Bäuerle, Nicole; Glauner, Alexander

Minimizing spectral risk measures applied to Markov decision processes

Mathematical Methods of Operations Research

Provided in Cooperation with:
Springer Nature

Suggested Citation: Bäuerle, Nicole; Glauner, Alexander (2021) : Minimizing spectral risk measures applied to Markov decision processes, Mathematical Methods of Operations Research, ISSN 1432-5217, Springer, Berlin, Heidelberg, Vol. 94, Iss. 1, pp. 35-69, https://doi.org/10.1007/s00186-021-00746-w

Mathematical Methods of Operations Research (2021) 94:35–69
https://doi.org/10.1007/s00186-021-00746-w

ORIGINAL ARTICLE

Minimizing spectral risk measures applied to Markov decision processes

Nicole Bäuerle[1] · **Alexander Glauner**[1]

Received: 8 December 2020 / Revised: 9 May 2021 / Accepted: 25 June 2021 /
Published online: 27 July 2021

Abstract
We study the minimization of a spectral risk measure of the total discounted cost generated by a Markov Decision Process (MDP) over a finite or infinite planning horizon. The MDP is assumed to have Borel state and action spaces and the cost function may be unbounded above. The optimization problem is split into two minimization problems using an infimum representation for spectral risk measures. We show that the inner minimization problem can be solved as an ordinary MDP on an extended state space and give sufficient conditions under which an optimal policy exists. Regarding the infinite dimensional outer minimization problem, we prove the existence of a solution and derive an algorithm for its numerical approximation. Our results include the findings in Bäuerle and Ott (Math Methods Oper Res 74(3):361–379, 2011) in the special case that the risk measure is Expected Shortfall. As an application, we present a dynamic extension of the classical static optimal reinsurance problem, where an insurance company minimizes its cost of capital.

Keywords Risk-sensitive Markov decision process · Spectral risk measure · Dynamic reinsurance

Mathematics Subject Classification 90C40 · 91G70 · 91G05

1 Introduction

In the last decade, there have been various proposals to replace the expectation in the optimization of Markov Decision Processes (MDPs) by risk measures. The idea

✉ Nicole Bäuerle
nicole.baeuerle@kit.edu

Alexander Glauner
alexander.glauner@kit.edu

[1] Department of Mathematics, Karlsruhe Institute of Technology (KIT), 76128 Karlsruhe, Germany

behind it is to take the risk-sensitivity of the decision maker into account. Using simply the expectation models a risk-neutral decision maker whose optimal policy sometimes can be very risky, for an example see e.g. Bäuerle and Ott (2011).

The literature can here be divided into two streams: Those papers which apply risk measures recursively and those which apply the risk measure to the total cost. The recursive approach for general MDP can for example be found in Ruszczyński (2010); Chu and Zhang (2014); Bäuerle and Glauner (2021). The theory for these kind of models is rather different to the ones where the risk measures is applied to the total cost, since in the recursive approach we still get a recursive solution procedure directly. In this paper, we contribute to the second model class, i.e. we assume that a cost process is generated over discrete time by a decision maker and she aims at minimizing the risk measure applied to either the cost over a finite time horizon or over an infinite time horizon. The class of risk measures we consider here are so-called *spectral risk measures* which form a class of coherent risk measures including the *Expected Shortfall* or *Conditional Value-at-Risk*. More precisely spectral risk measures are mixtures of Expected Shortfall at different levels.

For Expected Shortfall, the problem has already been treated e.g. in Bäuerle and Ott (2011), Chow et al. (2015), and Uğurlu (2017). Whereas in Chow et al. (2015) the authors use a decomposition result of the Expected Shortfall shown in Pflug and Pichler (2016), the authors of Bäuerle and Ott (2011) use the representation of Expected Shortfall as the solution of a global optimization problem over a real valued parameter, see Rockafellar and Uryasev (2000). Interchanging the resulting two infima from the optimization problems yields a two-step method to solve the decision problem. Using the recent representation of spectral risk measures as an optimization problem over functions involving the convex conjugate in Pichler (2015), we follow a similar approach here. The problem can again be decomposed into an inner and outer optimization problem. The inner problem is to minimize the expected convex function of the total cost. It can be solved with MDP techniques after a suitable extension of the original state space. Note that already here we get some difference to the Expected Shortfall problem. In contrast to the findings in Bäuerle and Ott (2011) who assume bounded cost or Uğurlu (2017) who assumes L^1 cost, we only require the cost to be bounded from below. No further integrability assumption is necessary here. Moreover, we allow for general Borel state and action spaces and give continuity and compactness conditions under which an optimal policy exists. The major challenge is now the outer optimization problem, since we have to minimize over a function space and the dependence of the value function of the MDP on the functions is involved. However, we are again able to prove the existence of an optimal policy and an optimal function in the representation of the spectral risk measure. Moreover, by approximating the function space in the right way, we are able to reduce the outer optimization problem to a finite dimensional problem with a predetermined error bound. This yields an algorithm for the solution of the original optimization problem. Using an example from optimal reinsurance we show how our results can by applied.

Note that for Expected Shortfall the authors in Chow and Ghavamzadeh (2014) and Tamar et al. (2015) have developed gradient-based methods for the numerical computation of the optimal value and policy. For finite state and action spaces (Li et al. 2017) provide an algorithm for quantile minimization of MDPs which is a

similar problem. However, the outer optimization problem for spectral risk measures is much more demanding, since it is infinite dimensional.

The paper is organized as follows: In the next section, we summarize definitions and properties of risk measures and introduce in particular the class of spectral risk measures which we consider here. In Sect. 3, we introduce the Markov Decision Model and give continuity and compactness assumptions which will later guarantee the existence of optimal policies. At the end of this section, we formulate the spectral risk minimization problem of the total cost. We also give some interpretations and show relations to other problems. In Sect. 4, we summarize our findings in a nutshell. The necessary state space extension is explained as well as the recursive solution algorithm for the inner optimization problem. Moreover, the existence of optimal policies is stated. Then, we treat the outer optimization problem and state the existence of an optimal function in the representation of the spectral risk measure. Afterwards, we deal with the numerical treatment of this problem. We show here that the infinite dimensional optimization problem can be approximated by a finite dimensional one. In Sect. 5, we extend our results to decision models with infinite planning horizon. Besides, if the state space is the real line we show that the restrictive assumption of the continuity of the transition function which we need in the general model, can be replaced by semicontinuity if some further monotonicity assumptions are satisfied. In the final Sect. 6, we apply our findings to an optimal dynamic reinsurance problem. Problems of this type have been treated in a static setting before, see e.g. Chi and Tan (2013), Cui et al. (2013), Lo (2017) and Bäuerle and Glauner (2018), but we consider them in a dynamic framework for the first time. The aim is to minimize the solvency capital calculated with a spectral risk measure by actively choosing reinsurance contracts for the next period. When the premium for the reinsurance contract is calculated by the expected premium principle, we show that the optimal reinsurance contracts are of stop loss type. All proofs and detailed derivations of our results are deferred to the appendix.

2 Spectral risk measures

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $L^0 = L^0(\Omega, \mathcal{A}, \mathbb{P})$ the vector space of real-valued random variables thereon. By L^1 we denote the subspace of integrable random variables and by $L^0_{\geq 0}$ the subspace which consists of non-negative random variables. We follow the convention of the actuarial literature that positive realizations of random variables represent losses and negative ones gains. Let $\mathcal{X} \subseteq L^0$ be a convex cone. A *risk measure* is a functional $\rho : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$. The following properties are relevant in this paper.

Definition 2.1 A risk measure $\rho : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is called

a) *law-invariant* if $\rho(X) = \rho(Y)$ for X, Y having the same distribution.
b) *monotone* if $X \leq Y$ implies $\rho(X) \leq \rho(Y)$.
c) *translation invariant* if $\rho(X + m) = \rho(X) + m$ for all $m \in \mathbb{R} \cap \mathcal{X}$.
d) *positive homogeneous* if $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda \in \mathbb{R}_+$.
e) *comonotonic additive* if $\rho(X + Y) = \rho(X) + \rho(Y)$ for all comonotonic X, Y.

f) *subadditive* if $\rho(X+Y) \leq \rho(X) + \rho(Y)$ for all X, Y.

A risk measure is referred to as *monetary* if it is monotone and translation invariant. It appears to be consensus in the literature that these two properties are a necessary minimal requirement for any risk measure. Monetary risk measures which are additionally positive homogeneous and subadditive are called *coherent*. Here, $F_X(x) = \mathbb{P}(X \leq x),\ x \in \mathbb{R}$, denotes the distribution function and $F_X^{-1}(u) = \inf\{x \in \mathbb{R} : F_X(x) \geq u\},\ u \in [0,1]$, the quantile function of a random variable X. We will focus on the following class of risk measures.

Definition 2.2 An increasing function $\phi : [0,1] \to \mathbb{R}_+$ with $\int_0^1 \phi(u)\mathrm{d}u = 1$ is called *spectrum* and the functional $\rho_\phi : L^0_{\geq 0} \to \mathbb{R} \cup \{\infty\}$ with

$$\rho_\phi(X) = \int_0^1 F_X^{-1}(u)\phi(u)\mathrm{d}u$$

is referred to as *spectral risk measure*.

Spectral risk measures were introduced by Acerbi (2002). They have all the properties listed in Definition 2.1. Properties a)–e) follow directly from respective properties of the quantile function. Verifying subadditivity is more involved, see Dhaene et al. (2000). As part of the proof they showed that spectral risk measures preserve the increasing convex oder. Spectral risk measures belong to the larger class of distortion risk measures.

Definition 2.3 An increasing right-continuous function $\varphi : [0,1] \to [0,1]$ with $\varphi(0) = 0$ and $\varphi(1) = 1$ is called *distortion function* and the functional $\rho_\varphi : L^0_{\geq 0} \to \mathbb{R} \cup \{\infty\}$ with

$$\rho_\varphi(X) = \int_0^1 F_X^{-1}(u)\mathrm{d}\varphi(u)$$

is referred to as *distortion risk measure*.

In the special case of a spectral risk measure, the distortion function is given by

$$\varphi(u) = \int_0^u \phi(s)\mathrm{d}s, \qquad u \in [0,1] \tag{2.1}$$

and is convex. This also shows that it is no restriction to assume ϕ being right continuous (as the right derivative of a convex function). Conversely, for a convex distortion function without a jump in 1, which implies continuity on $[0,1]$, one can always find a representation as in (2.1) with ϕ being a spectrum. Consequently, all distortion risk measures with convex and continuous distortion function are spectral. It has been proven by Dhaene et al. (2000) that the convexity of φ is equivalent to ρ_φ being subadditive.

Note that ρ_ϕ is finite on $L^1_{\geq 0}$ if the spectrum ϕ is bounded. On $L^0_{\geq 0}$ the value $+\infty$ is possible. Shapiro (2013) has shown that a finite risk measure on $L^1_{\geq 0}$ with all the properties in Definition 2.1 is already spectral with bounded spectrum.

Example 2.4 The most widely used spectral risk measure is *Expected Shortfall*

$$\mathrm{ES}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(u)\mathrm{d}u, \qquad \alpha \in [0,1).$$

Its spectrum $\phi(u) = \frac{1}{1-\alpha}\mathbb{1}_{[\alpha,1]}(u)$ is bounded. Especially in optimization, an infimum representation of Expected Shortfall going back to Rockafellar and Uryasev (2000) is very useful:

$$\mathrm{ES}_\alpha(X) = \inf_{q\in\mathbb{R}}\left\{q + \frac{1}{1-\alpha}\mathbb{E}[(X-q)^+]\right\}, \qquad X \in L^0_{\geq 0}. \tag{2.2}$$

The infimum is attained at $q = F_X^{-1}(\alpha)$.

Henceforth, we assume w.l.o.g. that ϕ is right-continuous. Then $\nu([0,t]) := \phi(t)$ defines a Borel measure on $[0,1]$. Let us define a further measure μ by $\frac{d\mu}{d\nu}(\alpha) := (1-\alpha)$. Every spectral risk measure can be expressed as a mixture of Expected Shortfall over different confidence levels, see e.g. Proposition 8.18 in McNeil et al. (2015).

Proposition 2.5 *Let ρ_ϕ be a spectral risk measure. Then μ is a probability measure on $[0,1]$ and ρ_ϕ has the representation*

$$\rho_\phi(X) = \int_0^1 ES_\alpha(X)\mu(\mathrm{d}\alpha), \qquad X \in L^0_{\geq 0}.$$

When we allow to take the supremum on the r.h.s. over all probability measures μ we would get the superclass of coherent risk measures, see Kusuoka (2001).

Using Proposition 2.5, the infimum representation (2.2) of Expected Shortfall can be generalized to spectral risk measures.

Proposition 2.6 *Let ρ_ϕ be a spectral risk measure with bounded spectrum. We denote by G the set of increasing convex functions $g : \mathbb{R} \to \mathbb{R}$. Then it holds for $X \in L^0_{\geq 0}$*

$$\rho_\phi(X) = \inf_{g\in G}\left\{\mathbb{E}[g(X)] + \int_0^1 g^*(\phi(u))\mathrm{d}u\right\},$$

where g^* is the convex conjugate of $g \in G$.

Proof For $X \in L^1_{\geq 0}$ the assertion has been proven by Pichler (2015). For non-integrable $X \in L^0_{\geq 0}$ it follows from Proposition 2.5

$$\rho_\phi(X) = \int_0^1 \mathrm{ES}_\alpha(X)\mu(\mathrm{d}\alpha) \geq \mathrm{ES}_0(X) = \mathbb{E}[X] = \infty.$$

Now let $g \in G$ and $U_X \sim \mathcal{U}(0,1)$ be the generalized distributional transform of X, i.e. $F_X^{-1}(U_X) = X$ a.s. By the definition of the convex conjugate it holds $g(X) + g^*(\phi(U_X)) \geq X\phi(U_X)$. Hence, we have

$$\mathbb{E}[g(X)] + \mathbb{E}[g^*(\phi(U_X))] \geq \mathbb{E}[X\,\phi(U_X)] = \mathbb{E}[F_X^{-1}(U_X)\,\phi(U_X)] = \rho_\phi(X) = \infty.$$

Since $g \in G$ was arbitrary, the assertion follows. □

Remark 2.7 The proof by Pichler (2015) shows that for $X \in L^1_{\geq 0}$ the infimum is attained in $g_{\phi,X} : \mathbb{R} \to \mathbb{R}$, $g_{\phi,X}(x) = \int_0^1 F_X^{-1}(\alpha) + \frac{1}{1-\alpha}\left(x - F_X^{-1}(\alpha)\right)^+ \mu(\mathrm{d}\alpha)$ with μ from Proposition 2.5 and that the derivative of this function is $g'_{\phi,X}(x) = \phi(F_X(x))$ a.e.

3 Markov decision model

We consider the following standard Markov Decision Process with general Borel state and action space. By Borel space we mean a Borel subset of a Polish space. The *state space* E is a Borel space with Borel σ-algebra $\mathcal{B}(E)$ and the *action space* A is a Borel space with Borel σ-Algebra $\mathcal{B}(A)$. The possible state-action combinations at time n form a measurable subset D_n of $E \times A$ such that D_n contains the graph of a measurable mapping $E \to A$. The x-section of D_n,

$$D_n(x) := \{a \in A : (x,a) \in D_n\},$$

is the set of admissible actions in state $x \in E$ at time n. Note that the sets $D_n(x)$ are non-empty. We assume that the dynamics of the MDP are given by measurable *transition functions* $T_n : D_n \times \mathcal{Z} \to E$ and depend on *disturbances* $Z_1, Z_2, \ldots$ which are independent random elements on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{Z}, \mathfrak{Z})$. When the current state is x_n, the controller chooses action $a_n \in D_n(x_n)$ and z_{n+1} is the realization of Z_{n+1}, then the next state is given by

$$x_{n+1} := T_n(x_n, a_n, z_{n+1}). \tag{3.1}$$

The *one-stage cost function* $c_n : D_n \times E \to \mathbb{R}_+$ gives the cost $c_n(x,a,x')$ for choosing action a if the system is in state x at time n and the next state is x'. The *terminal cost function* $c_N : E \to \mathbb{R}_+$ gives the cost $c_N(x)$ if the system terminates in state x. Note that instead of non-negative cost we can equivalently consider cost which are bounded from below.

The model data is supposed to have the following continuity and compactness properties.

Assumption 3.1 (i) The sets $D_n(x)$ are compact and $E \ni x \mapsto D_n(x)$ are upper semicontinuous, i.e. if $x_k \to x$ and $a_k \in D_n(x_k)$, $k \in \mathbb{N}$, then (a_k) has an accumulation point in $D_n(x)$.

(ii) The transition functions T_n are continuous in (x,a).

(iii) The one-stage cost functions c_n and the terminal cost function c_N are lower semi-continuous.

Under a finite planning horizon $N \in \mathbb{N}$, we consider the model data for $n = 0, \dots, N-1$. The decision model is called *stationary* if D, T, c do not depend on n and the disturbances are identically distributed. If the model is stationary and the terminal cost is zero, we allow for an *infinite time horizon* $N = \infty$.

For $n \in \mathbb{N}_0$ we denote by $\mathcal{H}_n$ the set of *feasible histories* h_n of the decision process up to time n where

$$h_n := \begin{cases} x_0, & \text{if } n = 0, \\ (x_0, a_0, x_1, \dots, x_n), & \text{if } n \geq 1, \end{cases}$$

with $a_k \in D_k(x_k)$ for $k \in \mathbb{N}_0$. In order for the controller's decisions to be implementable, they must be based on the information available at the time of decision making, i.e. be functions of the history of the decision process.

Definition 3.2 a) A measurable mapping $f_n : \mathcal{H}_n \to A$ with $f_n(h_n) \in D_n(x_n)$ for every $h_n \in \mathcal{H}_n$ is called *decision rule* at time n. A finite sequence $\sigma = (f_0, \dots, f_{N-1})$ is called *N-stage policy* and a sequence $\sigma = (f_0, f_1, \dots)$ is called *policy*.

b) A decision rule at time n is called *Markov* if it depends on the current state only, i.e. $f_n(h_n) = f_n(x_n)$ for all $h_n \in \mathcal{H}_n$. If all decision rules are Markov, the (N-stage) policy is called *Markov*.

c) An (N-stage) policy σ is called *stationary* if $\sigma = (f, \dots, f)$ or $\sigma = (f, f, \dots)$, respectively, for some Markov decision rule f.

With Π and Π^M we denote the sets of all policies and Markov policies, respectively. It will be clear from the context if N-stage or infinite stage policies are meant. An admissible policy always exists as D_n contains the graph of a measurable mapping.

Since risk measures are defined as real-valued mappings of random variables, we will work with a functional representation of the decision process. The law of motion does not need to be specified explicitly. We define for an initial state $x_0 \in E$ and a policy $\sigma \in \Pi$

$$X_0^\sigma = x_0, \qquad X_{n+1}^\sigma = T_n(X_n^\sigma, f_n(H_n^\sigma), Z_{n+1}).$$

Here, the process $(H_n^\sigma)_{n \in \mathbb{N}_0}$ denotes the history of the decision process viewed as a random element, i.e.

$$H_0^\sigma = x_0, \quad H_1^\sigma = \left(X_0^\sigma, f_0(X_0^\sigma), X_1^\sigma\right), \quad \dots, \quad H_n^\sigma = (H_{n-1}^\sigma, f_{n-1}(H_{n-1}^\sigma), X_n^\sigma).$$

Under a Markov policy the recourse on the random history of the decision process is not needed.

Even though the model is non-stationary we will explicitly introduce discounting by a factor $\beta > 0$ since for the following state space extension it is relevant if there is

discounting. Otherwise, stationary models with discounting would have to be treated separately. For a finite planning horizon $N \in \mathbb{N}$, the total discounted cost generated by a policy $\sigma \in \Pi$ if the initial state is $x \in E$, is given by

$$C_N^{\sigma x} := \sum_{k=0}^{N-1} \beta^k c_k(X_k^\sigma, f_k(H_k^\sigma), X_{k+1}^\sigma) + \beta^N c_N(X_N^\sigma).$$

If the model is stationary and the planning horizon infinite, the total discounted cost is given by

$$C_\infty^{\sigma x} := \sum_{k=0}^{\infty} \beta^k c(X_k^\sigma, f_k(H_k^\sigma), X_{k+1}^\sigma).$$

For a generic total cost regardless of the planning horizon we write $C^{\sigma x}$. Our aim is to find a policy $\sigma \in \Pi$ which attains

$$\inf_{\sigma \in \Pi} \rho_\phi(C_N^{\sigma x}) \tag{3.2}$$

or

$$\inf_{\sigma \in \Pi} \rho_\phi(C_\infty^{\sigma x}), \tag{3.3}$$

respectively, for a fixed spectral risk measure $\rho_\phi : L^0_{\geq 0} \to \mathbb{R} \cup \{\infty\}$ with $\phi(1) < \infty$, i.e. ϕ is bounded. We can apply Proposition 2.6 to reformulate the optimization problems (3.2) and (3.3) to

$$\begin{aligned} \inf_{\sigma \in \Pi} \rho_\phi\left(C^{\sigma x}\right) &= \inf_{\sigma \in \Pi} \inf_{g \in G} \left\{ \mathbb{E}[g(C^{\sigma x})] + \int_0^1 g^*(\phi(u)) \mathrm{d}u \right\} \\ &= \inf_{g \in G} \inf_{\sigma \in \Pi} \left\{ \mathbb{E}[g(C^{\sigma x})] + \int_0^1 g^*(\phi(u)) \mathrm{d}u \right\} \\ &= \inf_{g \in G} \left\{ \inf_{\sigma \in \Pi} \mathbb{E}[g(C^{\sigma x})] + \int_0^1 g^*(\phi(u)) \mathrm{d}u \right\}. \end{aligned} \tag{3.4}$$

For fixed $g \in G$ we will refer to

$$\inf_{\sigma \in \Pi} \mathbb{E}[g(C^{\sigma x})] \tag{3.5}$$

as *inner optimization problem*. In the following section we solve (3.5) as an ordinary MDP on an extended state space. If $C^{\sigma x} \in L^0_{\geq 0}$ but not in L^1, then $\rho_\phi(C^{\sigma x}) = \infty$. These policies are not interesting and can be excluded from the optimization.

Since an increasing convex function $g : \mathbb{R} \to \mathbb{R}$ can be viewed as a disutility function, optimality criterion (3.5) implies that the expected disutility of the total discounted cost in minimized. If g is strictly increasing, the optimization problem is

not changed by applying g^{-1}, i.e. minimizing the corresponding certainty equivalent $g^{-1}\big(\mathbb{E}[g(C^{\sigma x})]\big)$. For bounded one-stage cost functions such problems are solved in Bäuerle and Rieder (2014). The special case of the exponential disutility function $g(x) = \exp(\gamma x),\ \gamma > 0$, has been studied first by Howard and Matheson (1972) in a decision model with finite state and action space. The term *risk-sensitive MDP* goes back to them. The certainty equivalent corresponding to an exponential disutility is the entropic risk measure

$$\rho(X) = \frac{1}{\gamma} \log \mathbb{E}\left[e^{\gamma X}\right].$$

It has been shown by Müller (2007) that an exponential disutility is the only case where the certainty equivalent defines a monetary risk measure apart from expectation itself (linear disutility).

The concepts of spectral risk measures and expected disutilities (or corresponding certainty equivalents) can be combined to so-called *rank-dependent expected disutilities* of the form $\rho_\phi(u(X))$, where u is a disutility function. The corresponding certainty equivalent is $u^{-1}\big(\rho_\phi(u(X))\big)$. In fact, this concept works more generally for distortion risk measures and incorporates both expected disutilities (identity as distortion function) and distortion risk measures (identity as disutility function). The idea is that the expected disutility is calculated w.r.t. a distorted probability instead of the original probability measure. As long as the distorted probability is spectral, using a rank dependent disutility instead of ρ_ϕ leads to structurally the same inner problem as (3.5), only g is replaced by $g(u(\cdot))$. Our results apply here, too. The certainty equivalent of a rank-dependent expected disutility combining an exponential disutility with a spectral risk measure is itself a convex (but not coherent) risk measure. It has been introduced by Tsanakas and Desli (2003) as *distortion-exponential risk measure*.

4 Main results: finite planning horizon

4.1 Inner problem

Under a finite planning horizon $N \in \mathbb{N}$, we consider the non-stationary version of the decision model and our first aim is to solve

$$\inf_{\sigma \in \Pi} \mathbb{E}[g(C_N^{\sigma x})] \tag{4.1}$$

for an arbitrary but fixed increasing convex function $g \in G$. We assume that for all $x \in E$ there is at least one policy σ s.t. $C_N^{\sigma x} \in L^1$. Problem (4.1) is well-defined since the target function is bounded from below by $g(0)$. W.l.o.g. we assume $g \geq 0$. Note that the value $+\infty$ is possible.

As the functions $g \in G$ are in general non-linear, the optimization problem cannot be solved directly with dynamic programming techniques. This can be overcome by embedding the problem into an extended MDP following Bäuerle and Rieder (2014).

The state space of this extended MDP is

$$\mathbf{E} := E \times \mathbb{R}_+ \times (0, \infty)$$

with corresponding Borel σ-algebra. A generic element of $\mathbf{E}$ is denoted by (x, s, t). The idea is that s summarizes the cost accumulated to far and that t keeps track of the discounting. The action space A and the admissible state-action combinations D_n, $n = 0, \ldots, N-1$, remain unchanged. Formally, one defines

$$\mathbf{D}_n := \{(x, s, t, a) \in \mathbf{E} \times A : \ a \in D_n(x)\}, \qquad n = 0, \ldots, N-1$$

implying $\mathbf{D}_n(x, s, t) = D_n(x), \ (x, s, t) \in \mathbf{E}$. The transition function on the new state space is given by $\mathbf{T}_n : \mathbf{D}_n \times \mathcal{Z} \to \mathbf{E}$,

$$\mathbf{T}_n(x, s, t, a, z) := \begin{pmatrix} T_n(x, a, z) \\ s + t c_n(x, a, T_n(x, a, z)) \\ \beta t \end{pmatrix}, \qquad n = 0, \ldots, N-1.$$

Feasible histories of the decision model with extended state space up to time n have the form

$$\mathbf{h}_n := \begin{cases} (x_0, s_0, t_0), & n = 0, \\ (x_0, s_0, t_0, a_0, x_1, s_1, t_1, a_1, \ldots, x_n, s_n, t_n), & n \geq 1, \end{cases}$$

where $a_k \in D_k(x_k)$, $k = 0, \ldots, N-1$, and the set of such histories is denoted by $\mathcal{H}_n$. In particular, we have the same recursion (3.1) for the state process and when we start with $s_0 = 0, t_0 = 1$ we obtain:

$$s_n = \sum_{k=0}^{n-1} \beta^k c_k(x_k, a_k, x_{k+1}) \quad \text{and} \quad t_n = \beta^n, \qquad n = 1, \ldots, N. \tag{4.2}$$

By $\mathbf{\Pi}$ we denote the set of all history-dependent policies for the decision model with extended state space. Policies are denoted by $\pi = (d_0, d_1, \ldots, d_{N-1})$ with measurable decision rules $d_n : \mathcal{H}_n \to A$ satisfying $d_n(\mathbf{h}_n) \in D_n(x_n)$. By $\mathbf{\Pi}^M$ we denote the set of all Markov policies where decision rules are given by $d_n : \mathbf{E} \to A$ with $d_n(x_n, s_n, t_n) \in D_n(x_n)$. For $\pi = (d_0, \ldots, d_{N-1}) \in \mathbf{\Pi}$ the process $(\mathbf{H}_n^\pi)$ denotes the history of the extended MDP viewed as a random element, i.e.

$$\mathbf{H}_0^\pi = (X_0^\pi, \mathbf{s}_0^\pi, \mathbf{t}_0^\pi) = (x_0, s_0, t_0), \qquad \mathbf{H}_n^\pi = \big(\mathbf{H}_{n-1}^\pi, d_{n-1}(\mathbf{H}_{n-1}^\pi), X_n^\pi, \mathbf{s}_n^\pi, \mathbf{t}_n^\pi\big),$$

where

$$(X_n^\pi, \mathbf{s}_n^\pi, \mathbf{t}_n^\pi) = \mathbf{T}_{n-1}\big(X_{n-1}^\pi, \mathbf{s}_{n-1}^\pi, \mathbf{t}_{n-1}^\pi, d_{n-1}(\mathbf{H}_{n-1}^\pi), Z_n\big).$$

We will write $\mathbb{E}_{n\mathbf{h}_n}$ for a conditional expectation given $\mathbf{H}_n^\pi = \mathbf{h}_n, \ \mathbf{h}_n \in \mathcal{H}_n$. The value of a policy $\pi \in \mathbf{\Pi}$ with $\pi = (d_0, d_1, \ldots, d_{N-1})$ at time $n = 0, \ldots, N$ is defined as

$$V_{N\pi}(\mathbf{h}_N) := g(s_N + t_N c_N(x_N)),$$
$$V_{n\pi}(\mathbf{h}_n) := \mathbb{E}_{n\mathbf{h}_n}\left[g\left(s_n + t_n \left(\sum_{k=n}^{N-1} \beta^{k-n} c_k(X_k^\pi, d_k(\mathbf{H}_k^\pi), X_{k+1}^\pi) + \beta^{N-n} c_N(X_N^\pi) \right) \right) \right], \tag{4.3}$$

where $\mathbf{h}_n \in \mathcal{H}_n$. The corresponding value functions are

$$V_n(\mathbf{h}_n) := \inf_{\pi \in \mathbf{\Pi}} V_{n\pi}(\mathbf{h}_n), \qquad \mathbf{h}_n \in \mathcal{H}_n\,. \tag{4.4}$$

Obviously, we have $V_0(x, 0, 1) = \inf_{\sigma \in \Pi} \mathbb{E}[g(C_N^{\sigma x})]$. This means in the end, the quantity of interest is $V_0(x, 0, 1)$.

Remark 4.1 If there is no discounting or if the discounting is included in the non-stationary one-stage cost functions, the second summary variable t is obviously not needed. In the special case that ρ_ϕ is the Expected Shortfall, one only has to consider the functions $g_q(x) = (x - q)^+$, $q \in \mathbb{R}$, see (2.2). Due to their positive homogeneity in (x, q), it suffices to extend the state space by only one real-valued summary variable even if there is discounting, cf. Bäuerle and Ott (2011).

4.2 Solution of the extended MDP

We show next how to solve (4.4). It turns out that optimal policies can be found among Markov policies. Hence, let us now consider Markov policies $\pi \in \mathbf{\Pi}^M$, i.e. $\pi = (d_0, \ldots, d_{N-1})$ with $d_n : \mathbf{E} \to A$ such that $d_n(x, s, t) \in D_n(x)$. The function space

$$\begin{aligned} \mathbb{M} := \big\{ v : \mathbf{E} \to \mathbb{R}_+ \mid\ & v \text{ is lower semicontinuous,} \\ & v(x, \cdot, \cdot) \text{ is increasing for all } x \in E, \\ & v(x, s, t) \geq g(s) \text{ for } (x, s, t) \in \mathbf{E} \big\} \end{aligned}$$

turns out to be the set of potential value functions under such policies. In order to simplify the notation, we introduce the usual operators on $\mathbb{M}$. All $v \in \mathbb{M}$ are non-negative. Thus, integrals are well-defined with values in $\mathbb{R}_+ \cup \{\infty\}$.

Definition 4.2 For $v \in \mathbb{M}$ and a Markov decision rule $d : \mathbf{E} \to A$ we define

$$\begin{aligned} L_n v(x, s, t, a) &:= \mathbb{E}\Big[v\Big(\mathbf{T}_n(x, s, t, a, Z_{n+1}) \Big) \Big] \\ &= \mathbb{E}\Big[v\Big(T_n(x, a, Z_{n+1}),\ s + t c_n(x, a, T_n(x, a, Z_{n+1})),\ \beta t \Big) \Big], \qquad (x, s, t, a) \in \mathbf{D}_n, \\ \mathcal{T}_{nd} v(x, s, t) &:= L_n v(x, s, t, d(x, s, t)), \qquad (x, s, t) \in \mathbf{E}, \\ \mathcal{T}_n v(x, s, t) &:= \inf_{a \in D_n(x)} L_n v(x, s, t, a), \qquad (x, s, t) \in \mathbf{E}. \end{aligned}$$

The next result shows that $V_n(\mathbf{h}_n)$ depends only on (x_n, s_n, t_n), that V_n satisfies a Bellman equation and that an optimal policy exists and is Markov. All proofs are deferred to the appendix.

Theorem 4.3 *Let Assumption 3.1 be satisfied.*

a) *The value functions V_n only depend on (x_n, s_n, t_n), i.e. $V_n(\mathbf{h}_n) = J_n(x_n, s_n, t_n)$ for all $\mathbf{h}_n \in \mathcal{H}_n$ and $J_n \in \mathbb{M}$, $n = 0, \ldots, N$.*
b) *The J_n satisfy for $n = 0, \ldots, N$ the Bellman equation*

$$J_N(x, s, t) = g(s + tc_N(x)),$$
$$J_n(x, s, t) = \mathcal{T}_n J_{n+1}(x, s, t), \qquad (x, s, t) \in \mathbf{E}.$$

c) *There exist Markov decision rules $d_n^* : \mathbf{E} \to A$ for $n = 0, \ldots, N-1$ with $\mathcal{T}_{nd_n^*} J_{n+1} = \mathcal{T}_n J_{n+1}$ and every sequence of such minimizers constitutes an optimal policy $\pi^* = (d_0^*, \ldots, d_{N-1}^*) \in \mathbf{\Pi}^M$ for problem (4.4).*
d) *Given $\pi^* = (d_0^*, \ldots, d_{N-1}^*) \in \mathbf{\Pi}^M$ as in part c), an optimal policy $\sigma^* = (f_0^*, \ldots, f_{N-1}^*) \in \Pi$ for problem (4.1) is given by*

$$f_0^*(x_0) := d_0^*(x_0, 0, 1),$$
$$f_n^*(h_n) := d_n^*(x_n, s_n, t_n), \qquad n = 1, \ldots, N-1,$$

with s_n and t_n as in (4.2).

Remark 4.4 From Theorem 4.3 it follows that the sequence $\{(x_n, s_n, t_n)\}_{n=0}^{N-1}$ with

$$(s_n, t_n) = \left(\sum_{k=0}^{n-1} \beta^k c_k(x_k, a_k, x_{k+1}),\ \beta^n \right)$$

is a *sufficient statistic* of the decision model with the original state space in the sense of Hinderer (1970).

4.3 Outer problem: existence and numerical approximation

In this subsection, we study the existence of a solution to the outer optimization problem (3.4) under a finite planning horizon and its numerical approximation. We have assumed that for all $x \in E$ there exists a policy σ such that $C_N^{\sigma x} \in L^1$ and thus $\rho_\phi(C_N^{\sigma x}) =: \bar{\rho} < \infty$. Hence in what follows we can restrict to policies σ such that $\rho_\phi(C_N^{\sigma x}) \le \bar{\rho}$. In this case, we can further restrict the set G in the representation of Proposition 2.6.

Lemma 4.5 *It is sufficient to consider functions $g \in G$ in the representation of Proposition 2.6 which are $\phi(1)$-Lipschitz and satisfy*

$$0 \le g(x) \le \bar{g}(x) := \phi(1)x^+ + \bar{\rho}, \qquad x \in \mathbb{R}.$$

The space of such functions is denoted by $\mathcal{G}$.

In order to stress that the value function $V_0(x, 0, 1) = J_0(x, 0, 1)$ in Theorem 4.3 depends on g we write $J_0(g) := J_0(x, 0, 1)$ and suppress the dependence on the other variables. For initial state $x \in E$ and finite planning horizon $N \in \mathbb{N}$ the outer problem is given by

$$\inf_{g \in \mathcal{G}} J_0(g) + \int_0^1 g^*(\phi(u))\mathrm{d}u \tag{4.5}$$

We obtain now:

Theorem 4.6 *Under Assumption 3.1 there exists a solution $g \in \mathcal{G}$ for the outer optimization problem (4.5).*

As we know now that a solution to the outer optimization problem (4.5) exists, we aim to determine the solution numerically. The idea is to approximate the functions $g \in \mathcal{G}$ by piecewise linear ones and thereby obtain a finite dimensional optimization problem which can be solved with classical methods of global optimization. We are going to show that the minimal values converge when the approximation is continuously refined and give an error bound. Regarding the second summand of the objective function (4.5) our method coincides with the *Fast Legendre-Fenchel Transform (FLT)* algorithm studied for example by Corrias (1996).

For unbounded cost $C_N^{\sigma x}$ the functions $g \in \mathcal{G}$ would have to be approximated on the whole non-negative real line. This is numerically not feasible.

Assumption 4.7 We require additionally to Assumption 3.1 that c is bounded from above by a constant $\bar{c} \in \mathbb{R}_+$.

Consequently, it holds $0 \le C_N^{\sigma x} \le \hat{c} := \sum_{k=0}^N \beta^k \bar{c}$. The bounded cost allows for a further reduction of the feasible set of the outer problem. On the reduced feasible set, the second summand of the objective function is guaranteed to be finite and easier to calculate. Recall that the convex conjugate of $g \in \mathcal{G}$ is an $\mathbb{R} \cup \{\infty\}$-valued function defined by $g^*(y) := \sup_{s \in \mathbb{R}}\{sy - g(s)\},\ y \in \mathbb{R}$.

Lemma 4.8 a) *Under Assumption 4.7, a minimizer of the outer optimization problem (4.5) lies in*

$$\widehat{\mathcal{G}} := \left\{ g \in \mathcal{G} :\ g(s) = g(0) \text{ for } s < 0 \text{ and } g(s) = g(\hat{c}) + \phi(1)(s - \hat{c}) \text{ for } s > \hat{c} \right\}.$$

b) *For $g \in \widehat{\mathcal{G}}$ and $y \in [0, \phi(1)]$ it holds $g^*(y) = \max_{s \in [0,\hat{c}]}\{sy - g(s)\} < \infty$.*

The fact that the supremum of the convex conjugate reduces to the maximum of a continuous function over a compact set, opens the door for a numerical approximation with the FLT algorithm. By definition of $\widehat{\mathcal{G}}$, it is sufficient to approximate the functions $g \in \widehat{\mathcal{G}}$ on the interval $I := [0, \hat{c}]$. For the piecewise linear approximation we consider equidistant partitions $0 = s_1 < s_2 < \cdots < s_m = \hat{c}$, i.e. $s_k = (k-1)\frac{\hat{c}}{m-1},\ k = 1, \ldots, m,\ m \ge 2$. Let us define the mapping

$$p_m(g)(s) := g(s_k) + \frac{g(s_{k+1}) - g(s_k)}{s_{k+1} - s_k}(s - s_k), \qquad s \in [s_k, s_{k+1}],\ k = 1, \ldots, m-1,$$

which projects a function $g \in \widehat{\mathcal{G}}$ to its piecewise linear approximation and its image $\widehat{\mathcal{G}}_m := \{p_m(g) : g \in \widehat{\mathcal{G}}\}$. For considering the restriction of the outer optimization problem (4.5) to $\widehat{\mathcal{G}}_m$ it is convenient to define for $g \in \widehat{\mathcal{G}}$

$$K_m(g) := J_0(p_m(g)) + \int_0^1 p_m(g)^*(\phi(u))\mathrm{d}u$$

$$\text{and} \qquad K(g) := J_0(g) + \int_0^1 g^*(\phi(u))\mathrm{d}u.$$

Proposition 4.9 *It holds*

$$\left| \inf_{g\in\widehat{\mathcal{G}}} K_m(g) - \inf_{g\in\widehat{\mathcal{G}}} K(g) \right| \le \sup_{g\in\widehat{\mathcal{G}}} |K_m(g) - K(g)| \le 2\phi(1)\frac{\hat{c}}{m-1}.$$

The proposition shows that the infimum of K_m converges to the one of K. The error of restricting the outer problem (4.5) to $\widehat{\mathcal{G}}_m$ is bounded by $2\phi(1)\frac{\hat{c}}{m-1}$. The piecewise linear functions $g \in \widehat{\mathcal{G}}_m$ are uniquely determined by their values in the kinks $s_1, \ldots, s_m$. Hence, we can identify $\widehat{\mathcal{G}}_m$ with the compact set

$$\Gamma_m := \left\{ (y_1, \ldots, y_m) \in \mathbb{R}^m : \; y_1 \in I, \; 0 \le \frac{y_2 - y_1}{s_2 - s_1} \le \cdots \le \frac{y_m - y_{m-1}}{s_m - s_{m-1}} \le \phi(1) \right\}.$$

Note that due to translation invariance of ρ_ϕ it holds under Assumption 4.7 for $g \in \widehat{\mathcal{G}}$ that $g(0) \le \bar{g}(0) = \bar{\rho} \le \rho(\hat{c}) = \hat{c}$. Thus, the outer problem (4.5) restricted to $\widehat{\mathcal{G}}_m$ becomes finite dimensional:

$$\inf_{y\in\Gamma_m} J_0(g_y) + \int_0^1 g_y^*(\phi(u)), \tag{4.6}$$

where $g_y \in \widehat{\mathcal{G}}_m$ is the piecewise linear function induced by $y \in \Gamma_m$, i.e.

$$g_y(s) := y_k + \frac{y_{k+1} - y_k}{s_{k+1} - s_k}(s - s_k), \qquad s \in [s_k, s_{k+1}], \; k = 1, \ldots, m-1.$$

How to evaluate $J_0(\cdot)$ in g_y, $y \in \Gamma_m$, has been discussed in Sect. 4.1. The next Lemma simplifies the evaluation of the second summand of the objective function (4.6) to calculating the integrals $\int_{u_k}^{u_{k+1}} \phi(u)\mathrm{d}u$, where $u_0 := 0$, $u_k := \phi^{-1}\left(\frac{y_{k+1}-y_k}{s_{k+1}-s_k}\right)$, $k = 1, \ldots, m-1$ and $u_m := \phi(1)$.

Lemma 4.10 *The convex conjugate of g_y^*, $y \in \Gamma_m$, in $\xi \in [0, \phi(1)]$ is given by*

$$g_y^*(\xi) = \begin{cases} -y_1, & 0 \le \xi < \frac{y_2-y_1}{s_2-s_1}, \\ s_{k+1}\xi - y_{k+1}, & \frac{y_{k+1}-y_k}{s_{k+1}-s_k} \le \xi \le \frac{y_{k+2}-y_{k+1}}{s_{k+2}-s_{k+1}}, \; k = 1, \ldots, m-2 \\ s_m\xi - y_m, & \frac{y_m-y_{m-1}}{s_m-s_{m-1}} < \xi \le \phi(1). \end{cases}$$

The results of this section can be used to set up an algorithm for optimization problem (3.2). First we have to set $m := \left\lceil \frac{2\phi(1)\hat{c}}{\epsilon} \right\rceil + 1$ when we want to solve the problem with error estimate ϵ. Then choose $y_0 \in \Gamma_m$ and solve the inner problem with g_{y_0}. Use a global optimization procedure to select the next y_1, like, e.g. simulated annealing, and eventually determine the optimal value of (4.6). Note that we do not have convexity of (4.6) in y.

Algorithm 1: Outer problem

Data: Markov Decision Model

Output: Optimal policy σ^*, minimal risk-sensitive cost $\rho_\phi(C_N^{\sigma^* x})$

1. Select an approximation error $\epsilon > 0$ and set $m := \left\lceil \frac{2\phi(1)\hat{c}}{\epsilon} \right\rceil + 1$.
2. Solve (4.6) with an algorithm for global optimization.
For each evaluation of $J_0(\cdot)$ solve the inner problem (4.4) with Theorem 4.3.

It is worth noting that an optimal policy $\sigma^* = (f_0^*, \ldots, f_{N-1}^*) \in \Pi$ obtained with the algorithm is in general not *time consistent*. If one implements the policy σ^* and considers optimization problem (3.2) again at a later point in time $n \in \{1, \ldots, N-1\}$, one can disregard the cost $\sum_{k=0}^{n-1} \beta^k c_k(x_k, a_k, x_{k+1})$ which is already realized due to the translation invariance of ρ_ϕ and faces the remaining optimization problem

$$\inf_{\sigma \in \Pi} \rho_\phi \left(\sum_{k=n}^{N-1} \beta^k c_k(X_k^\sigma, f_k(H_k^\sigma), X_{k+1}^\sigma) + \beta^N c_N(X_N^\sigma) \right). \tag{4.7}$$

But for (4.7) the remaining policy $(f_n^*, \ldots, f_{N-1}^*)$ will in general not be optimal. The reason it that the optimal function g^* of the outer optimization problem will change due to Remark 2.7. However, for a fixed $g \in G$ the optimal solution of inner optimization problem is time consistent by the Bellman equation in Theorem 4.3. A more detailed discussion of time consistent policies for risk-sensitive MDP can be found in Shapiro (2009). Time consistency can alternatively be defined as a property of the risk measure. How this is related to the more general policy-based viewpoint is discussed in Shapiro and Uğurlu (2016).

5 Extensions and further results

5.1 Infinite planning horizon

In this subsection, we consider the risk-sensitive total cost minimization (3.3) under an infinite planning horizon. This is reasonable if the terminal period is unknown or if one wants to approximate a model with a large but finite planning horizon. Solving the infinite horizon problem will turn out to be easier since it admits a stationary optimal policy.

We study the stationary version of the decision model with no terminal cost, i.e. D, T, c do not depend on n, $c_N \equiv 0$ and the disturbances are identically distributed. Let Z be a representative of the disturbance distribution. Our first aim is to solve again

the inner problem

$$\inf_{\sigma\in\Pi} \mathbb{E}[g(C_\infty^{\sigma x})] \tag{5.1}$$

for an arbitrary but fixed increasing convex function $g \in G$. As in the previous section we assume w.l.o.g. that $g \geq 0$ and that for all $x \in E$ there exists a policy σ such that $C_\infty^{\sigma x} \in L^1$.

The remarks in Sect. 3 regarding connections to the minimization of (rank-dependent) expected disutilities and corresponding certainty equivalents apply in the infinite horizon case as well.

In order to obtain a solution by value iteration, the state space is extended to $\mathbf{E} := E \times \mathbb{R}_+ \times (0, \infty)$ as in Sect. 4. The action space A and the admissible state-action combinations $\mathbf{D}$ remain unchanged, i.e. $\mathbf{D} := \{(x, s, t, a) \in \mathbf{E} \times A : a \in D(x)\}$ and $\mathbf{D}(x, s, t) := D(x),\ (x, s, t) \in \mathbf{E}$. The transition function on the new state space is given by $\mathbf{T} : \mathbf{D} \times \mathcal{Z} \to \mathbf{E}$,

$$\mathbf{T}(x, s, t, a, z) := \begin{pmatrix} T(x, a, z) \\ s + tc(x, a, T(x, a, z)) \\ \beta t \end{pmatrix}.$$

Since the model with infinite planning horizon will be derived as a limit of the one with finite horizon, the consideration can be restricted to Markov policies $\pi = (d_1, d_2, \dots) \in \mathbf{\Pi}^M$ due to Theorem 4.3.

The value of a policy $\pi = (d_1, d_2, \dots) \in \mathbf{\Pi}^M$ under an infinite planning horizon is defined as

$$J_{\infty\pi}(x, s, t) := \mathbb{E}_{0x}\left[g\left(s + t \sum_{k=0}^{\infty} \beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi) \right) \right], \quad (x, s, t) \in \mathbf{E}.$$

Note that $J_{\infty\pi}$ is well-defined since $c \geq 0$. The infinite horizon value function is

$$J_\infty(x, s, t) := \inf_{\pi\in\mathbf{\Pi}^M} J_{\infty\pi}(x, s, t), \qquad (x, s, t) \in \mathbf{E}. \tag{5.2}$$

We obviously get that $\inf_{\sigma\in\Pi} \mathbb{E}[g(C_\infty^{\sigma x})] = J_\infty(x, 0, 1)$. The operators $\mathcal{T}$ and $\mathcal{T}_d$ which appear in the next theorem are defined as in Definition 4.2 for the stationary model data.

Theorem 5.1 *Let Assumption 3.1 be satisfied. Then it holds:*

a) *The infinite horizon value function J_∞ is the smallest fixed point of the Bellman operator $\mathcal{T}$ in $\mathbb{M}$.*
b) *There exists a Markov decision rule d^* such that $\mathcal{T}_{d^*} J_\infty = \mathcal{T} J_\infty$ and each stationary policy $\pi^* = (d^*, d^*, \dots) \in \mathbf{\Pi}^M$ induced by such a decision rule is optimal for optimization problem (5.2).*

c) *Given* $\pi^* = (d^*, d^*, \dots) \in \mathbf{\Pi}^M$ *as in part b), an optimal policy* $\sigma^* = (f_0^*, f_1^*, \dots) \in \Pi$ *for problem* (5.1) *is given by*

$$f_0^*(x_0) := d^*(x_0, 0, 1),$$
$$f_n^*(h_n) := d^*(x_n, s_n, t_n), \qquad n \in \mathbb{N},$$

with s_n *and* t_n *as in* (4.2).

The solution of the outer optimization problem

$$\inf_{g \in \mathcal{G}} J_\infty(g) + \int_0^1 g^*(\phi(u))\mathrm{d}u \tag{5.3}$$

follows the same lines as in the case of a finite time horizon. Again we can restrict to policies σ such that $\rho_\phi(C_\infty^{\sigma x}) \le \bar{\rho}$. Lemma 4.5 which reduces the outer optimization problem to $\mathcal{G}$ holds also in the infinite horizon case as well as Theorem 4.6 which states the existence of a solution to the outer problem.

The numerical approximation scheme for the infinite horizon works under the following assumption:

Assumption 5.2 In addition to Assumption 3.1 we require that c is bounded from above by a constant $\bar{c} \in \mathbb{R}_+$ and that $\beta \in (0, 1)$.

Hence, it holds that $0 \le C_\infty^{\sigma x} \le \hat{c}$ with $\hat{c} = \frac{\bar{c}}{1-\beta}$ and we obtain in the same way as Lemma 4.8:

Lemma 5.3 a) *Under Assumption* 5.2*, a minimizer of the outer optimization problem* (5.3) *lies in*

$$\widehat{\mathcal{G}} = \left\{ g \in \mathcal{G} : \ g(s) = g(0) \text{ for } s < 0 \text{ and } g(s) = g(\hat{c}) + \phi(1)(s - \hat{c}) \text{ for } s > \hat{c} \right\}.$$

b) *For* $g \in \widehat{\mathcal{G}}$ *and* $y \in [0, \phi(1)]$ *it holds* $g^*(y) = \max_{s \in [0,\hat{c}]}\{sy - g(s)\} < \infty$.

The remaining part of the numerical algorithm works as in the case of finite time horizon.

5.2 Relaxed assumptions for monotone models

The model has been introduced in Sect. 3 with a general Borel space as state space. In order to solve the optimization problem with finite or infinite time horizon we assumed a continuous transition function despite having a semicontinuous model. This assumption on the transition function can be relaxed to semicontinuity if the state space is the real line and the transition and one-stage cost function have some form of monotonicity. For notational convenience, we consider the stationary model with no terminal cost under both finite and infinite horizon in this section. We replace Assumption 3.1 by

Assumption 5.4 (i) The state space is the real line $E = \mathbb{R}$.
(ii) The sets $D(x)$ are compact and $\mathbb{R} \ni x \mapsto D(x)$ is upper semicontinuous and decreasing, i.e. $D(x) \supseteq D(y)$ for $x \leq y$.
(iii) The transition function T is lower semicontinuous in (x, a) and increasing in x.
(iv) The one-stage cost $c(x, a, T(x, a, z))$ is lower semicontinuous in (x, a) and increasing in x.

Requiring that the one-stage cost function c is lower semicontinuous in (x, a, x') and increasing in (x, x') is sufficient for Assumption 5.4 (iv) to hold due to part (iii) of the assumption.

How do the modified continuity assumptions affect the validity of the results in Sects. 4.1 and 5.1? The only two results that were proven using the continuity of the transition function T in (x, a) and not only its measurability are Theorems 4.3 and 5.1. All other statements are unaffected.

Proposition 5.5 *The assertions of Theorems 4.3 and 5.1 hold under Assumption 5.4, too. Moreover, the value functions J_n and J_∞ are increasing. The set of potential value functions can therefore be replaced by*

$$\mathbb{M} = \big\{v : \mathbf{E} \to \mathbb{R} \mid v \text{ is lower semicontinuous and increasing, } v(x, s, t) \geq g(s) \text{ for } (x, s, t) \in \mathbf{E}\big\}.$$

The monotonicity properties of Assumption 5.4 can be used to construct a convex model.

Lemma 5.6 *Let Assumption 5.4 be satisfied, A be a subset of a real vector space, the admissible state-action-combinations D be a convex set, the transition function T be convex in (x, a) and the one-stage cost $D \ni (x, a) \mapsto c(x, a, T(x, a, z))$ be a convex function for every $z \in \mathcal{Z}$. Then, the value functions $J_n(\cdot, \cdot, t)$ and $J_\infty(\cdot, \cdot, t)$ are convex for every $t > 0$.*

If c is increasing in x', it is sufficient to require that c and T are convex in (x, a). The monotonicity requirements in Assumption 5.4 are only one option. The following alternative is relevant in particular for the dynamic reinsurance model in Sect. 6. For a proof see Section 6.1.3 in Glauner (2020).

Corollary 5.7 *Change Assumption 5.4 (ii)–(iv) to*

(ii') *The sets $D(x)$ are compact and $\mathbb{R} \ni x \mapsto D(x)$ is upper semicontinuous and increasing.*
(iii') *T is upper semicontinuous in (x, a) and increasing in x.*
(iv') *$c(x, a, T(x, a, z))$ is lower semicontinuous in (x, a) and decreasing in x.*

Then, the assertions of Theorems 4.3 and 5.1 still hold with the value functions J_n and J_∞ being decreasing in x and increasing in (s, t).

If furthermore A is a subset of a real vector space, D a convex set, T concave in (x, a) and $D \ni (x, a) \mapsto c(x, a, T(x, a, z))$ convex for every $z \in \mathcal{Z}$, then the value functions $J_n(\cdot, \cdot, t)$ and $J_\infty(\cdot, \cdot, t)$ are convex for every $t > 0$.

6 Dynamic optimal reinsurance

As an application, we present a dynamic extension in discrete time of the static optimal reinsurance problem

$$\min_{\ell \in \mathcal{L}} \quad r_{\mathrm{CoC}} \cdot \rho\big(\ell(Y) + \pi_R(\ell)\big). \tag{6.1}$$

In this setting, the insurance company incurs an aggregate loss $Y \in L^1_{\geq 0}$ at the end of a fixed period due to insurance claims. In order to reduce its risk, the insurer concludes a reinsurance contract ℓ to transfer a part of its potential loss to a reinsurance company. The reinsurance contract ℓ determines the loss $\ell(Y(\omega))$ retained by the insurance company in each scenario $\omega \in \Omega$. For the risk transfer, the insurer has to compensate the reinsurer with a reinsurance premium $\pi_R(\ell) := \pi_R(Y - \ell(Y))$, where $\pi_R : L^1_{\geq 0} \to \mathbb{R}$ is a premium principle with properties similar to a risk measure. Most widely used is the expected premium principle $\pi_R(X) = (1+\theta)\mathbb{E}[X]$ with safety loading $\theta > 0$. In order to preclude moral hazard, it is standard in the actuarial literature to assume that both ℓ and the ceded loss function $\mathrm{id}_{\mathbb{R}_+} - \ell$ are increasing. Hence, the set of admissible retained loss functions is

$$\mathcal{L} = \{\ell : \mathbb{R}_+ \to \mathbb{R}_+ \mid \ell(y) \leq y \ \forall y \in \mathbb{R}_+,\ \ell \text{ increasing},\ \mathrm{id}_{\mathbb{R}_+} - \ell \text{ increasing}\}.$$

The insurer's target is to minimize its cost of solvency capital which is calculated as the cost of capital rate $r_{\mathrm{CoC}} \in (0, 1]$ times the solvency capital requirement determined by applying the risk measure ρ to the insurer's effective risk after reinsurance.

First research on the optimal reinsurance problem (6.1) dates back to the 1960s. Borch (1960) proved that a stop loss reinsurance contract minimizes the variance of the retained loss of the insurer given the premium is calculated with the expected value principle. A similar result has been derived in Arrow (1963) where the expected utility of terminal wealth of the insurer has been maximized. Since then a lot of generalizations of this problem have been considered. For a comprehensive literature overview, we refer to Albrecher et al. (2017). Since the 2000s, Expected Shortfall has become of special interest. Chi and Tan (2013) identified layer reinsurance contracts as optimal for Expected Shortfall under general premium principles. Their results were extended to general distortion risk measures by Cui et al. (2013). Other generalizations concerned additional constraints, see e.g. Lo (2017), or multidimensional settings induced by a macroeconomic perspective, see Bäuerle and Glauner (2018). We are not aware of any dynamic generalizations in the literature.

Reinsurance treaties are typically written for one year, cf. Albrecher et al. (2017). Hence, it is appropriate to model such an extension in discrete time. The insurer's annual surplus has the dynamics

$$X_0 = x, \qquad X_{n+1} = X_n + Z_{n+1} - \ell_n(Y_{n+1}) - \pi_R(\ell_n),$$

where the bounded, non-negative random variable $Z_{n+1} \in L^\infty_{\geq 0}$ represents the insurer's premium income from its customers in the n-th period. The premium principle π_R :

$L^p_{\geq 0} \to \mathbb{R}$ of the reinsurer is assumed to be law-invariant, monotone, normalized and to have the Fatou property. Normalization means that $\pi_R(0) = 0$ and the Fatou property is lower semicontinuity w.r.t. dominated convergence.

The Markov Decision Model is given by the state space $E = \mathbb{R}$, the action space $A = \mathcal{L}$, either no constraint or a budget constraint $D(x) = \{\ell \in \mathcal{L} : \pi_R(\ell) \leq x^+\}$, the independent disturbances $(Y_n, Z_n)_{n\in\mathbb{N}}$ with $Y_n \in L^1_{\geq 0}$ and $Z_n \in L^\infty_{\geq 0}$, the transition function $T(x, \ell, y, z) = x - \ell(y) - \pi_R(\ell) + z$ and the one-stage cost function $c(x, \ell, x') = x - x'$. A reinsurance policy is a sequence $\sigma = (f_0, \ldots, f_{N-1})$ of measurable decision rules $f_n : \mathcal{H}_n \to \mathcal{L}$ selecting the reinsurance contract at each stage based on the available information. The insurance companies target is to minimize its solvency cost of capital for the total discounted loss

$$\inf_{\sigma\in\Pi} \quad r_{\mathrm{CoC}} \cdot \rho_\phi \left(\sum_{k=0}^{N-1} \beta^k \Big(f_k(H_k^\sigma)(Y_{k+1}) + \pi_R(f_k(H_k^\sigma)) - Z_{k+1} \Big) \right), \tag{6.2}$$

where ρ_ϕ is a spectral risk measure with bounded spectrum ϕ, $\beta \in (0, 1]$ and $N \in \mathbb{N}$. As it is irrelevant for the minimization, we will in the sequel omit the cost of capital rate r_{CoC} and instead minimize the capital requirement. For $\beta = 1$ we have

$$\sum_{k=0}^{N-1} f_k(H_k^\sigma)(Y_{k+1}) + \pi_R(f_k(H_k^\sigma)) - Z_{k+1} = \sum_{k=0}^{N-1} X_k^\sigma - X_{k+1}^\sigma = x - X_N^\sigma,$$

i.e. due to translation invariance of spectral risk measures the objective reduces to minimizing the capital requirement for the loss (negative surplus) at the planing horizon $-X_N^\sigma$. This is reminiscent of the static reinsurance problem (6.1), however here the loss distribution at the planing horizon can be controlled by interim action. Throughout, we have required that the one-stage cost $c(x, \ell, T(x, \ell, Y, Z)) = \ell(Y) + \pi_R(\ell) - Z$ is non-negative. As $\ell(Y)$ and $\pi_R(\ell)$ are non-negative for all $\ell \in \mathcal{L}$ and $c(x, \mathrm{id}_{\mathbb{R}_+}, T(x, \mathrm{id}_{\mathbb{R}_+}, Y, Z)) = Y - Z$ due to normalization of π_R, the premium income Z would have to be non-positive. This makes no sense from an actuarial point of view, but since ρ_ϕ is translation invariant and $Z \in L^\infty$ we can add $\sum_{k=0}^{N-1} \beta^k \operatorname{ess\,sup}(Z)$ without influencing the minimization. This means that the one-stage cost function is changed to $\hat{c}(x, \ell, x') = x - x' + \operatorname{ess\,sup}(Z)$. The economic interpretation is that the one-stage cost

$$\hat{c}(x, \ell, T(x, \ell, Y, Z)) = \ell(Y) + \pi_R(\ell) + \operatorname{ess\,sup}(Z) - Z$$

now depends on the deviation from the maximal possible income instead of the actual income. For brevity we write $\hat{z} = \operatorname{ess\,sup}(Z)$.

As in (3.4) we separate an inner and outer reinsurance problem. For a structural analysis we can focus on the inner optimization problem

$$\inf_{\sigma\in\Pi} \mathbb{E}\left[g\left(\sum_{k=0}^{N-1} \beta^k \Big(f_k(H_k^\sigma)(Y_{k+1}) + \pi_R(f_k(H_k^\sigma)) + \hat{z} - Z_{k+1} \Big) \right) \right] \tag{6.3}$$

with arbitrary $g \in \mathcal{G}$, cf. Lemma 4.5. Note that for $\sigma = (f, \dots, f)$ with the constant decision rule $f \equiv \mathrm{id}_{\mathbb{R}_+}$ representing full retention at all stages we obtain $\rho_\phi(C_N^{\sigma x}) < \infty$. On the extended state space $\mathbf{E} = \mathbb{R} \times \mathbb{R}_+ \times (0, 1]$, the value of a policy $\pi = (d_0, \dots, d_{N-1}) \in \mathbf{\Pi}$ is defined as

$$V_{N\pi}(\mathbf{h}_N) = g(s_N),$$
$$V_{n\pi}(\mathbf{h}_n) = \mathbb{E}_{n\mathbf{h}_n}\left[g\left(s_n + t_n \sum_{k=n}^{N-1} \beta^{k-n} \Big(d_k(\mathbf{H}_k^\pi)(Y_{k+1}) + \pi_R(d_k(\mathbf{H}_k^\pi)) + \hat{z} - Z_{k+1} \Big) \right) \right],$$

for $n = 0, \dots, N$ and $\mathbf{h}_n \in \boldsymbol{\mathcal{H}}_n$. The corresponding value functions are

$$V_n(\mathbf{h}_n) = \inf_{\pi \in \mathbf{\Pi}} V_{n\pi}(\mathbf{h}_n), \qquad \mathbf{h}_n \in \boldsymbol{\mathcal{H}}_n .$$

Due to the real state space we want to apply Corollary 5.7 for solving the optimization problem. Note that the one-stage cost $\hat{c}$ is non-negative and the spectrum ϕ bounded by assumption. The following lemma shows that also the monotonicity, continuity and compactness assumptions of Corollary 5.7 are satisfied by the dynamic reinsurance model.

Lemma 6.1 a) *The retained loss functions $\ell \in \mathcal{L}$ are Lipschitz continuous with constant $L \le 1$. Moreover, $\mathcal{L}$ is a Borel space as a compact subset of the metric space $(C(\mathbb{R}_+), m)$ of continuous real-valued functions on $\mathbb{R}_+$ with the metric of compact convergence.*

b) *The functional $\pi_R : \mathcal{L} \to \mathbb{R}_+,\ \ell \mapsto \pi_R(\ell)$ is lower semicontinuous.*

c) *The transition function T is upper semicontinuous and increasing in x.*

d) *$D(x)$ is a compact subset of $\mathcal{L}$ for all $x \in \mathbb{R}$ and the set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is upper semicontinuous and increasing.*

e) *The one-stage cost $D \ni (x, \ell) \mapsto c(x, \ell, T(x, \ell, y, z))$ is lower semicontinuous and decreasing in x.*

Now, Corollary 5.7 yields that it is sufficient to minimize over all Markov policies and the value functions satisfy the Bellman equation

$$J_N(x, s, t) = g(s),$$
$$J_n(x, s, t) = \inf_{\ell \in D(x)} \mathbb{E}\Big[J_{n+1}\big(x - \ell(Y) - \pi_R(\ell) + Z,\ s + t\big(\ell(Y) + \pi_R(\ell) + \hat{z} - Z\big),\ \beta t \big) \Big] \tag{6.4}$$

for $(x, s, t) \in \mathbf{E}$ and $n = 0, \dots, N-1$. Moreover, there exists a Markov Decision rule $d_n^* : \mathbf{E} \to \mathcal{L}$ minimizing J_{n+1} and every sequence $\pi = (d_0^*, \dots, d_{N-1}^*) \in \mathbf{\Pi}^M$ of such minimizers is a solution to (6.3).

All structural properties of the optimal policy which do not depend on g are inherited by the optimal solution of the cost of capital minimization problem (6.2). The structural properties we will focus on in the rest of this section are induced by convexity. Therefore, we assume that the premium principle π_R is convex and that there is no

budget constraint. Note that $D(x)$ is non-convex even for convex π_R. Under these conditions, we have indeed a convex model: D is trivially convex, the transition function $T(x, \ell, y, z) = x - \ell(y) - \pi_R(\ell) + z$ is concave in (x, ℓ) as a sum of concave functions and the one-stage cost $(x, \ell) \mapsto \hat{c}(x, \ell, T(x, \ell, y, z)) = \ell(y) + \pi_R(\ell) + \hat{z} - z$ is convex as a sum of convex functions. Now, Corollary 5.7 yields that the value functions J_n are convex. Under the widely-used expected premium principle, the optimization problem can be reduced to finite dimension.

Example 6.2 Let $\pi_R(\cdot) = (1 + \theta)\mathbb{E}[\cdot]$ be the expected premium principle with safety loading $\theta > 0$ and assume there is no budget constraint. We will now show that the optimal reinsurance contracts (i.e. retained loss functions) can be chosen from the class of *stop loss* contracts

$$\ell(x) = \min\{x, a\}, \qquad a \in [0, \infty].$$

Due to the convexity of J_{n+1}, we can infer from the Bellman equation (6.4) that reinsurance contract ℓ_1 is better than ℓ_2 if

$$\ell_1(Y) + \pi_R(\ell_1) \leq_{cx} \ell_2(Y) + \pi_R(\ell_2),$$

where $\leq_{cx}$ denotes the convex order. Since $Y_1 \leq_{cx} Y_2$ implies $\mathbb{E}[Y_1] = \mathbb{E}[Y_2]$, it suffices to find an $a_\ell \in [0, \infty]$ such that

$$\min\{Y, a_\ell\} \leq_{cx} \ell(Y). \tag{6.5}$$

The mapping $[0, \infty] \ni a \mapsto \min\{Y(\omega), a\}$ is continuous for all $\omega \in \Omega$ and $0 \leq \min\{Y, a\} \leq Y \in L^1$. Thus, it follows from dominated convergence that $[0, \infty] \ni a \mapsto \mathbb{E}[\min\{Y, a\}]$ is continuous. Furthermore,

$$\mathbb{E}[\min\{Y, 0\}] \leq \mathbb{E}[\ell(Y)] \leq \mathbb{E}[\min\{Y, \operatorname{ess\,sup}(Y)\}].$$

Hence, by the intermediate value theorem there is an $a_\ell \in [0, \infty]$ such that $\mathbb{E}[\ell(Y)] = \mathbb{E}[\min\{Y, a_\ell\}]$. Let us compare the survival functions:

$$\begin{aligned} S_{\min\{Y, a_\ell\}}(y) &= \mathbb{P}(\min\{Y, a_\ell\} > y) = \mathbb{P}(Y > y)\mathbb{1}\{a_\ell > y\}, \\ S_{\ell(Y)}(y) &= \mathbb{P}(\ell(Y) > y) \leq \mathbb{P}(Y > y). \end{aligned}$$

The inequality holds since $\ell \leq \mathrm{id}_{\mathbb{R}_+}$. Hence, we have $S_{\min\{Y, a_\ell\}}(y) \geq S_{\ell(Y)}(y)$ for $y < a_\ell$ and $S_{\min\{Y, a_\ell\}}(y) \leq S_{\ell(Y)}(y)$ for $y \geq a_\ell$. The cut criterion 1.5.17 in Müller and Stoyan (2002) implies $\min\{Y, a_\ell\} \leq_{icx} \ell(Y)$ and (6.5) follows due to the equality in expectation, cf. Theorem 1.5.3 in Müller and Stoyan (2002). So the inner optimization problem (6.3) is reduced to finding an optimal nonnegative retention level of a stop loss contract at every stage. With this reduction, the dynamic reinsurance problem becomes numerically solvable without requiring a parametric approximation of the retained loss functions.

Table 1 Relative optimal retention level of the stop loss constract at time $t = 0$ for different parameters of the claim size distribution

	$\lambda = 1$	$\lambda = 0.5$	$\lambda = 0.25$	$\lambda = 0.125$
$\frac{a_0^*}{\text{ess sup}(Y)}$	0.96	0.98	0.99	0.99

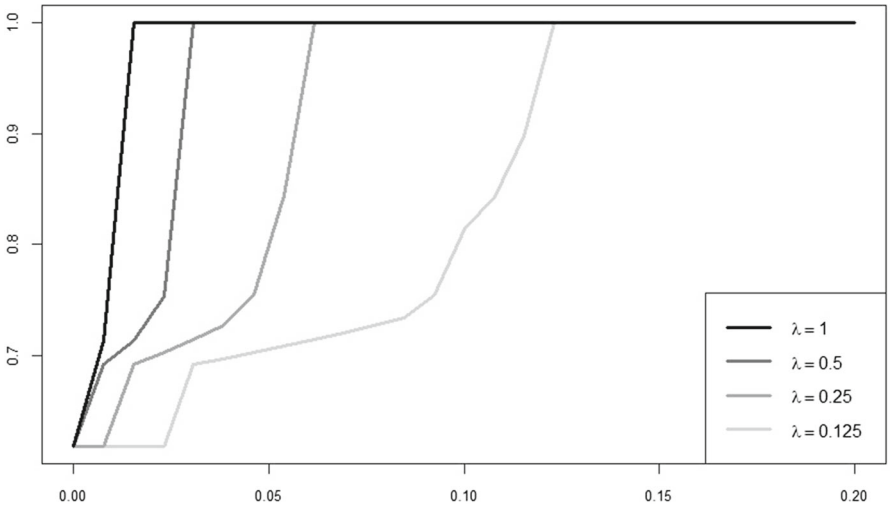

Fig. 1 Relative optimal retention level of the stop loss constract at time $t = 1$ as a function of the accumulated cost s for different parameters of the claim size distribution

Let us apply the algorithm derived at the end of Sect. 4.3 and study the optimal retention levels in a two-stage setting without discounting and deteministic income $Z \equiv z$. Due to the translation invariance of ρ_ϕ a deterministic income can be disregarded in the optimization. The stage-wise insurance claims Y are assumed to be exponentially distributed with parameter $\lambda > 0$, which is a classical choice in actuarial science. For numerical feasibility we truncate the distribution at the 99.9%-quantile. The safety loading of the expected premim principle is $\theta = 0.1$. As a concrete risk measure we consider Expected Shortfall $\text{ES}_{0.99}$ with the typical paramter $\alpha = 0.99$. Due to (2.2) we have to consider $\mathcal{G} = \{g_q : q \in \mathbb{R}\}$ with $g_q(s) = (s-q)^+$ and $\int_0^1 g^*(\phi(u))\mathrm{d}u = q$ for the outer optimization problem. We can infer recursively from the Bellman equation (6.4) that the value functions only depend on the accumulated cost s and not on the current capital x since there is no budget constraint. The same therefore holds for the optimal retention level.

Table 1 shows the optimal retention parameter at time $t = 0$ for different parameters of the truncated exponential distribution. For better comparability the parameter is shown relative to the maximal claim size ess sup(Z).

Figure 1 shows the optimal decision rule at time $t = 1$, i.e. the optimal retention level as a function of the accumulated cost s for different parameters of the truncated exponential distribution again relative to the maximal loss. Hence, the curves are given by $s \mapsto \frac{a_1^*(s)}{\text{ess sup}(Y)}$. It is worth noting that the parameter λ of the truncated exponential

distrubution has a structurally different influence than at time $t = 0$. At time $t = 1$, a smaller λ (i.e. a higher expected insurance claim) leads to a more convervative reinsurance contract in form of less risk retention, whereas at time $t = 0$ it increases the optimal retention level. By comparison, the influence of the safety loading θ and the level α of Expected Shortfall turned out to be negligible and is therefore not presented here.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendix

Proof for subsection 4.1

In order to prove Theorem 4.3 we need to show the value iteration for the value $V_{n\pi}$ of a fixed policy $\pi \in \mathbf{\Pi}$ first. This is done in the next proposition.

Proposition 7.1 *The value of a policy $\pi \in \mathbf{\Pi}$ can be calculated recursively for $n = 0, \ldots, N-1$ and $\mathbf{h}_n \in \mathcal{H}_n$ as*

$$\begin{aligned} V_{N\pi}(\mathbf{h}_N) &= g(s_N + t_N c_N(x_N)) \\ V_{n\pi}(\mathbf{h}_n) &= \mathbb{E}\Big[V_{n+1\pi}\Big(\mathbf{h}_n,\ d_n(\mathbf{h}_n),\ \mathbf{T}_n(x_n, s_n, t_n, d_n(\mathbf{h}_n), Z_{n+1})\Big)\Big] \\ &= \mathbb{E}\Big[V_{n+1\pi}\Big(\mathbf{h}_n,\ d_n(\mathbf{h}_n),\ T_n(x_n, d_n(\mathbf{h}_n), Z_{n+1}), \\ &\qquad s_n + t_n c_n(x_n, d_n(\mathbf{h}_n), T_n(x_n, d_n(\mathbf{h}_n), Z_{n+1})),\ \beta t\Big)\Big]. \end{aligned}$$

Proof The proof is by backward induction. At time N there is nothing to show. Now assume the assertion holds for $n+1$, then the tower property of conditional expectation yields

$$\begin{aligned} V_{n\pi}(\mathbf{h}_n) &= \mathbb{E}_{n\mathbf{h}_n}\Big[g\Big(s_n + t_n \Big(\sum_{k=n}^{N-1} \beta^{k-n} c_k(X_k^\pi, d_k(\mathbf{H}_k^\pi), X_{k+1}^\pi) + \beta^{N-n} c_N(X_N^\pi) \Big)\Big)\Big] \\ &= \mathbb{E}_{n\mathbf{h}_n}\Big[g\Big(s_n + t_n c_n(x_n, d_n(\mathbf{h}_n), T_n(x_n, d_n(\mathbf{h}_n), Z_{n+1})) \\ &\quad + t_n \beta \Big(\sum_{k=n+1}^{N-1} \beta^{k-(n+1)} c_k(X_k^\pi, d_k(\mathbf{H}_k^\pi), X_{k+1}^\pi) + \beta^{N-(n+1)} c_N(X_N^\pi) \Big)\Big)\Big] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{n\mathbf{h}_n}\Bigg[\mathbb{E}_{n+1(\mathbf{h}_n,d_n(\mathbf{h}_n),\mathbf{T}_n(x_n,s_n,t_n,d_n(\mathbf{h}_n),Z_{n+1}))}\Bigg[\\
&g\Bigg(s_n + t_n c_n(x_n, d_n(\mathbf{h}_n), T_n(x_n, d_n(\mathbf{h}_n), Z_{n+1}))\\
&\quad + t_n\beta\Bigg(\sum_{k=n+1}^{N-1}\beta^{k-(n+1)}c_k(X_k^\pi, d_k(\mathbf{H}_k^\pi), X_{k+1}^\pi) + \beta^{N-(n+1)}c_N(X_N^\pi)\Bigg)\Bigg)\Bigg]\Bigg]\\
&= \mathbb{E}\Big[V_{n+1\pi}\Big(\mathbf{h}_n,\ d_n(\mathbf{h}_n),\ \mathbf{T}_n(x_n, s_n, t_n, d_n(\mathbf{h}_n), Z_{n+1})\Big)\Big]. \qquad \square
\end{aligned}
$$

We are now in a position to prove Theorem 4.3. Note that the operators in Definition 4.2 are monotone in v. Under a Markov policy $\pi = (d_0, \ldots, d_{N-1}) \in \mathbf{\Pi}^M$ the value iteration in Proposition 7.1 can be expressed with the help of the operators. We denote the Markov value functions with J. More precisely, we set $J_{N\pi}(x, s, t) := g(s + tc_N(x)),\ (x, s, t) \in \mathbf{E}$ and we obtain according to Proposition 7.1 for $n = 0, \ldots, N-1$ and $(x, s, t) \in \mathbf{E}$

$$
\begin{aligned}
J_{n\pi}(x, s, t) &= \mathbb{E}\Big[J_{n+1\pi}\Big(T_n(x, d_n(x), Z_{n+1}),\ s + tc_n(x, d_n(x), T_n(x, d_n(x), Z_{n+1})),\ \beta t\Big)\Big]\\
&= \mathcal{T}_{nd_n} J_{n+1\pi}(x, s, t).
\end{aligned}
$$

The corresponding Markov value functions are defined for $n = 0, \ldots, N$ as

$$
J_n(x, s, t) := \inf_{\pi\in\Pi^M} J_{n\pi}(x, s, t), \qquad (x, s, t) \in \mathbf{E}.
$$

Proof of Theorem 4.3 The proof of parts a)-c) is by backward induction. At time N we have $V_N(\mathbf{h}_N) = J_N(x_N, s_N, t_N) = g(s_N + t_N c_N(x_N)),\ \mathbf{h}_N \in \mathcal{H}_N$, which is

- lower semicontinuous since g is increasing and continuous (as a convex function on $\mathbb{R}$) and c_N is lower semicontinuous,
- increasing in (s_N, t_N) since g is increasing and c_N is non-negative,
- bounded below by $g(s_N)$ since g is increasing and $t_N c_N(x_N) \geq 0$.

I.e. $J_N \in \mathbb{M}$. Assuming the assertion holds at time $n+1$ we have at time n for $\mathbf{h}_n \in \mathcal{H}_n$

$$
\begin{aligned}
V_n(\mathbf{h}_n) &= \inf_{\pi\in\mathbf{\Pi}} V_{n\pi}(\mathbf{h}_n)\\
&= \inf_{\pi\in\mathbf{\Pi}} \mathbb{E}\Big[V_{n+1\pi}\Big(\mathbf{h}_n,\ d_n(\mathbf{h}_n),\ \mathbf{T}_n(x_n, s_n, t_n, d_n(\mathbf{h}_n), Z_{n+1})\Big)\Big]\\
&\geq \inf_{\pi\in\mathbf{\Pi}} \mathbb{E}\Big[V_{n+1}\Big(\mathbf{h}_n,\ d_n(\mathbf{h}_n),\ \mathbf{T}_n(x_n, s_n, t_n, d_n(\mathbf{h}_n), Z_{n+1})\Big)\Big]\\
&= \inf_{\pi\in\mathbf{\Pi}} \mathbb{E}\Big[J_{n+1}\Big(\mathbf{T}_n(x_n, s_n, t_n, d_n(\mathbf{h}_n), Z_{n+1})\Big)\Big].\\
&= \inf_{a_n\in D_n(x_n)} \mathbb{E}\Big[J_{n+1}\Big(\mathbf{T}_n(x_n, s_n, t_n, a_n, Z_{n+1})\Big)\Big]. \qquad (7.1)
\end{aligned}
$$

The last equality holds since the minimization does not depend on the entire policy but only on $a_n = d_n(\mathbf{h}_n)$. Here, objective and constraint depend on the history of the process only through (x_n, s_n, t_n). Thus, given existence of a minimizing Markov decision rule d_n^*, (7.1) equals $\mathcal{T}_{nd_n^*} J_{n+1}(x_n, s_n, t_n)$. Again by the induction hypothesis, there exists an optimal Markov policy $\pi^* \in \mathbf{\Pi}^M$ such that $J_{n+1} = J_{n+1\pi^*}$. Hence, we have

$$V_n(\mathbf{h}_n) \geq \mathcal{T}_{nd_n^*} J_{n+1\pi^*}(x_n, s_n, t_n) = J_{n\pi^*}(x_n, s_n, t_n) \geq J_n(x_n, s_n, t_n) \geq V_n(\mathbf{h}_n).$$

It remains to show the existence of a minimizing Markov decision rule d_n^* and that $J_n \in \mathbb{M}$. We want to apply Proposition 2.4.3 of Bäuerle and Rieder (2011). The set-valued mapping $\mathbf{E} \ni (x, s, t) \mapsto D_n(x)$ is compact-valued and upper semicontinuous. Next, we show that $\mathbf{D}_n \ni (x, s, t, a) \mapsto L_n v(x, s, t, a)$ is lower semicontinuous for every $v \in \mathbb{M}$. Let $\{(x_k, s_k, t_k, a_k)\}_{k\in\mathbb{N}}$ be a convergent sequence in $\mathbf{D}_n$ with limit $(x^*, s^*, t^*, a^*) \in \mathbf{D}_n$. The mapping

$$\mathbf{D}_n \ni (x, s, t, a) \mapsto v\Big(T_n(x, a, Z_{n+1}(\omega)),\ s + tc_n(x, a, T_n(x, a, Z_{n+1}(\omega))),\ \beta t\Big)$$

is lower semicontinuous. Since $v \geq g \geq 0$, we can apply Fatou's Lemma which yields

$$\begin{aligned}
&\liminf_{k\to\infty} L_n v(x_k, s_k, t_k, a_k) \\
&= \liminf_{k\to\infty} \mathbb{E}\Big[v\Big(T_n(x_k, a_k, Z_{n+1}),\ s_k + t_k c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})),\ \beta t_k\Big)\Big] \\
&\geq \mathbb{E}\Big[\liminf_{k\to\infty} v\Big(T_n(x_k, a_k, Z_{n+1}),\ s_k + t_k c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})),\ \beta t_k\Big)\Big] \\
&\geq \mathbb{E}\Big[v\Big(T_n(x^*, a^*, Z_{n+1}),\ s^* + t^* c_n(x^*, a^*, T_n(x^*, a^*, Z_{n+1})),\ \beta t^*\Big)\Big] \\
&= L_n v(x^*, s^*, t^*, a^*).
\end{aligned}$$

I.e. $L_n v$ is lower semicontinuous. Proposition 2.4.3 in Bäuerle and Rieder (2011) implies the existence of a minimizing decision rule d_n^* and the lower semicontinuity of $\mathcal{T}_n v$.

Now fix $x \in E$. The fact that $(s, t) \mapsto \mathcal{T}_n v(x, s, t)$ is increasing follows as in Theorem 2.4.14 in Bäuerle and Rieder (2011). The inequality $\mathcal{T}_n v(x, s, t) \geq g(s),\ (x, s, t) \in \mathbf{E}$, is obvious. Altogether, we have $\mathcal{T}_n v \in \mathbb{M}$.

Part d) follows from the construction of the policies and the proof is complete. □

Proofs for subsection 4.3

Lemma 4.5 states that the outer optimization problem (3.4) can be reduced to the space $\mathcal{G}$.

Proof of Lemma 4.5 Set $C = C_N^{\sigma x}$ to simplify the notation and assume that $\rho_\phi(C) \le \bar{\rho}$. We know from Remark 2.7 that the optimal $g \in G$ corresponding to C is

$$g_{\phi,C}(x) = \int_0^1 F_C^{-1}(\alpha) + \frac{1}{1-\alpha}\left(x - F_C^{-1}(\alpha)\right)_+ \mu(\mathrm{d}\alpha), \qquad x \in \mathbb{R},$$

with μ from Proposition 2.5. Since $C \ge 0$ it follows

$$g_{\phi,C}(x) \ge \int_0^1 F_C^{-1}(\alpha)\mu(\mathrm{d}\alpha) \ge 0.$$

Furthermore, we have

$$\begin{aligned} g_{\phi,C}(x) &= \int_0^1 F_C^{-1}(\alpha)\mu(\mathrm{d}\alpha) + \int_0^1 \frac{1}{1-\alpha}\left(x - F_C^{-1}(\alpha)\right)^+ \mu(\mathrm{d}\alpha) \\ &\le \int_0^1 \mathrm{ES}_\alpha(C)\mu(\mathrm{d}\alpha) + x^+ \int_0^1 \frac{1}{1-\alpha}\mu(\mathrm{d}\alpha) \\ &= \rho_\phi(C) + \phi(1)x^+ \le \bar{\rho} + \phi(1)x^+ = \bar{g}(x). \end{aligned}$$

The first inequality uses $F_C^{-1}(\alpha) = \mathrm{VaR}_\alpha(C) \le \mathrm{ES}_\alpha(C)$ and $C \ge 0$. The identity

$$\int_0^1 \frac{1}{1-\alpha}\mu(\mathrm{d}\alpha) = \phi(1)$$

is by definition of μ. As a convex function, $g_{\phi,C}$ is almost everywhere differentiable with derivative $g'_{\phi,C}(x) = \phi(F_C(x)) \le \phi(1)$, cf. Remark 2.7. This establishes the Lipschitz continuity with constant $L = \phi(1)$. □

Next, we prove Theorem 4.6 which states the existence of a solution of the outer problem. For this, we need some preliminary results. As a first step we study the dependence of the value functions of the inner problem on g. In order to do so, we need some structure on $\mathcal{G}$.

Lemma 7.2 *$(\mathcal{G}, m)$ is a compact metric space, where*

$$m(g_1, g_2) = \sum_{j=1}^{\infty} 2^{-j} \frac{\max_{|s|\le j} |g_1(s) - g_2(s)|}{1 + \max_{|s|\le j} |g_1(s) - g_2(s)|}$$

is the metric of compact convergence.

Proof Since $\mathcal{G} \subseteq C(\mathbb{R}, \mathbb{R})$, it suffices to show that $\mathcal{G}$ is closed w.r.t. m and verify the assumptions of the Arzelà–Ascoli theorem. Note that convergence w.r.t. m implies pointwise convergence. Convexity, monotonicity, the common Lipschitz constant $\phi(1)$, non-negativity and the pointwise upper bound $\bar{g}$ are all preserved even under pointwise convergence. Hence, $\mathcal{G}$ is closed w.r.t. m. Moreover, $\mathcal{G}$ is pointwise bounded and the common Lipschitz constant implies that it is uniformly equicontinuous. □

For clarity we index the value functions with g. The value functions J_0^g of the finite horizon inner problem depend semicontinuously on g.

Lemma 7.3 *Let Assumption 3.1 be satisfied. Then the functional* $\mathcal{G}\times\mathbf{E} \ni (g,x,s,t) \mapsto J_n^g(x,s,t)$ *is lower semicontinuous for all* $n = 0,\dots,N$.

Proof The proof is by backward induction. At time N we have to verify that $J_N^g(x,s,t) = g(s+tc_N(x))$ is lower semicontinuous in (g,x,s,t). First, note that $\mathcal{G}\times\mathbb{R}_+ \ni (g,s) \mapsto g(s)$ is continuous since if $(g_k,s_k)\to(g,s)$, then g converges especially pointwise and

$$\begin{aligned}|g_k(s_k)-g(s)| &= |g_k(s_k)-g_k(s)+g_k(s)-g(s)| \le |g_k(s_k)-g_k(s)| + |g_k(s)-g(s)|\\ &\le \phi(1)\,|s_k-s| + |g_k(s)-g(s)| \to 0 \quad \text{as } k\to\infty.\end{aligned}$$

Now let $(g_k,x_k,s_k,t_k)\to(g,x,s,t)$ and define the increasing sequence $\{c_k\}_{k\in\mathbb{N}}$ through $c_k = \inf_{\ell\ge k} c_N(x_\ell)$.

Case 1: $\{c_k\}_{k\in\mathbb{N}}$ is bounded above and therefore convergent with limit $\hat{c}$. Then

$$\hat{c} = \lim_{k\to\infty} c_k = \lim_{k\to\infty}\inf_{\ell\ge k} c_N(x_\ell) = \liminf_{k\to\infty} c_N(x_k) \ge c_N(x)$$

since c_N is lower semicontinuous. As the functions $\{g_k\}_{k\in\mathbb{N}}$ and g are all increasing, we get

$$\liminf_{k\to\infty} g_k(s_k+t_kc_N(x_k)) \ge \lim_{k\to\infty} g_k(s_k+t_kc_k) = g(s+t\hat{c}) \ge g(s+tc_N(x)).$$

Case 2: $\{c_k\}_{k\in\mathbb{N}}$ is unbounded above. Then there exists $K\in\mathbb{N}$ such that $c_k \ge c_N(x)$ for all $k\ge K$ and

$$\liminf_{k\to\infty} g_k(s_k+t_kc_N(x_k)) \ge \liminf_{k\to\infty} g_k(s_k+t_kc_k) \ge \lim_{k\to\infty} g_k(s_k+t_kc_N(x)) = g(s+tc_N(x)).$$

Now assume the assertion holds for $n+1$. By Theorem 4.3 we have at time n

$$J_n^g(x,s,t) = \inf_{a\in D(x)} \mathbb{E}\Big[J_{n+1}^g\Big(T_n(x,a,Z_{n+1}),\ s+tc_n(x,a,T_n(x,a,Z_{n+1})),\ \beta t\Big)\Big].$$

The integrand $J_{n+1}^g\Big(T_n(x,a,Z_{n+1}(\omega)),\ s+tc_n(x,a,T_n(x,a,Z_{n+1}(\omega))),\ \beta t\Big)$ is lower semicontinuous in (g,x,s,t,a) for every $\omega\in\Omega$ by the induction hypothesis. Hence, if $(g_k,x_k,s_k,t_k)\to(g,x,s,t)$, Fatou's lemma and the monotonicity of expectation yield

$$\begin{aligned}&\liminf_{k\to\infty} \mathbb{E}\Big[J_{n+1}^{g_k}\Big(T_n(x_k,a_k,Z_{n+1}),\ s_k+t_kc_n(x_k,a_k,T_n(x_k,a_k,Z_{n+1})),\ \beta t_k\Big)\Big]\\ &\ge \mathbb{E}\Big[\liminf_{k\to\infty} J_{n+1}^{g_k}\Big(T_n(x,a,Z_{n+1}),\ s+tc_n(x,a,T_n(x,a,Z_{n+1})),\ \beta t\Big)\Big]\\ &\ge \mathbb{E}\Big[J_{n+1}^{g}\Big(T_n(x,a,Z_{n+1}),\ s+tc_n(x,a,T_n(x,a,Z_{n+1})),\ \beta t\Big)\Big]\end{aligned}$$

I.e. $(g, x, s, t) \mapsto L_n J_{n+1}^g(x, s, t, a)$ is lower semicontinuous. As the set-valued mapping $E \ni x \mapsto D(x)$ is compact valued and upper semicontinuous,

$$(g, x, s, t) \mapsto J_n^g(x, s, t, a) = \inf_{a \in D(x)} L_n J_{n+1}^g(x, s, t, a)$$

is lower semicontinuous by Proposition 2.4.3 in Bäuerle and Rieder (2011).

Now we are in a position to prove the existence result for the outer optimization problem.

Proof of Theorem 4.6 We want to apply Weierstraß' extreme value theorem. In view of Lemmata 7.2 and 7.3 it suffices to show that the functional $\mathcal{G} \ni g \mapsto \int_0^1 g^*(\phi(u))\mathrm{d}u$ is lower semicontinuous. Let $\{g_k\}_{k\in\mathbb{N}} \subseteq \mathcal{G}$ be a convergent sequence with limit $g \in \mathcal{G}$. It holds for all $u \in [0, 1]$

$$\begin{aligned}
\liminf_{k\to\infty} g_k^*(\phi(u)) &= \lim_{k\to\infty} \inf_{\ell\geq k} g_\ell^*(\phi(u)) = \lim_{k\to\infty} \inf_{\ell\geq k} \sup_{s\in\mathbb{R}} \left\{\phi(u)s - g_\ell(s)\right\} \\
&\geq \lim_{k\to\infty} \sup_{s\in\mathbb{R}} \inf_{\ell\geq k} \left\{\phi(u)s - g_\ell(s)\right\} = \sup_{s\in\mathbb{R}} \lim_{k\to\infty} \inf_{\ell\geq k} \left\{\phi(u)s - g_\ell(s)\right\} \\
&= \sup_{s\in\mathbb{R}} \left\{\phi(u)s - \limsup_{k\to\infty} g_k(s)\right\} = \sup_{s\in\mathbb{R}} \left\{\phi(u)s - g(s)\right\} = g^*(\phi(u)).
\end{aligned} \tag{7.2}$$

The inequality holds generally for the interchange of infimum and supremum, the equality thereafter by Lemma A.1.6 in Bäuerle and Rieder (2011) and the last but one equality since the sequence $\{g_k\}_{k\in\mathbb{N}}$ is especially pointwise convergent. Moreover note that for all $k \in \mathbb{N}$ and $u \in [0, 1]$ it holds

$$g_k^*(\phi(u)) = \sup_{s\in\mathbb{R}} \left\{\phi(u)s - g_k(s)\right\} \geq -g_k(0) \geq -\bar{g}(0) > -\infty.$$

Now, Fatou's lemma and (7.2) yield together with

$$\liminf_{k\to\infty} \int g_k^*(\phi(u))\mathrm{d}u \geq \int \liminf_{k\to\infty} g_k^*(\phi(u))\mathrm{d}u \geq \int g^*(\phi(u))\mathrm{d}u$$

the assertion. □

Next, we can turn to the numerical approximation scheme.

Proof of Lemma 4.8 a) Fix $\sigma \in \Pi$, $x \in E$ and set $C = C_N^{\sigma x}$ to simplify the notation. We know from Remark 2.7 that the optimal $g \in \mathcal{G}$ corresponding to C is

$$g_{\phi,C}(s) = \int_0^1 F_C^{-1}(\alpha) + \frac{1}{1-\alpha}\left(s - F_C^{-1}(\alpha)\right)_+ \mu(\mathrm{d}\alpha), \qquad s \in \mathbb{R},$$

with μ from Proposition 2.5. Clearly, it is sufficient to consider functions $g \in G$ which are optimal for at least one $C = C_N^{\pi x}$. Since $0 \leq C \leq \hat{c}$ we have $0 \leq$

$F_C^{-1}(\alpha) \le \hat{c}$. Consequently, it holds for $s < 0$

$$g_{\phi,C}(s) = \int_0^1 F_C^{-1}(\alpha)\mu(\mathrm{d}\alpha) = g(0).$$

As a convex function, $g_{\phi,C}$ is almost everywhere differentiable with derivative $g'_{\phi,C}(s) = \phi(F_C(s))$, cf. Remark 2.7, and for $s > \hat{c}$ it holds $F_C(s) = 1$.

b) Let $g \in \widehat{\mathcal{G}}$ and $y \in [0, \phi(1)]$. For $s \ge \hat{c}$ the function

$$s \mapsto sy - g(s) = (y - \phi(1))s - g(\hat{c}) + \phi(1)\hat{c}$$

is decreasing and for $s \le 0$ the function

$$s \mapsto sy - g(s) = sy - g(0)$$

is increasing. Hence, it suffices to consider the supremum over $[0, \hat{c}]$. □

Proof of Proposition 4.9 The first inequality is obvious and it remains to prove the second. We have for $N \in \mathbb{N} \cup \{\infty\}$, $x \in E$ and $g \in \widehat{\mathcal{G}}$

$$\begin{aligned} |J_0(p_m(g)) - J_0(g)| &= \left| \inf_{\sigma\in\Pi} \mathbb{E}[p_m(g)(C_N^{\sigma x})] - \inf_{\sigma\in\Pi} \mathbb{E}[g(C_N^{\sigma x})] \right| \\ &\le \sup_{\sigma\in\Pi} \mathbb{E}\left| p_m(g)(C_N^{\sigma x}) - g(C_N^{\sigma x}) \right| \\ &\le \sup_{s\in I} |p_m(g)(s) - g(s)|. \end{aligned}$$

Moreover, it holds for $y \in [0, \phi(1)]$

$$|p_m(g)^*(y) - g^*(y)| = \left| \sup_{s\in I}\{sy - g(s)\} - \sup_{s\in I}\{sy - p_m(g)(s)\} \right| \le \sup_{s\in I} |p_m(g)(s) - g(s)|.$$

Finally, the assertion follows with

$$\begin{aligned} |K_m(g) - K(g)| &\le |J_0(p_m(g)) - J_0(g)| + \int \left| p_m(g)^*(\phi(u)) - g^*(\phi(u)) \right| \mathrm{d}u \le 2 \sup_{s\in I} |p_m(g)(s) - g(s)| \\ &= 2 \max_{k=1,\dots,m-1} \max_{s\in[s_k,s_{k+1}]} \left| g(s) - g(s_k) - \frac{g(s_{k+1}) - g(s_k)}{s_{k+1} - s_k}(s - s_k) \right| \\ &\le 2 \max_{k=1,\dots,m-1} |g(s_{k+1}) - g(s_k)| \le 2\phi(1)\frac{\hat{c}}{m-1}. \end{aligned}$$

□

Proof of Lemma 4.10 By Lemma 4.8 b) we have $g_y^*(\xi) = \max_{s\in I} s\xi - g_y(s)$. Note that the slopes $c_k = \frac{y_{k+1}-y_k}{s_{k+1}-s_k}$, $k = 1, \dots, m-1$, are increasing. It follows

$$g_y^*(\xi) = \sup_{s\in[0,\hat{c}]} s\xi - g_y(s) = \max_{k=1,\dots,m-1} \max_{s\in[s_k,s_{k+1}]} s(\xi - c_k) - y_k + c_k s_k.$$

Let us distinguish three cases. Firstly, assume $\xi \in [c_\ell, c_{\ell+1}]$ for some $\ell \in \{1, \dots, m-2\}$. Then

$$\begin{aligned} g_y^*(\xi) &= \max\left\{ \max_{k=1,\dots,\ell} s_{k+1}(\xi - c_k) - y_k + c_k s_k,\ \max_{k=\ell+1,\dots,m-1} s_k(\xi - c_k) - y_k + c_k s_k \right\} \\ &= \max\left\{ \max_{k=1,\dots,\ell} s_{k+1}\xi - y_{k+1},\ \max_{k=\ell+1,\dots,m-1} s_k\xi - y_k \right\} \\ &= s_{\ell+1}\xi - y_{\ell+1}. \end{aligned}$$

The last equality holds, since $c_1 \le \dots \le c_{m-1}$ and $c_\ell \le \xi \le c_{\ell+1}$ is equivalent to $\xi s_\ell - y_\ell \le \xi s_{\ell+1} - y_{\ell+1} \ge \xi s_{\ell+2} - y_{\ell+2}$. Secondly, assume $\xi < c_1$. Then

$$g_y^*(\xi) = \max_{k=1,\dots,m-1} s_k(\xi - c_k) - y_k + c_k s_k = \max_{k=1,\dots,m-1} s_k\xi - y_k = s_1\xi - y_1 = -y_1.$$

Again, $\xi < c_1$ is equivalent to $\xi s_2 - y_2 < \xi s_1 - y_1$. Since $c_1 \le \dots \le c_{m-1}$, this implies the last equality. The third case $c_{m-1} < \xi$ is analogous. □

Proofs for section 5

We start with the proof of the solution algorithm for the model with infinite planning horizon. Since the model with infinite planning horizon is derived as a limit of the one with finite horizon, the consideration can be restricted to Markov policies $\pi = (d_1, d_2, \dots) \in \mathbf{\Pi}^M$ due to Theorem 4.3. When calculating limits, it is more convenient to index the value functions with the distance to the time horizon rather than the point in time. This is also referred to as *forward form* of the value iteration and is only possible under Markov policies in a stationary model. There, the two ways of indexing are equivalent. The value of a policy $\pi = (d_0, d_1 \dots) \in \mathbf{\Pi}^M$ up to a planning horizon $N \in \mathbb{N}$ is

$$\begin{aligned} J_{0\pi}(x, s, t) &= g(s) \\ J_{N\pi}(x, s, t) &= \mathbb{E}_{0x}\left[g\left(s + t \sum_{k=0}^{N-1} \beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi) \right) \right]. \end{aligned}$$

The change of indexing makes it necessary to write the value iteration in terms of the *shifted policy* $\overrightarrow{\pi} = (d_1, d_2, \dots)$ corresponding to $\pi = (d_0, d_1, \dots) \in \mathbf{\Pi}^M$:

$$\begin{aligned} J_{N\pi}(x, s, t) &= E\Big[J_{N-1\overrightarrow{\pi}}\Big(T(x, d_0(x, s, t), Z),\ s + tc(x, d_0(x, s, t), T(x, d_0(x, s, t), Z)),\ \beta t \Big) \Big] \\ &= \mathcal{T}_{d_0} J_{N-1\overrightarrow{\pi}}(x), \qquad (x, s, t) \in \mathbf{E}. \end{aligned} \tag{7.3}$$

The value function for finite planning horizon $N \in \mathbb{N}$ is given by

$$J_N(x, s, t) = \inf_{\pi \in \mathbf{\Pi}^M} J_{N\pi}(x, s, t), \qquad (x, s, t) \in \mathbf{E},$$

and satisfies due to Theorem 4.3 the Bellman equation

$$J_N(x,s,t) = \mathcal{T}J_{N-1}(x,s,t) = \mathcal{T}^N g(x,s,t), \qquad (x,s,t) \in \mathbf{E}.$$

The value of a policy $\pi \in \mathbf{\Pi}^M$ under an infinite planning horizon is then for $(x,s,t) \in \mathbf{E}$

$$J_{\infty\pi}(x,s,t) := \lim_{N\to\infty} J_{N\pi}(x,s,t) = \mathbb{E}_{0x}\left[g\left(s + t\sum_{k=0}^{\infty} \beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right)\right].$$

The second equality holds by monotone convergence and the continuous mapping theorem. The infinite horizon value function is

$$J_\infty(x,s,t) := \inf_{\pi\in\Pi^M} J_{\infty\pi}(x,s,t), \qquad (x,s,t) \in \mathbf{E}, \tag{7.4}$$

and the limit value function is

$$J(x,s,t) := \lim_{N\to\infty} J_N(x,s,t), \qquad (x,s,t) \in \mathbf{E},$$

which again exists since J_N is increasing. Note that $\mathbb{M}$ is closed under pointwise convergence and hence $J \in \mathbb{M}$. Having introduced these notions we can now turn to the proof.

Proof of Theorem 5.1 a) First, we show that $J_\infty = J$. For all $N \in \mathbb{N}$ we have $J_{N\pi} \geq J_N$. Taking the limit $N \to \infty$ we obtain $J_{\infty\pi} \geq J$ for policies $\pi \in \mathbf{\Pi}^M$. Thus $J_\infty \geq J$.

For the reverse inequality we start with $J_{N\pi} \leq J_{\infty\pi}$ which is true for all policies $\pi \in \mathbf{\Pi}^M$ due to the fact that $c \geq 0$. Taking the infimum over all policies yields $J_N \leq J_\infty$ and taking the limit $N \to \infty$ we obtain $J \leq J_\infty$. It total, we have $J = J_\infty$.

That J is a fixed point of $\mathcal{T}$ follows from Theorem A.1.5 in (Bäuerle and Rieder 2011) in case $J(x,s,t) < \infty$. The case $J(x,s,t) = \infty$ follows directly.

Let now $v \in \mathbb{M}$ be another fixed point of $\mathcal{T}$, i.e. $v = \mathcal{T}v$. Iterating this equality yields $v = \mathcal{T}^n v$ for all $n \in \mathbb{N}$. Since $v \in \mathbb{M}$ we have $v \geq g$ and because of the monotonicity of the Bellman operator we get $v = \mathcal{T}^n v \geq \mathcal{T}^n g$. Letting $n \to \infty$ finally implies $v \geq J = J_\infty$, thus J_∞ is the smallest fixed point of the Bellman operator.

b) Since $J_\infty \in \mathbb{M}$, the existence of a minimizing Markov decision rule follows as in the proof of Theorem 4.3. Furthermore, it holds $J_\infty(x,s,t) \geq g(s)$, $(x,s,t) \in \mathbf{E}$, since $J_\infty \in \mathbb{M}$. Consequently, we have

$$J_\infty = \lim_{N\to\infty} \mathcal{T}_{d^*}^N J_\infty \geq \lim_{N\to\infty} \mathcal{T}_{d^*}^N g = \lim_{N\to\infty} J_{N\pi^*} = J_{\infty\pi^*} \geq J_\infty.$$

i.e. π^* is optimal. The first equality is by part a), the inequality thereafter by the monotonicity of the operator $\mathcal{T}_{d^*}$ and the second equality by the value iteration (7.3).

c) The last part follows from the construction of the policies.

□

For the proof of the existence of a solution to the outer optimization problem we have to show the lower semicontinuity of the infinite horizon value functions:

Lemma 7.4 *Let Assumption 3.1 be satisfied. Then the functional* $\mathcal{G}\times\mathbf{E} \ni (g, x, s, t) \mapsto J_\infty^g(x, s, t)$ *is lower semicontinuous for all* $(x, s, t) \in \mathbf{E}$.

Proof The value functions J_N^g are lower semicontinuous in (g, x, s, t) by Lemma 7.3. Note that the induction basis holds especially for $c_N \equiv 0$. Since $J_N^g \uparrow J_\infty^g$ as $N \to \infty$, the assertion follows from Lemma A.1.4 in Bäuerle and Rieder (2011).

The proof of Lemma 5.3 follows exactly the same lines as in the finite horizon case. Finally we state the proofs of the remaining two results in Sect. 5.2.

Proof of Proposition 5.5 In Theorem 4.3, the continuity of T is used to show that $\mathbf{D} \ni (x, s, t, a) \mapsto Lv(x, s, t, a)$ is lower semicontinuous for every $v \in \mathbb{M}$. Due to the monotonicity assumptions, the mapping

$$\mathbf{D} \ni (x, s, t, a) \mapsto v\Big(T(x, a, Z(\omega)),\ s + tc(x, a, T(x, a, Z(\omega))),\ \beta t\Big)$$

is lower semicontinuous for every $\omega \in \Omega$ as a composition of an increasing lower semicontinuous function with a lower semicontinuous one. Now, the lower semicontinuity of $\mathbf{D} \ni (x, s, t, a) \mapsto Lv(x, s, t, a)$ and the existence of a minimizing decision rule follow as in the proof of Theorem 4.3. The fact that $\mathcal{T}v$ is increasing for every $v \in \mathbb{M}$ follows as in Theorem 2.4.14 in Bäuerle and Rieder (2011). In Theorem 5.1, the continuity of T is only used indirectly through Theorem 4.3. Note that $J_\infty \in \mathbb{M}$ since the pointwise limit of increasing functions remains increasing. □

Proof of Theorem 5.6 We prove by induction that J_n is convex in (x, s) for $n \in \mathbb{N}_0$. Then J_∞ is convex as a pointwise limit of convex functions. For $n = 0$ we know that $J_0(x, s, t) = g(s)$ is convex in (x, s). Now assume that J_n is convex in (x, s). Recall that J_n increasing by Proposition 5.5. Hence, for every $\omega \in \Omega$ and $t > 0$ the function

$$(x, s, a) \mapsto J_n\Big(T(x, a, Z(\omega)),\ s + tc(x, a, T(x, a, Z(\omega))),\ \beta t\Big)$$

is convex as a composition of an increasing convex with a convex function. By the linearity of expectation $(x, s, a) \mapsto LJ_n(x, s, t, a)$ is convex, too, for every $t > 0$. Now, the convexity of J_{n+1} follows from Proposition 2.4.18 in Bäuerle and Rieder (2011). □

Proofs for section 6

Proof of Lemma 6.1 a) Let $\ell \in \mathcal{L}$. Since $\mathrm{id}_{\mathbb{R}_+} - \ell$ is increasing, it holds for $0 \le x \le y$ that $x - \ell(x) \le y - \ell(y)$. Rearranging and using that ℓ is increasing, too, yields with $|\ell(x) - \ell(y)| = \ell(y) - \ell(x) \le y - x = |x - y|$ the Lipschitz continuity with common constant 1. Moreover, $\mathcal{L}$ is pointwise bounded by $\mathrm{id}_{\mathbb{R}_+}$ and closed under pointwise convergence. Hence, $(\mathcal{L}, m)$ is a compact metric space by the Arzelà-Ascoli theorem and as such also complete and separable, i.e. a Borel space.

b) Let $\{\ell_k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{L}$ such that $\ell_k \to \ell \in \mathcal{L}$ for $k \to \infty$. Especially, it holds $\ell_k(y) \to \ell(y)$ for all $y \in \mathbb{R}_+$ and $Y - \ell_k(Y) \to Y - \ell(Y)$ $\mathbb{P}$-a.s. Since $Y - \ell_k(Y) \le Y \in L^1$ for all $k \in \mathbb{N}$, the Fatou property of π_R implies

$$\liminf_{k\to\infty} \pi_R(\ell_k) = \liminf_{k\to\infty} \pi_R\big(Y - \ell_k(Y)\big) \ge \pi_R\big(Y - \ell(Y)\big) = \pi_R(\ell).$$

c) We show that the mapping $\mathcal{L} \times \mathbb{R}_+ \ni (\ell, y) \mapsto \ell(y)$ is continuous. Then, the transition function T is upper semicontinuous as a sum of upper semicontinuous functions due to part b). Let $\{(\ell_k, y_k)\}_{k\in\mathbb{N}}$ be a convergent sequence in $\mathcal{L}\times\mathbb{R}_+$ with limit (ℓ, y). Since convergence w.r.t. the metric m implies pointwise convergence and all ℓ_k have the Lipschitz constant $L = 1$, it follows

$$|\ell_k(y_k) - \ell(y)| \le |\ell_k(y_k) - \ell_k(y)| + |\ell_k(y) - \ell(y)| \le |y_k - y| + |\ell_k(y) - \ell(y)| \to 0.$$

The fact that T is increasing in x is obvious.

d) Due to a), we only have to consider the budget-constrained case. Since $\mathcal{L}$ is compact it suffices to show that $D(x) = \{\ell \in \mathcal{L} : \pi_R(\ell) \le x^+\}$ is closed. This is the case since $D(x)$ is a sublevel set of the lower semicontinuous function $\pi_R : \mathcal{L} \to \mathbb{R}_+$, cf. Lemma A.1.3 in Bäuerle and Rieder (2011). Furthermore, we show that D is closed to obtain the upper semicontinuity from Lemma A.2.2 in Bäuerle and Rieder (2011). From the lower semicontinuity of π_R it follows that the epigraph

$$\mathrm{epi}(\pi_R) = \{(x, \ell) \in \mathbb{R}_+ \times \mathcal{L} : \pi_R(\ell) \le x\}$$

is closed. Thus, $D = \mathrm{epi}(\pi_R) \cup (\mathbb{R}_- \times D(0))$ is closed, too. That $x \mapsto D(x)$ is increasing is clear.

e) The one-stage cost $c(x, \ell, T(x, \ell, y, z)) = x - T(x, \ell, y, z) = \ell(y) + \pi_R(\ell) - z$ is lower semicontinuous in (x, ℓ) as a sum of lower semicontinuous functions and decreasing in x since it does not depend on x.

□

References

Acerbi C (2002) Spectral measures of risk: a coherent representation of subjective risk aversion. J Bank Finance 26(7):1505–1518

Albrecher H, Beirlant J, Teugels JL (2017) Reinsurance: actuarial and statistical aspects. Wiley, Hoboken

Arrow K (1963) Uncertainty and the welfare economics of medical care. Am Econ Rev 53(5):941–973

Bäuerle N, Glauner A (2018) Optimal risk allocation in reinsurance networks. Insure Math Econ 82:37–47
Bäuerle N, Glauner A (2021) Markov decision processes with recursive risk measures. Eur J Oper Res. https://doi.org/10.1016/j.ejor.2021.04.030
Bäuerle N, Ott J (2011) Markov decision processes with average-value-at-risk criteria. Math Methods Oper Res 74(3):361–379
Bäuerle N, Rieder U (2011) Markov decision processes with applications to finance. Springer, Berlin, Heidelberg
Bäuerle N, Rieder U (2014) More risk-sensitive Markov decision processes. Math Oper Res 39(1):105–120
Borch K (1960) An attempt to determine the optimum amount of stop loss reinsurance. Trans XVI Int Congr Actuar I:597–610
Chi Y, Tan KS (2013) Optimal reinsurance with general premium principles. Insur Math Econ 52(2):180–189
Chow Y, Ghavamzadeh M (2014) Algorithms for CVaR optimization in MDPs. Adv Neural Inf Process Syst 3509–3517
Chow Y, Tamar A, Mannor S, Pavone M (2015) Risk-sensitive and robust decision-making: a CVaR optimization approach. Adv Neural Inf Process Syst 1522–1530
Chu S, Zhang Y (2014) Markov decision processes with iterated coherent risk measures. Int J Control 87(11):2286–2293
Corrias L (1996) Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws. SIAM J Numer Anal 33(4):1534–1558
Cui W, Yang J, Wu L (2013) Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles. Insur Math Econ 53(1):74–85
Dhaene J, Wang S, Young V, Goovaerts MJ (2000) Comonotonicity and maximal stop-loss premiums. Bull Swiss Assoc Actuar 2000(2):99–113
Glauner A (2020) Robust and risk-sensitive Markov decision processes with applications to dynamic optimal reinsurance. PhD thesis, Karlsruhe Institute of Technology. https://doi.org/10.5445/IR/1000126170
Hinderer K (1970) Foundations of non-stationary dynamic programming with discrete time parameter. Springer, Berlin, Heidelberg
Howard RA, Matheson JE (1972) Risk-sensitive Markov decision processes. Manage Sci 18(7):356–369
Kusuoka S (2001) On law invariant coherent risk measures. In: Advances in mathematical economics. Springer, pp 83–95
Li X, Zhong H, Brandeau ML (2017) Quantile Markov decision process. arXiv:1711.05788
Lo A (2017) A Neyman–Pearson perspective on optimal reinsurance with constraints. ASTIN Bull 47(2):467–499
McNeil AJ, Frey R, Embrechts P (2015) Quantitative risk management: concepts. Princeton University Press, revised edition, Techniques and Tools
Müller A (2007) Certainty equivalents as risk measures. Braz J Probab Stat 21(1):1–12
Müller A, Stoyan D (2002) Comparison methods for stochastic models and risks. Wiley, Chichester
Pflug GC, Pichler A (2016) Time-consistent decisions and temporal decomposition of coherent risk functionals. Math Oper Res 41(2):682–699
Pichler A (2015) Premiums and reserves, adjusted by distortions. Scand Actuar J 2015(4):332–351
Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. J Risk 2(3):21–41
Ruszczyński A (2010) Risk-averse dynamic programming for Markov decision processes. Math Program 125(2):235–261
Shapiro A (2009) On a time consistency concept in risk averse multistage stochastic programming. Oper Res Lett 37(3):143–147
Shapiro A (2013) On Kusuoka representation of law invariant risk measures. Math Oper Res 38(1):142–152
Shapiro A, Uğurlu K (2016) Decomposability and time consistency of risk averse multistage programs. Oper Res Lett 44(5):663–665
Tamar A, Chow Y, Ghavamzadeh M, Mannor S (2015) Policy gradient for coherent risk measures. Adva Neural Inf Process Syst 1468–1476
Tsanakas A, Desli E (2003) Risk measures and theories of choice. Br Actuar J 9(4):959–991
Uğurlu K (2017) Controlled Markov decision processes with AVaR criteria for unbounded costs. J Comput Appl Math 319:24–37