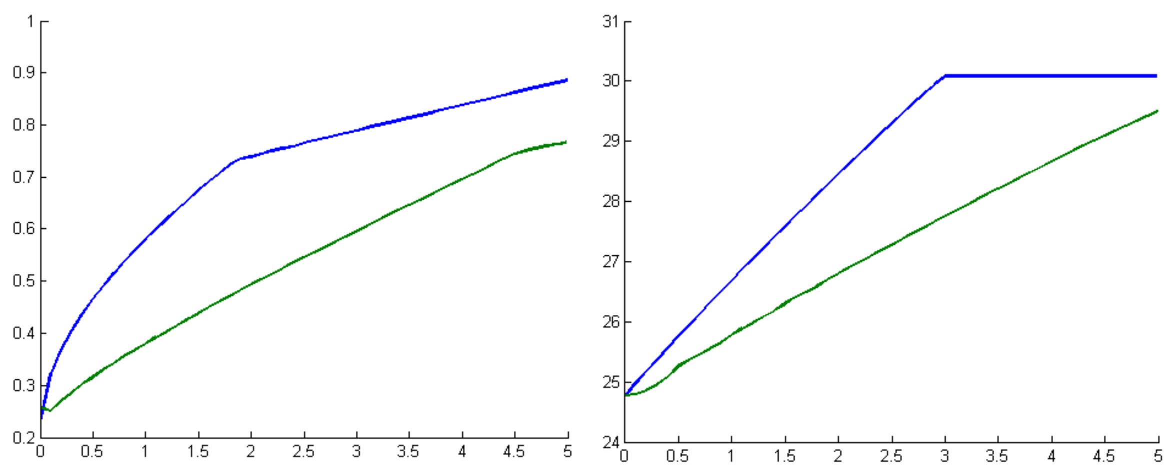
$$\begin{array}{rll} \underset{(\text{in } \pi \text{ and } Z)}{\text{maximize}} & \sum_{i,j} \pi_{i,j} L_j Z_j & \tilde{\mathbb{E}}\, ZL \\ \text{subject to} & \sum_j \pi_{i,j} = p_i, & \pi\left(A \times X\right) = P\left(A\right) \\ & \pi_{i,j} \geq 0, & \pi \text{ a probability measure} \\ & \sum_{i,j} \pi_{i,j} d_{i,j}^r \leq K^r & \mathsf{d}_r\left(P, \tilde{P}\right) \leq K \\ & 0 \leq Z_j \leq \frac{1}{1-\alpha} & 0 \leq Z \leq \frac{1}{1-\alpha} \\ & \sum_{i,j} \pi_{i,j} Z_j = 1 & \tilde{\mathbb{E}} Z = 1, \end{array} \tag{25}$$



(a) The CTE_α (small values of α) for the pure endowment ${}_{20}E_{40} = 64.6\,\%$ (left) and the annuity $\ddot{a}_{40:\overline{20}|} = 16.6\,\%$ (right).



(b) The CTE_α for the life insurance ${}_{20}A_{40} = 6.8\,\%$ (left) and the endowment $A_{40:\overline{20}|} = 71.4\,\%$ (right).



(c) The Conditional Tail Expectation at level $\alpha = 0.3$, $\mathrm{CTE}_{0.3}$, for the life insurance ${}_{20}A_{40}$ (left) and the annuity $\ddot{a}_{60}$ (right; notice, that $\ddot{a}_{\overline{40}|} \approx 29$). Displayed is the concave map $K \mapsto \mathrm{CTE}_{0.3;K}(L)$ for values of K up to 5 for the respective loss functions. The order for the Wasserstein distance is $r = 1$ and $r = 2$.

Figure 1: Conditional Tail Expectation CTE_α. The computations are made for the German life table 2008T M/F (for which ${}_{20}q_{40} = 8.6\,\%$), and an interest rate of $1.75\,\%$.

where we have indicated the conditions on the right again, subject to the new measure $\tilde{P} = \sum_j \tilde{p}_j \delta_j$ ($\tilde{p}_j = \sum_i \pi_{i,j}$), which is at a distance of $\mathsf{d}_r\left(P, \tilde{P}\right) \le K$. It should be noted that (25) is not a linear program, as it involves multiplications as $\pi_{i,j} \cdot Z_j$ in its constraints as well as in the objective (the problem is *bi*linear). This fact significantly complicates the computations.

To escape this difficulty and to obtain reasonable computing times we propose to handle the problem in the same way, as bilinear problems are usually treated, that is by iteratively improving (25): In an iteration step fix Z in (25) and compute π, next keep π fixed in (25) and compute Z (or – even quicker – compute Z by employing (14)). Any of these problems then is a linear program, which improves the objective. The described procedure can be proved to converge, provided that $L \ge 0$, what is the case for a typical payoff in insurance.

8. Summary and outlook

In this paper we address the problem of pricing insurance contracts under ambiguity. An ambiguous situation is an every day situation for an actuary, as the distribution of a loss function is not known precisely in a typical situation.

In a first part of the paper we address general risk measures, in particular the Conditional Tail Expectation as a pricing strategy. These risk measures incorporate designed loadings for the premium of an insurance contract. To illustrate the impact of the design parameter for the CTE the computations are made explicit for usual types of life insurance contracts, as, surprisingly, simple explicit expressions are available in these situations by changing the life table.

The second part of the paper is dedicated to evaluate these risk measures in ambiguous situations. Closed forms are not available here, but sharp bounds allow to precisely estimate the results. These bounds are designed in a way to serve as a quick guideline for an actuary when pricing a contract in a situation, where the distribution of the loss function is known only at an insufficient level. The concept to have a notion of distance on probability measures available is the Wasserstein distance.

Finally some relevant examples are used to demonstrate the impact of these pricing methodologies in concrete situations.

Ambiguity was studied in a different context – in finance – recently, it is an essential ingredient to prove key properties of the $1/N$ investment strategy [PPW12]. These fundamental results demonstrate the particular importance of considerations on ambiguity and justify further research in the direction, they reveal very natural, useful and capable estimates.

9. Acknowledgment

We wish to express our gratitude to Prof. Pflug for his continual advice and thank the referees for their constructive criticism.

The plots have been produced by use of Mathematica and Matlab.

10. Bibliography

[ADEH99] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent Measures of Risk. *Mathematical Finance*, 9:203–228, 1999. 1, 2

[AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Birkhäuser Verlag AG, Basel, Switzerland, 2nd edition, 2005. 10

[Den89] D. Denneberg. Distorted probabilities and insurance premiums. *Proceedings of the 14th SOR*, 1989. 1

[Ger97] Hans U. Gerber. *Life Insurance Mathematics.* Springer-Verlag Berlin and Heidelberg GmbH&Co. K, 3rd edition, 1997. 4

[GLS03] Hans U. Gerber, Bartholomew P. K. Leung, and Elias S. W. Shiu. Indicator function and Hattendorff theorem. *North American Actuarial Journal*, 7(1):38–47, 2003. 1

[Hat68] K. Hattendorff. Das Risico bei der Lebensversicherung. *Maisius' Rundschau der Versicherungen*, 18:169–183, 1868. German. 1

[KRS09] Bangwon Ko, Ralph P. Russo, and Nariankadu D. Shyamalkumar. A note on the nonparametric estimation of the CTE. *Astin Bulletin*, 39(2):717–734, 2009. 2

[Kus01] Shigeo Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001. 7

[LDSSL12] Miguel López-Díaz, Miguel A. Sordo, and Alfonso Suárez-Llorens. On the L_p-metric between a probability distribution and its distortion. *Insurance: Mathematics and Economics*, 51:257–264, 2012. 8

[Pfl06] Georg Ch. Pflug. On distortion functionals. *Statistics and Decisions*, 24:45–60, 2006. 7

[PP11] Georg Ch. Pflug and Alois Pichler. *Approximations for Probability Distributions and Stochastic Optimization Problems*, volume 163 of *International Series in Operations Research & Management Science*, chapter 15, pages 343–387. Springer New York, 2011. 8

[PP12a] Georg Ch. Pflug and Alois Pichler. Decomposion of risk measures. optimization online, 2012. 7

[PP12b] Georg Ch. Pflug and Alois Pichler. On dynamic decomposition of multistage stochastic programs. Optimization Online preprint, 2012. 7

[PPW12] Georg Ch. Pflug, Alois Pichler, and David Wozabal. The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36:410–417, 2012. 17

[PR07] Georg Ch. Pflug and Werner Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, River Edge, NJ, 2007. 3, 8

[PW09] Georg Ch. Pflug and David Wozabal. *Ambiguity in Portfolio Selection*, chapter 18, pages 337–391. Financial Mathematics Series. Chapmann and Hall/ CRC, London, 2009. 13

[Rac91] Svetlozar T. Rachev. *Probability metrics and the stability of stochastic models*. John Wiley and Sons Ltd., West Sussex PO19, 1UD, England, 1991. 8

[RSF11] Svetlozar T. Rachev, Stoyan V. Stoyanov, and Frank J. Fabozzi. *A Probability Metrics Approach to Financial Risk Measures*. Wiley, 2011. 8

[RU00] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–41, 2000. 3

[SDR09] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming*. MPS-SIAM Series on Optimization 9, 2009. 8

[Sha12] Alexander Shapiro. Time consistency of dynamic risk measures. Optimization Online preprint, 2012. 7, 8

[Ver06] Anatoly M. Vershik. Kantorovich metric: Initial history and little-known applications. *Journal of Mathematical Sciences*, 133(4):1410–1417, 2006. 8

[Vil03] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003. 8, 10

[Wan00] Shaun S. Wang. A class of distortion operations for financial and insurance risk. *The Journal of Risk and Insurance*, 67(1):15–36, 2000. 1