Robust and Risk-Sensitive Markov Decision Processes with Applications to Dynamic Optimal Reinsurance

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von

Alexander Harald Glauner, M.Sc.
aus Neuenbürg

Tag der mündlichen Prüfung: 11. November 2020

Referentin: Prof. Dr. Nicole Bäuerle
Korreferentin: Prof. Dr. Anna Jaśkiewicz

To my father

Acknowledgments

This thesis has been written during my employment as a research associate at the Institute of Stochastics at Karlsruhe Institute of Technology (KIT). At this point, I would like to take the opportunity to express my gratitude to several people who supported me in the process.

First and foremost, I want to thank my advisor Prof. Dr. Nicole Bäuerle for her excellent guidance over the past three years. She took a lot of time for our discussions, was constantly available for questions, made numerous helpful suggestions, and always asked the right questions. My special thanks go to Prof. Dr. Anna Jaśkiewicz for acting as second referee.

I feel fortunate to have worked in such a pleasant atmosphere at the Institute of Stochastics. For that, I am very grateful to my present and former colleagues, many of whom have become friends.

Furthermore, I am deeply indebted to my family and especially my parents for their loving support and encouragement. Last but certainly not least, I would like to thank Carina for making me so happy every day.

Prior Publications

Parts of Chapter 4 are direct quotes from the prior publication

> Bäuerle, N. and Glauner, A. (2020). Distributionally robust Markov decision processes and their connection to risk measures. arXiv:2007.13103.

Contents

CHAPTER 1

Introduction

The pioneering monographs of Bellman (1957) and Howard (1960) mark the beginning of highly active research on *Markov Decision Processes (MDP)* and their applications for more than half a century. The term itself goes back to Bellman (1954) while the formalization in the present form was introduced by Blackwell (1965). Further steps towards more general models are represented by the books of Hinderer (1970) and Bertsekas and Shreve (1978). Their common optimality criterion of minimizing the expected total cost became standard in the literature and is a suitable choice in many applications. Moreover, the tower property of conditional expectation is a key feature enabling the application of dynamic programming techniques.

But there are also circumstances in which the use of this optimality criterion is either not possible or not appropriate. Two of them are addressed in this thesis. Firstly, we consider the case that the transition law of the decision process is not fully known. In the literature this is referred to as *ambiguity* whereas uncertainty relates to random quantities with known distribution. Secondly, we study risk-averse decision-makers who are willing to accept a higher expected cost in order to reduce the risk of an extremely adverse outcome. Such preferences are referred to as *risk-sensitive*. It will turn out that in some cases ambiguity and risk-sensitivity lead to the same optimal decision. Our guiding example is a model for dynamic reinsurance in discrete time.

1.1. Literature Overview

In many applications, the transition law of a Markov Decision Process is subject to misspecification since it is either based on expert opinion or estimation from historical data. Consequently, the controller may optimize with respect to a transition law deviating from the true one. Applying the so-obtained optimal policies in the real system may lead to a significant degeneration of performance. One way of dealing with an ambiguous transition law is the *robust approach*, where the controller selects a policy which is optimal with respect to the most adverse transition law of a respective family in each scenario. Satia and Lave (1973) as well as White and Eldeib (1994) have studied this approach for MDP with finite state and action spaces assuming that the rows of the unknown transition matrix lie in prespecified polytopes. Later on, Iyengar (2005) developed a solution theory for distributionally robust MDP with countable state and action spaces under general constraints on the transition kernels. Contemporaneously, Nilim and El Ghaoui (2005) reached similar findings, however, limited to finite state and action spaces. In both works, a rectangularity condition on the respective sets of probability measures turned out to be a key assumption for deriving a Bellman equation. This property and possibilities to weaken it were further investigated by Wiesemann et al. (2012) and Shapiro (2016).

The robust approach can also be interpreted as a dynamic zero-sum Stackelberg game with nature as the controllers opponent selecting the transition law in each scenario. This perspective provides more clarity about based on what information the transition law is selected and turns out to be helpful when dealing with measurability issues in more general settings with Borel state and action spaces. Such a dynamic game set-up, where the topology of convergence in distribution is used for the space of probability measures, can be found in González-Trejo et al. (2002), however lacking the rigorous derivation of a Bellman equation. Their results are complemented by Jaśkiewicz and Nowak (2011, 2014).

A seemingly different problem is to incorporate risk-sensitive preferences of the controller into the decision model. While minimizing the expected cost implies a risk-neutral attitude, empirical evidence suggests that many agents tend to be risk-averse or are even forced to be so by regulators, e.g. in the finance or insurance industry. The study of so-called *risk-sensitive Markov Decision Processes* was pioneered by Howard and Matheson (1972), who replaced expectation by the certainty equivalent of an exponential utility in a decision model with finite state and action space. In the sequel, the study of similar optimality criteria was extensively pursued in the literature. A comprehensive treatment with general utilities and Borel state and action spaces can be found in the paper by Bäuerle and Rieder (2014). The exponential certainty equivalent is also known as entropic risk measure. Especially with regard to financial and insurance applications it is of interest to replace it by other monetary risk measures. Exemplarily, Bäuerle and Ott (2011) consider the problem with Expected Shortfall. The main difficulty of maximizing a certainty equivalent or monetary risk measure is to obtain a value iteration since these functionals do not have a tower property like conditional expectation. A solution dating back to Kreps (1977a,b) and since then frequently applied in the literature is to extend the state space and introduce summary variables.

Other risk-sensitive optimality criteria avoid this issue. Since the 2000s, dynamic risk measures, which typically have some sort of tower property, were increasingly studied in the literature. For an overview see Föllmer and Schied (2016). In order to apply the concept to MDP and preserve Markovian value functions, Ruszczyński (2010) constructed the subclass of so-called Markov risk measures. A different approach was taken by Bäuerle and Jaśkiewicz (2017, 2018) and Asienkiewicz and Jaśkiewicz (2017), who recursively apply a static risk measure, namely the entropic risk measure, at each stage. Here, a value iteration holds by construction. While the recursive procedure induces sensible decisions at each stage, it lacks a global interpretation of the objective function. The approach is motivated by the economic literature, where the representation of preferences by recursive utility functions (here exponential ones) has been widely studied with notable contributions by Kreps and Porteus (1978) and Epstein and Zin (1989). For an overview see Chapter 20 in Miao (2014). The key feature of recursive utilities is that they allow separating intertemporal preferences from risk aversion. The risk-sensitive recursive approach turns out to induce the same optimal policy as the robust approach in many cases. Through this connection, a global interpretation of the recursively defined objective function as a risk measure can be obtained. Osogami (2012) and Shapiro (2012) outlined this connection exemplarily in stylized settings.

Research on the static counterpart of our actuarial application dates back to the 1960s. The objective is to minimize an insurance company's cost of capital or capital requirement for the retained loss including the cost of reinsurance. The capital requirement is determined by a risk measure applied to the effective risk after reinsurance and the cost of capital is given as a cost of capital rate times the capital requirement. Borch (1960) proved that a stop-loss reinsurance treaty minimizes the variance of the retained loss of the insurer given the reinsurance premium is calculated with the expected value principle. A similar result has been derived in Arrow (1963) where the insurer's expected utility of terminal wealth 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). Due to developments in the regulatory framework like Solvency II, the risk measures Value-at-Risk and Expected Shortfall are of special interest since the 2000s. Cai and Tan (2007) optimized the retention levels of stop-loss contracts for these risk measures under the expected premium principle. Later on, Chi and Tan (2013) identified layer reinsurance contracts as optimal within a large nonparametric class of treaties 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) for quite general results, or multidimensional settings induced by a macroeconomic perspective, see Bäuerle and Glauner (2018).

For dynamic extensions of the optimal reinsurance problem, it is necessary to model the development of the insurer's surplus over time. Until now, such problems were almost exclusively studied in continuous time. A very popular optimality criterion is to maximize the expected total dividend payments to the insurance company's shareholders. Albrecher and Thonhauser (2009) provide a good overview of the relevant literature. The only treatment of this problem in discrete time we are aware of is Chen and Assa (2019). Cost of capital minimization has so far not been studied in a dynamic setting.

1.2. Outline of the Thesis

This thesis is structured as follows. In Chapter 2, we recall some important results about risk measures and with regard to our application also about the related concept of premium principles. We mainly focus on the large class of distortion risk measures and its subclass of spectral risk measures. The most widely used risk measures in practice, Value-at-Risk and Expected Shortfall, belong to this class. Of particular interest for our purposes are continuity properties (Section 2.2) and dual representations (Section 2.3).

In Chapter 3, we introduce the Markov Decision Model which we are working with throughout. It has Borel state and action spaces and allows for unbounded cost functions. Since it is more convenient in our setting, we are using a functional representation for the dynamics of the state process. The general continuity and compactness properties with variants for special cases of the model are stated. In Section 3.2, a dynamic reinsurance model in discrete time is introduced in two versions. One with a focus on cost of capital and the other one with dividend payments. The latter one is similar to the model in Chen and Assa (2019). Since reinsurance treaties are typically written for one year (Albrecher et al.; 2017, p. 1) and dividends are paid annually, modeling in discrete time is appropriate when focusing on the management of the insurer's surplus by means of reinsurance and dividend payments while neglecting the possible use of capital market instruments.

Chapter 4 treats robust minimization of the expected total cost with ambiguity concerning the distribution of the disturbances generalizing the results of Iyengar (2005) to a model with Borel spaces and unbounded cost function. In order to deal with the arising measurability issues, we borrow from the dynamic game setup in González-Trejo et al. (2002) and Jaśkiewicz and Nowak (2011). The major difference of our contribution compared to these two papers is the design of the distributional ambiguity where we replace the topology of convergence in distribution on the ambiguity sets by the weak* topology $\sigma(L^q, L^p)$. Our formulation leads to a Stackelberg game against nature. Under suitable integrability assumptions and a finite planning horizon, we derive a robust cost iteration for a fixed policy of the decision-maker and a Bellman equation for the robust optimization problem. Moreover, we show the existence of optimal deterministic policies for both players. This is in contrast to classical zero-sum games where one usually obtains randomized optimal policies. The results are then extended to an infinite planning horizon. In Section 4.3, we study the special case that the state space is the real line, which allows us to introduce monotonicity properties for the model data and weaken the continuity assumption. Under additional convexity assumptions, we show that it is possible to interchange infimum and supremum in the Bellman equation and outline the game-theoretical implications. Finally, we discuss special choices for the ambiguity sets which have computational advantages and where the robust optimization problem coincides with the minimization of a coherent risk measure. As applications, we consider a robust LQ problem and robust maximization of the expected total dividend payment for an insurance company. Here, we prove the existence of an optimal dividend and reinsurance policy. For the setting of Chen and Assa (2019), whose proof contains a fundamental error, the existence is verified as a special case.

In the subsequent Chapter 5, we study risk-sensitive recursive cost minimization in our decision model. Under a finite planning horizon and some integrability assumptions, we extend the findings of Bäuerle and Jaśkiewicz (2017, 2018) and Asienkiewicz and Jaśkiewicz (2017) to general law-invariant monetary risk measures with the Fatou property. A corresponding Bellman equation is derived and the existence of Markovian optimal policies ensured. For infinite planning horizons, we additionally have to require coherence to obtain a contracting model and can use the weaker initial assumptions only in special cases. As in the previous chapter, the real line as state space allows us to introduce monotonicity properties for the model data and weaken the continuity assumption. In Section 5.3, we discuss connections to the distributionally robust cost minimization. Under some technical assumptions, the two optimality criteria are indeed equivalent. Thus, we obtain a global interpretation of the recursively defined objective functions. The comparison is more general than in Osogami (2012) and Shapiro (2012). Especially, we find that the corresponding global (or composite) risk measure depends on the controller's policy apart from special cases. As an application, we study the cost of capital minimization of an insurance company in discrete time closing a gap in the actuarial literature. We ensure the existence of an optimal reinsurance policy under general conditions and determine it explicitly for Value-at-Risk as risk measure. Here, the optimal reinsurance treaties have a one-layer form.

The final Chapter 6 treats the minimization of a spectral risk measure applied to the total cost. This can be seen as a reverse approach to the minimization of a recursively applied spectral risk measure, which is in some cases equivalent to the minimization of a non-standard risk measure applied to the total cost, cf. Section 4.3.2. We adopt the approach of Bäuerle and Ott (2011) to separate the minimization in an outer and inner problem and extend their findings to general spectral risk measures with bounded spectrum and unbounded above costs. The inner optimization problem is solved as an ordinary MDP on an extended state space under both finite and infinite horizon given some integrability assumptions. The real line as state space allows again to introduce monotonicity properties for the model data and weaken the continuity assumption. For spectral risk measures, the outer optimization problem becomes infinite dimensional. We ensure existence in the general setting. We also discuss an algorithmic approximation for bounded cost functions and prove its convergences. As an application, we introduce an alternative dynamic extension in discrete time of the static cost of capital minimization problem for an insurance company. The existence of an optimal reinsurance policy is proven under general conditions. For the expected premium principle we show that it is optimal to choose stop-loss contracts.

CHAPTER 2

Risk Measures and Premium Principles

Let an atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a real number $p \in [1, \infty)$ be fixed. With $q \in (1, \infty]$ we denote the conjugate index satisfying $\frac{1}{p} + \frac{1}{q} = 1$ under the convention $\frac{1}{\infty} = 0$. Henceforth, $L^p = L^p(\Omega, \mathcal{A}, \mathbb{P})$ denotes the vector space of real-valued random variables thereon which have an integrable p-th moment. L^p_+ is the subset 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. With $\mathbb{R}_+$ we denote the non-negative real numbers.

A *risk measure* is a functional $\rho : L^p \to \bar{\mathbb{R}}$. The notion of a *premium principle* $\pi : L^p_+ \to \bar{\mathbb{R}}$ is mathematically closely related but the applications are different. While the former determines the necessary solvency capital to bear a risk, the latter gives the price of (re)insuring it. In contrast to general financial risks, insurance risks are typically non-negative. Hence, it suffices to consider premium principles on L^p_+. The properties of risk measures discussed in the sequel apply to premium principles analogously.

Definition 2.1. A risk measure $\rho : L^p \to \bar{\mathbb{R}}$ is called

a) *law-invariant* if $\rho(X) = \rho(Y)$ for random variables X, Y with 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 $m \in \mathbb{R}$.

d) *positive homogeneous* if $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda \in \mathbb{R}_+$.

e) *normalized* if $\rho(0) = 0$.

f) *finite* if $\rho(L^p) \subseteq \mathbb{R}$.

g) *comonotonic additive* if $\rho(X+Y) = \rho(X)+\rho(Y)$ for all comonotonic random variables X, Y.

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

i) *convex* if $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for $\lambda \in [0,1]$.

Throughout, we will only consider law-invariant risk measures and premium principles. A risk measure is called *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. However, the attribute monetary is rather unusual for premium principles since most of them are monotone but often not translation invariant. Monetary risk measures which are additionally positive homogeneous and subadditive are referred to as *coherent*. Further, note that

- given law invariance, monotonicity is equivalent to preservation of the usual stochastic order.
- due to translation invariance, assuming normalization is no structural restriction for a monetary risk measure.
- positive homogeneity implies normalization.
- given positive homogeneity, convexity and subadditivity are equivalent.

The next Lemma derives another property from the axioms discussed above.

Lemma 2.2 (Pichler; 2013, Prop. 6)**.** *A coherent risk measure ρ satisfies the triangular inequality*

$$|\rho(X) - \rho(Y)| \leq \rho(|X - Y|).$$

Proof. Using subadditivity and monotonicity one obatains

$$\rho(X) = \rho\left(Y + (X - Y)\right) \leq \rho(Y) + \rho(X - Y) \leq \rho(Y) + \rho(|X - Y|).$$

Consequently, it holds $\rho(X) - \rho(Y) \leq \rho(|X - Y|)$. Interchanging the roles of X and Y yields the assertion. □

2.1. Distortion Risk Measures

Many established risk measures and premium principles belong to the large class of distortion risk measures. This class is based on the well-known representation of the expectation of $X \in L^p$

$$E[X] = \int_0^\infty S_X(x)\,\mathrm{d}\,x - \int_{-\infty}^0 1 - S_X(x)\,\mathrm{d}\,x,$$

where $S_X(x) = 1 - F_X(x) = \mathbb{P}(X > x),\ x \in \mathbb{R}$, denotes the survival function of X.

Definition 2.3. a) An increasing function $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$ is called *distortion function.*

b) For a distortion function g, the function

$$g(S_X) : \mathbb{R} \to [0,1], \quad x \mapsto g(S_X(x))$$

is called *distorted survival function*.

c) The *distortion risk measure* w.r.t. a distortion function g is defined by $\rho_g : L^p \to \bar{\mathbb{R}}$,

$$\rho_g(X) = \int_0^\infty g(S_X(x))\,\mathrm{d}\,x - \int_{-\infty}^0 1 - g(S_X(x))\,\mathrm{d}\,x$$

whenever at least one of the integrals is finite.

d) The *Wang premium principle* w.r.t. a distortion function g is defined by $\pi_g : L^p_+ \to \bar{\mathbb{R}}$,

$$\pi_g(X) = (1+\theta) \int_0^\infty g(S_X(x))\,\mathrm{d}\,x, \quad \theta \geq 0.$$

Note that for left-continuous g the distorted survival function is itself a survival function. While expectation is usually not regarded as an appropriate risk measure since it does not distinguish between gain and loss, ρ_g and π_g can outweigh this by an appropriate distortion function. For simplicity the following discussion is in terms of distortion risk measures. Mutatis mutandis the results apply to Wang premium principles as well.

There is an alternative representation of distortion risk measures in terms of Lebesgue-Stieltjes integrals based on the quantile function $F_X^{-1}(u) = \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$, $u \in (0,1)$ in lieu of the survival function. Following the convention in Klenke (2014) we consider the Lebesgue-Stieltjes integral for right-continuous integrators and on half-open intervals of the form $(a,b]$ for real numbers $a \leq b$. I.e.

$$\int_a^b f(x)\,\mathrm{d}\,g(x) = \int_{(a,b]} f(x)\mu_g(\mathrm{d}\,x),$$

where μ_g denotes the Lebesgue-Stieltjes measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ induced by an increasing and right-continuous function g through $\mu_g((a,b]) = g(b) - g(a)$. There are different versions of the following result, where the necessary requirement of a directional continuity of g is often neglected in the literature. A precise proof of two other versions can be found in Dhaene et al. (2012) as Theorems 4 and 6.

Proposition 2.4. *For a distortion risk measure ρ_g with left-continuous distortion function g it holds*

$$\rho_g(X) = \int_0^1 F_X^{-1}(u)\,\mathrm{d}\,\bar{g}(u), \tag{2.1}$$

where $\bar{g}(u) = 1 - g(1-u)$, $u \in [0,1]$, is the dual distortion function.

Proof. First note that g is left-continuous if and only if $\bar{g}$ is right continuous. By definition of the Lebesgue-Stieltjes integral we have

$$\int_{1-S_X(x)}^1 \mathrm{d}\,\bar{g}(u) = \bar{g}(1) - \bar{g}(1 - S_X(x)) = g(S_X(x)),$$
$$\int_0^{1-S_X(x)} \mathrm{d}\,\bar{g}(u) = \bar{g}(1 - S_X(x)) - \bar{g}(0) = 1 - g(S_X(x)).$$

Inserting this in the definition of the distortion risk measure, we get

$$\begin{aligned}
\rho_g(X) &= \int_0^\infty \int_{1-S_X(x)}^1 \mathrm{d}\,\bar{g}(u)\,\mathrm{d}\,x - \int_{-\infty}^0 \int_0^{1-S_X(x)} \mathrm{d}\,\bar{g}(u)\,\mathrm{d}\,x \\
&= \int_0^\infty \int_0^1 \mathbb{1}\{F_X(x) < u\}\,\mathrm{d}\,\bar{g}(u)\,\mathrm{d}\,x - \int_{-\infty}^0 \int_0^1 \mathbb{1}\{u \le F_X(x)\}\,\mathrm{d}\,\bar{g}(u)\,\mathrm{d}\,x \\
&= \int_0^\infty \int_0^1 \mathbb{1}\{x < F_X^{-1}(u)\}\,\mathrm{d}\,\bar{g}(u)\,\mathrm{d}\,x - \int_{-\infty}^0 \int_0^1 \mathbb{1}\{F_X^{-1}(u) \le x\}\,\mathrm{d}\,\bar{g}(u)\,\mathrm{d}\,x \\
&= \int_0^1 \left(\int_0^\infty \mathbb{1}\{x < F_X^{-1}(u)\}\,\mathrm{d}\,x - \int_{-\infty}^0 \mathbb{1}\{F_X^{-1}(u) \le x\}\,\mathrm{d}\,x \right) \mathrm{d}\,\bar{g}(u) \\
&= \int_0^1 F_X^{-1}(u)\,\mathrm{d}\,\bar{g}(u).
\end{aligned}$$

Here, the third equality is by Lemma B.8 and the fourth by Tonelli's Theorem B.2. □

Many of the properties introduced in Definition 2.1 are fulfilled by distortion risk measures.

Lemma 2.5 (Sereda et al.; 2010, 25.4)**.** *a) The distortion risk measure is law invariant, monotone, positive homogeneous and comonotonic additive.*

b) The distortion risk measure is additionally translation invariant, i.e. monetary. The Wang premium principle has this property only if $\theta = 0$*.*

c) A distortion risk measure with concave distortion function g *preserves the increasing convex order, i.e.* $X \le_{icx} Y \Rightarrow \rho_g(X) \le \rho_g(Y)$*.*

d) A distortion risk measure is subadditive if and only if the distortion function g *is concave.*

The proof of parts a) and b) is by simple calculations which can be found in the cited reference. Part d) is more involved. It was proven by Dhaene and Wang (1998) relying on an incorrect proof of part c) by Wang (1996). Part c) was correctly proven later on by Dhaene et al. (2000).

Remark 2.6. Some authors refer to the dual distortion function $\bar{g}$ in Lemma 2.4 as the distortion function. Then subadditivity holds for *convex* distortions. This ambiguity can be avoided by using a different Lebesgue-Stieltjes representation

$$\rho_g(X) = \int_0^1 F_X^{-1}(1-u)\,\mathrm{d}\,g(u) \tag{2.2}$$

of distortion risk measures with left-continuous distortion function which does not involve the dual distortion function, cf. Dhaene et al. (2012). However, (2.2) requires another notion of Lebesgue-Stieltjes integrals and is less convenient when working with the parametrization of Value-at-Risk and Expected Shortfall commonly used in insurance (see below). So we will stick to (2.1).

Part d) of the Lemma 2.5 gives rise to defining a subclass of distortion risk measures.

Remark and Definition 2.7. For a continuous concave distortion function $g : [0,1] \to [0,1]$, the dual distortion function $\bar{g} : [0,1] \to [0,1]$ is continuous convex and can be written as $\bar{g}(u) = \int_0^u \phi(s)\,\mathrm{d}\,s$ for an increasing right-continuous function $\phi : [0,1] \to \mathbb{R}_+$, which is called *spectrum*. By the properties of the Lebesgue-Stieltjes integral, (2.1) can then be transformed to

$$\rho_g(X) = \rho_\phi(X) = \int_0^1 F_X^{-1}(u)\phi(u)\,\mathrm{d}\,u. \tag{2.3}$$

Therefore, distortion risk measures with continuous concave distortion function are referred to as *spectral risk measures*. Note that continuity of g is an additional requirement only in 0, since an increasing concave function on $[0,1]$ is already continuous on $(0,1]$.

Originally, spectral risk measures were defined by Acerbi (2002) without explicitly considering the dual distortion function. For this approach, every increasing right-continuous function $\phi : [0,1] \to \mathbb{R}_+$ with $\int_0^1 \phi(u)\,\mathrm{d}\,u = 1$ is an admissible spectrum.

In the following example, the best-known distortion risk measures and Wang premium principles are introduced.

Example 2.8. The most widely used risk measure in finance and insurance *Value-at-Risk*

$$\mathrm{VaR}_\alpha(X) = F_X^{-1}(\alpha), \qquad \alpha \in (0,1),$$

is a distortion risk measure with distortion function $g(u) = \mathbb{1}_{(1-\alpha,1]}(u)$. Since the distortion function is not concave, Value-at-Risk is not coherent and especially not a spectral risk measure. Within our parametrization, which is standard in insurance (cf. e.g. Denuit et al.; 2005; Rüschendorf; 2013), α is chosen close to 1. For instance, Solvency II requires $\alpha = 0.995$. The frequently criticized lack of coherence can be overcome by using *Expected Shortfall*

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

The corresponding distortion function $g(u) = \min\{\frac{u}{1-\alpha}, 1\}$ is concave and Expected Shortfall thus coherent. It is also spectral with $\phi(u) = \frac{1}{1-\alpha}\mathbb{1}_{[\alpha,1]}(u)$.

The *Proportional Hazard (PH) premium principle*

$$\pi(X) = (1+\theta)\int_0^\infty S_X(x)^\gamma\,\mathrm{d}\,x, \qquad \theta \geq 0,\ \gamma \in (0,1],$$

is a well-known example from the class of Wang premium principles. Note that the distortion function $g(x) = x^\gamma$, $\gamma \in (0,1]$ is concave. For $\gamma = 1$ the *Expected premium principle*

$$\pi(X) = (1+\theta)\mathbb{E}(X), \qquad \theta \geq 0,$$

is a special case.

Every spectral risk measure can be expressed as a mixture of Expected Shortfall over different confidence levels. The following result combines Proposition 8.18 of McNeil et al. (2015) and Remark 3 in Shapiro (2013). Fix a spectrum $\phi : [0,1] \to \mathbb{R}_+$. Then $\nu([0,t]) = \phi(t)$ defines a Borel measure on $[0,1]$ since ϕ is increasing and right continuous. Let us define a further measure μ by $\frac{d\mu}{d\nu}(\alpha) = (1-\alpha)$.

Proposition 2.9. *a) Let ρ_ϕ be a spectral risk measure. Then μ is a probability measure on $[0,1]$ and ρ_ϕ has the representation*

$$\rho_\phi(X) = \int_0^1 \mathrm{ES}_\alpha(X)\mu(\mathrm{d}\,\alpha).$$

b) Conversely, if a risk measure ρ can be represented as in a) with a probability measure μ, it is spectral and the spectrum $\phi : [0,1] \to \mathbb{R}_+$ is given by

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

Proof. a) Using the integration by parts rule for Lebesgue-Stieltjes integrals, one gets

$$\begin{aligned}\mu([0,1]) &= \int_0^1 1-\alpha\ \nu(\mathrm{d}\,\alpha) = \int_0^1 1-\alpha\ \mathrm{d}\,\phi(\alpha) = \phi(1) - \int_0^1 \alpha\,\mathrm{d}\,\phi(\alpha)\\ &= \phi(1) - [\alpha\phi(\alpha)]_0^1 + \int_0^1 \phi(\alpha)\,\mathrm{d}\,\alpha = \bar{g}(1) = 1,\end{aligned}$$

i.e. μ is a probability measure. Now, we have

$$\begin{aligned}\int_0^1 \mathrm{ES}_\alpha(X)\mu(\mathrm{d}\,\alpha) &= \int_0^1 \left(\frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(u)\,\mathrm{d}\,u\right)(1-\alpha)\,\nu(\mathrm{d}\,\alpha)\\ &= \int_0^1\int_0^1 F_X^{-1}(u)\mathbb{1}\{\alpha \le u\}\,\mathrm{d}\,u\ \nu(\mathrm{d}\,\alpha)\\ &= \int_0^1\int_0^1 F_X^{-1}(u)\mathbb{1}\{\alpha \le u\}\,\nu(\mathrm{d}\,\alpha)\,\mathrm{d}\,u\\ &= \int_0^1 F_X^{-1}(u)\int_0^u \nu(\mathrm{d}\,\alpha)\,\mathrm{d}\,u\\ &= \int_0^1 F_X^{-1}(u)\phi(u)\,\mathrm{d}\,u.\end{aligned}$$

The third equality is by Funini's theorem which can be applied since $X \in L^p$.

b) Define a Borel measure ν on $[0,1]$ by $\frac{d\nu}{d\mu}(\alpha) = \frac{1}{1-\alpha}$. Then

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

and the assertion follows from the calculation in the proof of part a). □

Remark 2.10. It has been shown by Shapiro (2013, Theorem 2) that every finite, law-invariant, coherent and comonotonic additive risk measure on L^p is already spectral. Note that comonotonic additivity is a natural extension of the properties translation invariance and positive homogeneity which are already included in coherence. Therefore, it is not a strong restriction to focus on spectral risk measures instead of considering a general coherent risk measure.

Especially in optimization, an infimum representation of Expected Shortfall going back to Rockafellar and Uryasev (2000) is very useful:

$$\mathrm{ES}_\alpha(X) = \inf_{q\in\mathbb{R}} \left\{ q + \frac{1}{1-\alpha}\mathbb{E}[(X-q)^+] \right\}, \qquad X \in L^p, \tag{2.4}$$

where the infimum is attained at $q = F_X^{-1}(\alpha)$. Pichler (2015) has proven an infimum representation for spectral risk measures generalizing the one of Expected Shortfall. We give an adapted version of the result which does not require $\phi \in L^q$.

Proposition 2.11. *Let ρ_ϕ be a spectral risk measure and $X \in L^p$ a random variable which is bounded from below. With G we denote the set of increasing convex functions $g : \mathbb{R} \to \mathbb{R}$. Then it holds*

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

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

Proof. Let $g \in G$, $X \in L^p$ and $U_X \sim \mathcal{U}(0,1)$ be the generalized distributional transform of X. By the definition of the convex conjugate it holds $g(X) + g^*(\phi(U_X)) \geq X\,\phi(U_X)$. Hence, we have

$$\begin{aligned}
\mathbb{E}[g(X)] + \mathbb{E}[g^*(\phi(U_X))] &\geq \mathbb{E}[X\,\phi(U_X)] \\
&= \mathbb{E}[F_X^{-1}(U_X)\,\phi(U_X)] \\
&= \int_0^1 F_X^{-1}(u)\phi(u)\,\mathrm{d}\,u = \rho_\phi(X)
\end{aligned}$$

whenever the expectations on the left hand side exist. Since $g \in G$ was arbitrary, it follows

$$\rho_\phi(X) \leq \inf_{g\in G} \left\{ \mathbb{E}[g(X)] + \int_0^1 g^*(\phi(u))\,\mathrm{d}\,u \right\}. \tag{2.5}$$

The function $g_{\phi,X} : \mathbb{R} \to \mathbb{R}$, $g_{\phi,X}(x) = \int_0^1 F_X^{-1}(\alpha) + \frac{1}{1-\alpha}\left(x - F_X^{-1}(\alpha)\right)^+ \mu(\mathrm{d}\,\alpha)$ with μ from Proposition 2.9 is increasing and convex. Using this proposition one obtains

$$\begin{aligned}
\rho_\phi(X) &= \int_0^1 \mathrm{ES}_\alpha(X)\mu(\mathrm{d}\,\alpha) \\
&= \int_0^1 F_X^{-1}(\alpha) + \frac{1}{1-\alpha}\mathbb{E}[(X - F_X^{-1}(\alpha))^+]\mu(\mathrm{d}\,\alpha) \\
&= \mathbb{E}\left[\int_0^1 F_X^{-1}(\alpha) + \frac{1}{1-\alpha}(X - F_X^{-1}(\alpha))^+\mu(\mathrm{d}\,\alpha)\right] \\
&= \mathbb{E}\left[g_{\phi,X}(X)\right].
\end{aligned} \tag{2.6}$$

Here, the second equality is due to (2.4) and the third due to Tonelli's Theorem B.2 since the integrand is bounded from below. Again, by Tonelli's Theorem B.2 and Lemma B.8 we have

$$\begin{aligned}\int_0^1 \frac{1}{1-\alpha}\left(x - F_X^{-1}(\alpha)\right)^+ \mu(\mathrm{d}\,\alpha) &= \int_0^1 \int_{-\infty}^x \frac{1}{1-\alpha}\mathbb{1}\left\{F_X^{-1}(\alpha) \le z\right\} \mathrm{d}\,z\,\mu(\mathrm{d}\,\alpha)\\ &= \int_{-\infty}^x \int_0^1 \frac{1}{1-\alpha}\mathbb{1}\left\{\alpha \le F_X(z)\right\} \mu(\mathrm{d}\,\alpha)\,\mathrm{d}\,z\\ &= \int_{-\infty}^x \int_0^{F_X(z)} \frac{1}{1-\alpha}\mu(\mathrm{d}\,\alpha)\,\mathrm{d}\,z.\end{aligned}$$

Hence, $g'_{\phi,X}(x) = \int_0^{F_X(x)} \frac{1}{1-\alpha}\mu(\mathrm{d}\,\alpha) = \phi(F_X(x))$ a.e., where the last equality is due to the definition of μ. For the convex conjugate $g^*_{\phi,X}(\phi(u)) = \sup_{x\in\mathbb{R}}\{\phi(u)x - g_{\phi,X}(x)\}$ the supremum is therefore attained at every x satisfying $\phi(u) = g'_{\phi,X}(x) = \phi(F_X(x))$, i.e.

$$g^*_{\phi,X}(\phi(u)) = \phi(u)F_X^{-1}(u) - g_{\phi,X}\left(F_X^{-1}(u)\right) \quad a.e.$$

Integrating with respect to u and using (2.6) yields

$$\int_0^1 g^*_{\phi,X}(\phi(u))\,\mathrm{d}\,u = \rho_\phi(X) - \mathbb{E}\left[g_{\phi,X}(X)\right] = 0.$$

Consequently, the lower bound in (2.5) is attained and the proof is complete. □

Besides Wang premium principles, so-called certainty equivalents are another large class of premium principles. Recall that a *disutility function* is a strictly increasing, continuous and convex function $u : \mathbb{R}_+ \to \mathbb{R}_+$. A certainty equivalent is a deterministic outcome c that yields the same disutility as a risk $X \in L^p$, i.e.

$$u(c) = \mathbb{E}[u(X)]. \tag{2.7}$$

Since u can be inverted, we have the following definition.

Definition 2.12. The *certainty equivalent (CE) premium principle* with respect to a disutility function u is given by $\pi : L^p_+ \to \bar{\mathbb{R}}$,

$$\pi(X) = u^{-1}\left(\mathbb{E}[u(X)]\right).$$

Clearly, CE premium principles are law-invariant and monotone. However, it was shown by Müller (2007) that they are translation invariant only if u is either affine or exponential and coherent only if u is the identity. Therefore, certainty equivalents are rather used as premium principles than as risk measures, with one notable exception.

Example 2.13. The exponential disutility function $u(x) = \exp(\gamma x)$, $\gamma > 0$, is well-defined for $x \in \mathbb{R}$. The respective certainty equivalent

$$X \mapsto \frac{1}{\gamma}\log\mathbb{E}\left[e^{\gamma X}\right]$$

is translation invariant and referred to as *entropic risk measure* for $X \in L^p$ as well as *exponential premium principle* for $X \in L^p_+$.

Remark 2.14. We have seen that both Wang premium principles and certainty equivalents yield reasonable premiums by modifying expectation: The former adjust probabilities while the latter adjust outcomes. The two approaches are unified by so-called *rank-dependent expected disutilities*

$$\pi(u(X)) = (1+\theta)\int_0^\infty g(\mathbb{P}(u(X) > x))\,\mathrm{d}\,x$$

where π is a Wang premium principle and u a disutility function. The concept naturally generalizes to risk measures.

2.2. Continuity Properties

The properties of risk measures and premium principles discussed so far have an obvious economic interpretation. In this section, we will consider two continuity properties which are mainly of mathematical interest.

Definition 2.15. A risk measure $\rho : L^p \to \bar{\mathbb{R}}$ has the

a) *Fatou property* if for every sequence $\{X_n\}_{n\in\mathbb{N}} \subseteq L^p$ with $|X_n| \le Y$ $\mathbb{P}$-a.s. for some $Y \in L^p$ and $X_n \to X$ $\mathbb{P}$-a.s. for some $X \in L^p$ it holds

$$\liminf_{n\to\infty} \rho(X_n) \ge \rho(X).$$

b) *Lebesgue property* if for every sequence $\{X_n\}_{n\in\mathbb{N}} \subseteq L^p$ with $|X_n| \le Y$ $\mathbb{P}$-a.s. for some $Y \in L^p$ and $X_n \to X$ $\mathbb{P}$-a.s. for some $X \in L^p$ it holds

$$\lim_{n\to\infty} \rho(X_n) = \rho(X).$$

Proposition 2.16. *Finite convex risk measures $\rho : L^p \to \mathbb{R}$ have both the Fatou and the Lebesgue property.*

For a proof we refer to Rüschendorf (2013), Theorem 7.24. This result covers many spectral risk measures and including Expected Shortfall.

Corollary 2.17. *Spectral risk measures $\rho_\phi : L^p \to \bar{R}$ with spectrum $\phi \in L^q$ have both the Fatou and the Lebesgue property.*

Proof. In order to apply Proposition 2.16, we only have to show that ρ_ϕ is finite on L^p. It follows from Hölder's inequality that

$$|\rho_\phi(X)| = \left|\int_0^1 F_X^{-1}(u)\phi(u)\,\mathrm{d}\,u\right| \le \int_0^1 |F_X^{-1}(u)|\phi(u)\,\mathrm{d}\,u = \left(\mathbb{E}|F_X^{-1}(U)|^p\right)^{\frac{1}{p}}\left(\mathbb{E}|\phi(U)|^q\right)^{\frac{1}{q}} < \infty$$

where $U \sim \mathcal{U}([0,1])$ is arbitrary. □

To the best of our knowledge, it has surprisingly not been investigated in the literature whether Value-at-Risk as the most widely used risk measure has the Fatou property.

Proposition 2.18. *Value-at-Risk has the Fatou property.*

Proof. Assume the contrary. Then there exists a sequence $\{X_n\}_{n\in\mathbb{N}} \subseteq L^p$ with $|X_n| \leq Y$ $\mathbb{P}$-a.s. for some $Y \in L^p$ and $X_n \to X$ $\mathbb{P}$-a.s. for some $X \in L^p$ such that

$$\liminf_{n\to\infty} \mathrm{VaR}_\alpha(X_n) < \mathrm{VaR}_\alpha(X).$$

I.e. there is an $\epsilon > 0$ such that for every $\delta \in (0, \epsilon)$

$$\liminf_{n\to\infty} F_{X_n}^{-1}(\alpha) \leq F_X^{-1}(\alpha) - \delta.$$

Hence, there exists a subsequence $\{F_{X_{N_k}}^{-1}(\alpha)\}_{k\in\mathbb{N}}$ such that for all $k \in \mathbb{N}$ and $\delta \in (0, \epsilon)$

$$F_{X_{n_k}}^{-1}(\alpha) \leq F_X^{-1}(\alpha) - \delta$$

or equivalently (cf. Lemma B.8)

$$\alpha \leq F_{X_{n_k}}(F_X^{-1}(\alpha) - \delta).$$

Since F_X has at most countably many discontinuities, we can choose $\delta_0 \in (0, \epsilon)$ such that $F_X^{-1}(\alpha) - \delta_0$ is a point of continuity of F_X. Then, by the definition of convergence in distribution

$$\alpha \leq \lim_{k\to\infty} F_{X_{n_k}}(F_X^{-1}(\alpha) - \delta_0) = F_X(F_X^{-1}(\alpha) - \delta_0).$$

Again by Lemma B.8 this is equivalent to

$$F_X^{-1}(\alpha) \leq F_X^{-1}(\alpha) - \delta_0,$$

a contradiction. □

Since premium principles are applied to non-negative risks, Fatou's Lemma B.1 yields the following continuity properties.

Lemma 2.19. *a) For a left-continuous distortion function g, the Wang premium principle has the Fatou property.*

b) The CE premium principle has the Fatou property.

Proof. Let $\{X_n\}_{n\in\mathbb{N}} \subseteq L_+^p$ with $X_n \leq Y$ $\mathbb{P}$-a.s. for some $Y \in L_+^p$ and $X_n \to X$ $\mathbb{P}$-a.s. for $X \in L_+^p$.

a) Especially, $X_n \to X$ in distribution. Therefore, $S_{X_n}(x) \to S_X(x)$ for almost every $x \in R_+$. Since g is left-continuous and increasing it is lower semicontinuous, i.e.

$\liminf_{n\to\infty} g(S_{X_n}(x)) \geq g(S_X(x))$ for almost every $x \in R_+$. Now by Fatou's Lemma B.1,

$$\begin{aligned}\liminf_{n\to\infty} \pi(X_n) &= \liminf_{n\to\infty} (1+\theta) \int_0^\infty g(S_{X_n}(x)) \,\mathrm{d}\, x \\ &\geq (1+\theta) \int_0^\infty g(S_X(x)) \,\mathrm{d}\, x = \pi(X).\end{aligned}$$

b) By the continuous mapping theorem, $u(X_n) \to u(X)$ $\mathbb{P}$-a.s. Then Fatou's Lemma B.1 yields $\liminf_{n\to\infty} \mathbb{E}[u(X_n)] \geq \mathbb{E}[u(X)]$. Since u^{-1} is continuous as well, we finally have

$$\liminf_{n\to\infty} \pi(X_n) = \liminf_{n\to\infty} u^{-1}(\mathbb{E}[u(X_n)]) \geq u^{-1}(\mathbb{E}[u(X)]) = \pi(X). \qquad \square$$

2.3. Dual Representation

Convex risk measures can be expressed as worst-case expectations minus a penalty term. Often, this is referred to as *robust representation*. For coherent risk measures the representation becomes particularly nice since the penalty term vanishes. These observations were first made for risk measures defined on L^∞. A detailed account can be found in Föllmer and Schied (2016). The results were later generalized to L^p-spaces with $p \in [1, \infty]$ using the Fenchel-Moreau Theorem from convex analysis. This connection to duality gives rise to the alternative denomination of robust representations in the headline. With regard to our purposes the following presentation is restricted to risk measures on L^p with $p \in [1, \infty)$.

Denote by $\mathcal{M}_1(\Omega, \mathcal{A}, \mathbb{P})$ the set of probability measures on $(\Omega, \mathcal{A})$ which are absolutely continuous with respect to $\mathbb{P}$ and let

$$\mathcal{M}_1^q(\Omega, \mathcal{A}, \mathbb{P}) = \left\{ \mathbb{Q} \in \mathcal{M}_1(\Omega, \mathcal{A}, \mathbb{P}) : \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} \in L^q(\Omega, \mathcal{A}, \mathbb{P}) \right\}.$$

Moreover, recall that a $\bar{\mathbb{R}}$-valued convex functional is called *proper* if it never attains $-\infty$ and is strictly smaller than $+\infty$ in at least one point.

Proposition 2.20 (Rüschendorf; 2013, 7.14)**.** *Let $\rho : L^p \to \bar{R}$ be a proper convex risk measure with the Fatou property, then*

$$\rho(X) = \sup_{\mathbb{Q} \in \mathcal{M}_1^q(\Omega,\mathcal{A},\mathbb{P})} \left(\mathbb{E}^{\mathbb{Q}}[X] - \rho^*(\mathbb{Q}) \right), \qquad X \in L^p.$$

Here, $\rho^*(\mathbb{Q}) = \sup_{X \in L^p} \left(\mathbb{E}^{\mathbb{Q}}[X] - \rho(X) \right)$ denotes the convex conjugate of ρ.

Proof. The result is an immediate consequence of the Fenchel-Moreau Theorem. $\square$

The representation simplifies if the risk measure is additionally positive homogeneous.

Proposition 2.21 (Rüschendorf; 2013, 7.20)**.** *A functional $\rho : L^p \to \bar{R}$ is a proper coherent risk measure with the Fatou property if and only if there exists a subset $\mathcal{Q} \subseteq \mathcal{M}_1^q(\Omega, \mathcal{A}, \mathbb{P})$*

such that

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X], \qquad X \in L^p. \tag{2.8}$$

The supremum is attained since the subset $\mathcal{Q} \subseteq \mathcal{M}_1^q(\Omega, \mathcal{A}, \mathbb{P})$ can be chosen $\sigma(L^q, L^p)$-compact and the functional $\mathbb{Q} \mapsto \mathbb{E}^{\mathbb{Q}}[X]$ is $\sigma(L^q, L^p)$-continuous.

Proof. Let $\rho : L^p \to \bar{R}$ be a proper coherent risk measure with the Fatou property. Due to the positive homogeneity of ρ it holds for all $\mathbb{Q} \in \mathcal{M}_1^q(\Omega, \mathcal{A}, \mathbb{P})$ and $\lambda \in \mathbb{R}_+$ that

$$\begin{aligned}
\rho^*(\mathbb{Q}) &= \sup_{X\in L^p} \mathbb{E}^{\mathbb{Q}}[X] - \rho(X) \\
&= \sup_{\lambda X\in L^p} \mathbb{E}^{\mathbb{Q}}[\lambda X] - \rho(\lambda X) \\
&= \lambda\rho^*(\mathbb{Q}),
\end{aligned}$$

i.e. $\rho^*(\mathbb{Q}) \in \{0, \infty\}$. Setting

$$\mathcal{Q} = \{\mathbb{Q} \in M_1^q(\Omega, \mathcal{A}, \mathbb{P}) : \rho^*(\mathbb{Q}) = 0\},$$

Proposition 2.20 yields

$$\rho(X) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X], \qquad X \in L^p,$$

and it remains to show that the supremum is attained. The functional $\mathbb{Q} \mapsto \mathbb{E}^{\mathbb{Q}}[X]$ is continuous for every $X \in L^p$ by definition of the weak* topology $\sigma(L^q, L^p)$. In Proposition 7.19 of Rüschendorf (2013) it is shown that $\mathcal{Q}$ is $\sigma(L^q, L^p)$-compact. Hence, Weierstraß' Extreme Value Theorem yields the assertion.

Conversely, let $\rho : L^p \to \bar{R}$ be representable as in (2.8) with some $\sigma(L^q, L^p)$-compact subset $\mathcal{Q} \subseteq \mathcal{M}_1^q(\Omega, \mathcal{A}, \mathbb{P})$. Then it is a coherent risk measure since the properties of monotonicity, translation invariance, positive homogeneity and additivity are trivially satisfied. As the supremum is attained in (2.8), we find for fixed $X \in L^p$ a probability measure $\mathbb{Q}_X \in \mathcal{Q}$ such that

$$\rho(|X|) = \mathbb{E}^{\mathbb{Q}_X}|X| \le (\mathbb{E}|X|^p)^{\frac{1}{p}} \left(\mathbb{E}\left|\frac{\mathrm{d}\,\mathbb{Q}_X}{\mathrm{d}\,\mathbb{P}}\right|^q\right)^{\frac{1}{q}} < \infty.$$

Hence, ρ is finite by Lemma 2.2 and especially proper. Due to finiteness, the Fatou property follows from Proposition 2.16. □

Remark 2.22. Since ρ^* is convex as a conjugate function and takes values in $\{0, \infty\}$, the set

$$\mathcal{Q} = \{\mathbb{Q} \in M_1^q(\Omega, \mathcal{A}, \mathbb{P}) : \rho^*(\mathbb{Q}) = 0\} = \{\mathbb{Q} \in M_1^q(\Omega, \mathcal{A}, \mathbb{P}) : \rho^*(\mathbb{Q}) \le 0\}$$

of the dual representation is convex, too, as a sublevel set of a convex function.

With the dual representation (2.8) we can prove a complementary inequality to subadditivity.

Lemma 2.23. *A proper coherent risk measure with the Fatou property $\rho : L^p \to \bar{R}$ satisfies the inequality*

$$\rho(X+Y) \geq \rho(X) - \rho(-Y) \qquad \text{for all } X, Y \in L^p.$$

Proof. By Proposition 2.21 it holds for $X, Y \in L^p$

$$\begin{aligned}
\rho(X+Y) &= \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X+Y] = \sup_{\mathbb{Q}\in\mathcal{Q}} \Big(\mathbb{E}^{\mathbb{Q}}[X] + \mathbb{E}^{\mathbb{Q}}[Y]\Big) \\
&\geq \sup_{\mathbb{Q}\in\mathcal{Q}} \Big(\mathbb{E}^{\mathbb{Q}}[X] + \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[Y]\Big) = \sup_{\mathbb{Q}\in\mathcal{Q}} \Big(\mathbb{E}^{\mathbb{Q}}[X] - \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[-Y]\Big) \\
&= \sup_{\mathbb{Q}\in\mathcal{Q}} \Big(\mathbb{E}^{\mathbb{Q}}[X] - \rho(-Y)\Big) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X] - \rho(-Y) \\
&= \rho(X) - \rho(-Y).
\end{aligned}$$

□

For spectral risk measures the dual representation becomes more explicit. The original proof by Pichler (2015) is along the line of the general case, i.e. a calculation of the convex conjugate ρ_ϕ^* and its null set. That means part (iii) of the following Proposition was proven first and the other parts where then derived as Corollaries. Proceeding the other way round, we give a shorter alternative proof.

Proposition 2.24. *A spectral risk measure $\rho_\phi : L^p \to \mathbb{R}$ with spectrum $\phi \in L^q$ can be represented as*

(i)

$$\rho_\phi(X) = \sup_{U\sim\mathcal{U}(0,1)} \mathbb{E}[X\phi(U)].$$

(ii)

$$\rho_\phi(X) = \sup\Big\{\mathbb{E}[XY] : Y \in L^q,\ Y \leq_{cx} \phi(U),\ U \sim \mathcal{U}(0,1)\Big\}.$$

(iii)

$$\rho_\phi(X) = \sup\Bigg\{\mathbb{E}[XY] : Y \in L^q,\ E[Y] = 1, \\ \mathrm{ES}_\alpha(Y) \leq \frac{1}{1-\alpha}\int_\alpha^1 \phi(u)\,\mathrm{d}\,u,\ 0 \leq \alpha \leq 1\Bigg\}$$

The suprema are attained and the maximizer is given by $\phi(U_X)$, where U_X is the generalized distributional transform of X.

Proof. (i) We can reformulate the definition of a spectral risk measure to

$$\rho_\phi(X) = \int_0^1 F_X^{-1}(u)\phi(u)\,\mathrm{d}\,u = \mathbb{E}\left[F_X^{-1}(U_X)\phi(U_X)\right] = \mathbb{E}\left[X\phi(U_X)\right],$$

where the last equality is by Lemma B.10. For random vectors (X_1, X_2) and (Y_1, Y_2) with the same marginals it follows from the upper Fréchet-Hoeffding bound that

$$\mathbb{E}[X_1 X_2] \le \mathbb{E}[Y_1 Y_2]$$

if (Y_1, Y_2) is comonotonic and the expectations exist, cf. Müller and Stoyan (2002, 3.1.1, 3.8.2). Recalling that X and U_X are comonotonic and ϕ is increasing yields (i) and the assertion regarding the maximizer.

(ii) Let $X \in L^p$ and $Y \in L^q$ with $Y \le_{cx} \phi(U),\ U \sim \mathcal{U}(0,1)$. We proceed in three steps.

Step 1: $X \ge 0$
Then $\sigma : [0,1] \to \mathbb{R}_+,\ \sigma(u) = \frac{q_X^+(u)}{\mathbb{E}[X]}$ is increasing and right continuous with normed integral, i.e. a spectrum. Therefore,

$$\begin{aligned}\mathbb{E}[XY] &\le \mathbb{E}[q_X^+(U_Y)Y] = \mathbb{E}[X] \cdot \rho_\sigma(Y)\\ &\le \mathbb{E}[X] \cdot \rho_\sigma(\phi(U_Y)) = \mathbb{E}[q_X^+(U_Y)\phi(U_Y)]\\ &= E[X\phi(U_X)]\end{aligned}$$

The first inequality is by the same argument as in (i) and the second one holds since the spectral risk measure $\rho_\sigma : L^q \to \mathbb{R}$ has the Fatou property (Proposition 2.17 for $q < \infty$ or Jouini et al. (2006) for $q = \infty$) and therefore preserves the convex order (Bäuerle and Müller; 2006, 4.3).

Step 2: $X \ge -N$ for some $N \in \mathbb{N}$
Since Y is a density by Remark 2.25 it follows from step 1 that

$$\mathbb{E}[XY] = \mathbb{E}[(X+N)Y] - N \le E[(X+N)\phi(U_X)] - N = E[X\phi(U_X)].$$

Step 3: general case
By step 2 it holds

$$\mathbb{E}[\max\{X, -N\}Y] \le E[\max\{X, -N\}\phi(U_X)], \qquad N \in \mathbb{N}.$$

Since $|\max\{X, -N\}| \le X \in L^p,\ N \in \mathbb{N}$ and $\max\{X, -N\} \to X$ $\mathbb{P}$-a.s. as $N \to \infty$, the claim follows with dominated convergence.

(iii) It holds $\mathbb{E}[\phi(U)] = \int_0^1 \phi(u)\,\mathrm{d}\,u = 1$ by the definition of a spectrum. Moreover, ϕ is increasing and right-continuous, i.e. an upper quantile function function. Now, (iii) is only a reformulation of the convex order $\le_{cx}$ in terms of ordered integrated quantile functions and equal means, cf. Shaked and Shanthikumar (2007, 3.A.5). Note that we replaced the lower quantile function of $\phi(U)$ with the upper one which is given by ϕ. Due to equality a.e. this does not change the integrals. □

Remark 2.25. The ordering $Y \leq_{cx} \phi(U)$ implies $\mathbb{E}[Y] = \mathbb{E}[\phi(U)] = 1$ and $Y \geq 0$ $\mathbb{P}$-a.s. Indeed, assume that $\alpha = \frac{1}{2}\mathbb{P}(Y < 0) > 0$. Then $F_Y^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F_Y(x) \geq \alpha\} < 0$ and since quantile functions are increasing $\int_0^\alpha F_Y^{-1}(d)\,\mathrm{d}\,u < 0$. But by the quantile representation of the convex order one gets

$$\int_0^\alpha F_Y^{-1}(u)\,\mathrm{d}\,u = 1 - \int_\alpha^1 F_Y^{-1}(u)\,\mathrm{d}\,u \geq 1 - \int_\alpha^1 \phi(u)\,\mathrm{d}\,u \geq 0,$$

a contradiction. Consequently, all suprema in Proposition 2.24 are taken over densities and we have a dual representation in the classical sense.

The well-known dual representation of Expected Shortfall is a special case.

Corollary 2.26. *The Expected Shortfall can be represented as*

$$\mathrm{ES}_\alpha(X) = \sup_{\mathbb{Q} \in \mathcal{Q}_\alpha} \mathbb{E}^{\mathbb{Q}}[X], \qquad X \in L^1,$$

where $\mathcal{Q}_\alpha = \{\mathbb{Q} \in M_1^\infty(\Omega, \mathcal{A}, \mathbb{P}) : \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}} \leq \frac{1}{1-\alpha}\}$.

Proof. Expected Shortfall has the spectrum $\phi(u) = \frac{1}{1-\alpha}\mathbb{1}[\alpha, 1](u)$, i.e. the representation in Proposition 2.24 becomes

$$\mathrm{ES}_\alpha(X) = \sup\Bigg\{\mathbb{E}[XY] : Y \in L^\infty,\ \mathbb{E}[Y] = 1,$$
$$(1-\beta)\,\mathrm{ES}_\beta(Y) \leq \frac{1}{1-\alpha}\int_{\max\{\alpha,\beta\}}^1 \mathrm{d}\,u,\ 0 \leq \beta \leq 1\Bigg\}.$$

The last constraint is equivalent to

$$\begin{aligned}
&(1-\beta)\,\mathrm{ES}_\beta(Y) \leq \frac{\min\{1-\alpha, 1-\beta\}}{1-\alpha} \\
\iff\quad &(1-\beta)\,\mathrm{ES}_\beta(Y) \leq \begin{cases} \frac{1-\beta}{1-\alpha}, & \beta \geq \alpha \\ 1, & \beta < \alpha \end{cases} \\
\iff\quad &\begin{cases} \mathrm{ES}_\beta(Y) \leq \frac{1}{1-\alpha}, & \beta \geq \alpha \\ \int_\beta^1 F_Y^{-1}(u)\,\mathrm{d}\,u \leq 1 & \beta < \alpha. \end{cases}
\end{aligned}$$

Since ES_β is increasing in β, the constraint in the first case is equivalent to

$$\mathrm{ES}_1(Y) = \operatorname{ess\,sup}(Y) \leq \frac{1}{1-\alpha}.$$

Moreover, the constraint in the second case is redundant because $\beta \mapsto \int_\beta^1 F_Y^{-1}(u)\,\mathrm{d}\,u$ is decreasing for non-negative Y and $\mathbb{E}[Y] = \int_0^1 F_Y^{-1}(u)du = 1$ holds for every density Y. □

CHAPTER 3

Markov Decision Model

We consider stochastic systems which evolve in discrete time and can be influenced by sequential decisions of a controller. The decisions incur a cost at each stage and influence the conditions for future decision-making. This is formalized by a state process with random transitions. Given the current state, the controller chooses an admissible action which influences the transition to the next state. The cost incurred at each stage may depend on the current state, the chosen action, and the next state. Hence, the decision-maker has to take into account the impact of his action on the current as well as on future costs and balance possible opposite effects.

In this chapter, we first introduce an abstract cost model which will subsequently be considered under different optimality criteria. Furthermore, we specify the decision-making of the controller. The dynamic reinsurance model introduced afterward is a special case and will serve as a running example. From the actuarial perspective, it is novel and therefore of interest on its own.

3.1. Abstract Cost Model

Under the term *Borel space* we understand a Borel subset S of a Polish space, i.e. complete, separable metric space equipped with the metric and the Borel σ-algebra $\mathcal{B}(S)$. Note that in the literature such spaces are occasionally referred to as *Standard Borel spaces* when Borel space only means topological space with Borel σ-algebra.

The abstract cost model is a Markov Decision Model with general Borel state and action spaces. We define the model components (or model data) distinguishing finite and infinite planning horizon. Properties of the components which are listed here are required to hold throughout all subsequent chapters, unless explicitly stated otherwise. Additional assumptions will be made later on specifically for the different optimality criteria.

The model with *finite planning horizon* $N \in \mathbb{N}$ has the following components for $n = 0, \ldots, N-1$:

- The *state space* E is a Borel space with Borel σ-algebra $\mathcal{B}(E)$. The elements $x \in E$ are called *states*.

- The *action space* A is a Borel space with Borel σ-Algebra $\mathcal{B}(A)$. The elements $a \in A$ are referred to as *actions*.

- The possible state-action combinations D_n at time n form a measurable subset 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. It induces a set-valued mapping $E \ni x \mapsto D_n(x)$.

- The *disturbances* $Z_1, \ldots, Z_N$ are independent random elements on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{Z}, \mathfrak{Z})$. Their influence on the next state is formalized by a measurable *transition function* $T_n : D_n \times \mathcal{Z} \to E$. When the current state is x_n, the controller chooses action a_n and z_{n+1} is the realization of Z_{n+1}, then the next state is given by
$$x_{n+1} = T_n(x_n, a_n, z_{n+1}).$$

- The *one-stage cost function* $c_n : D_n \times E \to \mathbb{R}$ gives the cost $c_n(x, a, x')$ which the controller incurs 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 $c_N(x)$ which the controller incurs if the system terminates in state x.

Since D_n contains the graph of a measurable map, the set of admissible actions $D_n(x)$ is non-empty for every state $x \in E$. Moreover,
$$D_n = \{(x, a) \in E \times A : a \in D_n(x)\},$$
i.e. D_n is the graph of the set-valued mapping $D_n(\cdot)$.

The model data is supposed to have the following continuity and compactness properties. In the subsequent chapters it will be stated explicitly, which of the three assumptions is made on the transition function.

Properties 3.1. (i) The set-valued mapping $E \ni x \mapsto D_n(x)$ is upper semicontinuous (see Definition A.12) and compact-valued for $n = 0, \ldots, N-1$.

(ii) Regarding the transition function we distinguish three cases:

Case 1: T_n is continuous in (x, a) for $n = 0, \dots, N-1$.
Case 2: T_n is lower semicontinuous in (x, a) for $n = 0, \dots, N-1$.
Case 3: T_n is upper semicontinuous in (x, a) for $n = 0, \dots, N-1$.

(iii) The one-stage cost $D_n \ni (x, a) \mapsto c_n(x, a, T_n(x, a, z))$ is lower semicontinuous for every $z \in \mathcal{Z}$ and $n = 0, \dots, N-1$ and so is the terminal cost function $E \ni x \mapsto c_N(x)$.

Note that in Case 1 it is sufficient due to Lemma A.4 a) to require that the one-stage cost function $c_n : D_n \times E \to \mathbb{R}$ is lower semicontinuous in order to obtain lower semicontinuity of the composition $c_n(\cdot, \cdot, T_n(\cdot, \cdot, z))$.

The abstract cost model is called *stationary* if $D,\ T$ do not depend on n, the disturbances are identically distributed, the one-stage cost functions are of the form $c_n = \beta^n c$, $n = 0, \dots, N-1$, and the terminal cost function is $\beta^N c_N$, where $\beta \in (0, 1]$ is a discount factor. In that case, Z denotes a representative of the disturbance distribution. If the model is stationary and the terminal cost is zero, we allow for an *infinite time horizon* $N = \infty$. For a non-stationary model, one may think of the discount factor being included in the cost functions.

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

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

where $a_k \in D(x_k)$ for $k \in N_0$. The set $\mathcal{H}_\infty$ is defined accordingly. 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. This axiomatic requirement is referred to as *non-anticipativity.*

Definition 3.2. a) A *randomized policy* is a sequence $\pi = (\pi_0, \pi_1, \dots)$ of stochastic kernels π_n from $\mathcal{H}_n$ to the action space A satisfying the constraint

$$\pi_n(D(x_n)|h_n) = 1, \qquad h_n \in \mathcal{H}_n.$$

A finite sequence $\pi = (\pi_0, \dots, \pi_{N-1})$ is referred to as *randomized N-stage policy.*

b) A measurable mapping $d_n : \mathcal{H}_n \to A$ with $d_n(h_n) \in D(x_n)$ for every $h_n \in \mathcal{H}_n$ is called *decision rule* at time n.

c) A decision rule at time n is called *Markov* if it only depends on the current state, i.e. $d_n(h_n) = d_n(x_n)$ for all $h_n \in \mathcal{H}_n$.

d) A sequence of decision rules $\pi = (d_0, d_1, \dots)$ is called *deterministic policy* and a finite sequence $\pi = (d_0, \dots, d_{N-1})$ is called *deterministic N-stage policy.*

e) If all decision rules are Markov, the deterministic (N-stage) policy is called *Markov.*

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

For convenience, deterministic policies may simply be referred to as policy. With $\Pi^R \supseteq \Pi \supseteq \Pi^M \supseteq \Pi^S$ we denote the sets of all randomized policies, deterministic policies, Markov policies and stationary policies. The first inclusion is by identifying deterministic decision rules d_n with the corresponding Dirac kernels

$$\pi_n(\cdot|h_n) = \delta_{d_n(h_n)}(\cdot), \qquad h_n \in \mathcal{H}_n.$$

It will be clear from the context if N-stage or infinite stage policies are meant. An admissible policy always exists since D_n contains the graph of a measurable mapping.

At each stage $n \in \mathbb{N}_0$, the transition function T_n and the disturbance Z_{n+1} induce a stochastic kernel

$$Q_n(B|x,a) = \int \mathbb{1}_B(T_n(x,a,Z_{n+1}(\omega)))\mathbb{P}(\mathrm{d}\,\omega), \qquad B \in \mathcal{B}(E),\ (x,a) \in D_n \tag{3.1}$$

from D_n to E characterizing the transition law. Note that (3.1) indeed defines a stochastic kernel: Firstly, $Q_n(\cdot|x,a)$ defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for every $(x,a) \in D_n$. Secondly, the function $D_n \times \Omega \ni (x,a,\omega) \mapsto \mathbb{1}_B(T_n(x,a,Z(\omega)))$ is measurable as a composition of measurable functions and hence $D_n \ni (x,a) \mapsto Q_n(B|x,a)$ is measurable for every $B \in \mathcal{B}(E)$ by Tonelli's Theorem B.2. The *decision process* $(X_n)_{n\in\mathbb{N}_0}$ can now be defined by the following *canonical construction* as for instance in Hernández-Lerma and Lasserre (1996). We directly consider an infinite time horizon with the N-stage version simply being a truncation. Define a measurable space by the sample space $\overline{\mathcal{H}}_\infty = (E \times A)^\infty$ and the product σ-algebra

$$\bigotimes_{n=0}^{\infty} (\mathcal{B}(E) \otimes \mathcal{B}(A)).$$

Elements of $\overline{\mathcal{H}}_\infty$ are of the form $\omega = (x_0, a_0, x_1, a_1, \dots)$. We define the *state process* $(X_n)_{n\in\mathbb{N}_0}$ and the *action process* $(A_n)_{n\in\mathbb{N}_0}$ on $\overline{\mathcal{H}}_\infty$ as projections

$$X_n(\omega) = x_n, \qquad A_n(\omega) = a_n, \qquad n \in \mathbb{N}_0.$$

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

$$H_0 = X_0, \quad H_1 = (X_0, A_0, X_1), \quad H_2 = (X_0, A_0, X_1, A_1, X_2), \dots$$

By the Theorem of Ionescu-Tulcea (Klenke; 2014, 14.32), each initial state $x \in E$ and policy $\pi \in \Pi^R$ of the controller induce a unique probability measure

$$\mathbb{Q}_x^\pi = \delta_x \otimes \pi_0 \otimes Q_0 \otimes \pi_1 \otimes Q_1 \otimes \dots \tag{3.2}$$

on $\overline{\mathcal{H}}_\infty$ called the *law of motion*. It satisfies for all $n \in \mathbb{N}_0$, $B \in \mathcal{B}(E)$ and $C \in \mathcal{B}(A)$

$$\mathbb{Q}_x^\pi(X_0 \in B) = \delta_x(B),$$
$$\mathbb{Q}_x^\pi(A_n \in C | H_n = h_n) = \pi_n(C|h_n),$$
$$\mathbb{Q}_x^\pi(X_{n+1} \in B | H_n = h_n, A_n = a_n) = Q_n(B|x_n, a_n).$$

Henceforth, we denote with $\mathbb{E}_x^\pi$ the expectation operator with respect to $\mathbb{Q}_x^\pi$ and with $\mathbb{E}_{nh_n}^\pi$ or $\mathbb{E}_{nx}^\pi$ the respective conditional expectation given $H_n = h_n$ or $X_n = x$. Clearly, for this canonical construction it was not necessary to define disturbances and transition functions since one could more generally start directly with transition kernels Q_n. In the following chapters, we will, however, rely on this functional representation of the transition law. The canonical construction is only needed to allow for randomized policies. Under a deterministic policy $\pi = (d_0, d_1, \dots) \in \Pi$ we do not need to specify a law of motion but can define the decision process directly by the functional representation

$$X_0^\pi = x_0, \qquad X_{n+1}^\pi = T(X_n^\pi, d_n(H_n^\pi), Z_{n+1}). \tag{3.3}$$

In this setting, expectations can be calculated with respect to the probability measure of the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ of the disturbances. Hence, we index the decision process and its random history with the policy rather than the expectation operator.

3.2. Dynamic Reinsurance Model

As an application, we study dynamic reinsurance of an insurance company in discrete time. Optimality criteria are minimization of solvency capital requirements or cost of solvency capital as well as robust maximization of expected total dividends. Since reinsurance treaties are typically written for one year (Albrecher et al.; 2017, p. 1) and dividends are paid annually, modeling in discrete time is appropriate when focusing on the management of the insurer's surplus by means of reinsurance and dividend payments while neglecting the possible use of capital market instruments.

3.2.1. Solvency Capital

The cost of solvency capital is given by the solvency capital requirement times the insurer's cost of capital rate. Hence, minimizing the two quantities is structurally the same. The model introduced here is a dynamic generalization of a static optimal reinsurance problem extensively studied in the literature, starting with Cai and Tan (2007) and generalizations i.a. by Chi and Tan (2013) and Cui et al. (2013). The insurer is endowed with an initial capital $x \in \mathbb{R}_+$. At the end of each period $[n, n+1)$, $n \in \mathbb{N}_0$, he incurs aggregate claims Y_{n+1} for that period and receives the total premium income Z_{n+1} for the next period. Both quantities are stochastic and therefore modeled by non-negative random variables. Thus, the insurer's uncontrolled surplus process is given recursively by

$$X_0 = x, \qquad X_{n+1} = X_n - Y_{n+1} + Z_{n+1}.$$

In order to reduce the downside risk of its surplus process, the insurance company can underwrite a reinsurance treaty at the beginning of each period. A reinsurance treaty is described by a *retained loss function* $f : \mathbb{R}_+ \to \mathbb{R}_+$. When purchasing reinsurance f_n at time n, the insurance company retains the portion $f_n(Y_{n+1})$ of the claims Y_{n+1} arriving at time $n+1$ and the reinsurer covers $Y_{n+1} - f_n(Y_{n+1})$. In return, the insurer has to pay a reinsurance premium $\pi_R\big(Y_{n+1} - f_n(Y_{n+1})\big)$. The admissible retained loss functions are

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

The first condition $0 \leq f \leq \mathrm{id}_{\mathbb{R}_+} \Leftrightarrow 0 \leq \mathrm{id}_{\mathbb{R}_+} - f \leq \mathrm{id}_{\mathbb{R}_+}$ ensures that only actual losses can be reinsured. The second and third condition ensure that the retained loss $f_n(Y_{n+1})$ and the *ceded loss* $Y_{n+1} - f_n(Y_{n+1})$ are comonotonic random variables, i.e. that both the insurer and the reinsurer suffer from higher claims. Otherwise, the insurer might have an incentive to misreport losses or accept unjustified claims. This form of moral hazard is precluded by the constraint which is also referred to as *incentive compatibility* condition in the literature. Additionally, one may introduce a budget constraint. The dynamic of the controlled surplus process is given by

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

Let us now formulate the reinsurance model as a stationary Markov Decision Process and embed it in the abstract cost model. Important properties are summarized below in Lemma 3.3.

- The state space is the real line $\mathbb{R}$ with Borel σ-algebra $\mathcal{B}(\mathbb{R})$.
- The action space is $\mathcal{F}$ with Borel σ-algebra $\mathcal{B}(\mathcal{F})$.
- The disturbance space is $\mathbb{R}_+^2$ with Borel σ-algebra $\mathcal{B}(\mathbb{R}_+^2)$ and the disturbances are $(Y_n, Z_n)_{n\in\mathbb{N}}$. It is assumed that claims $(Y_n)_{n\in\mathbb{N}}$ and premium income $(Z_n)_{n\in\mathbb{N}}$ are non-negative, independent and defined on a common atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as well as

$$\begin{aligned} Y_1, Y_2, \ldots &\overset{\text{iid}}{\sim} Y \in L^p(\Omega, \mathcal{A}, \mathbb{P}) \\ Z_1, Z_2, \ldots &\overset{\text{iid}}{\sim} Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P}) \end{aligned}$$

  for some $p \in [1, \infty)$.

- The transition function $T : \mathbb{R} \times \mathcal{F} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$T(x, f, y, z) = x - f(y) - \pi_R(f) + z.$$

  Here, $\pi_R : L^p_+ \to \bar{\mathbb{R}}$ is a law-invariant, monotone and normalized premium principle having the Fatou property and satisfying $\pi_R(Y) < \infty$. Due to the identical distribution of the claims we can use the shorthand notation $\pi_R(f) = \pi_R(Y - f_n(Y))$, $f \in \mathcal{F}$.

- Regarding the admissible actions $D(x)$ in state $x \in \mathbb{R}$ we will consider two cases:

  Unconstrained: $D(x) = \mathcal{F}$ for all $x \in \mathbb{R}$.

  Budget-constrained: $D(x) = \{f \in \mathcal{F} : \pi_R(f) \leq x^+\}$ for all $x \in \mathbb{R}$.

  The set of all state-action combinations is $D = \{(x, f) \in \mathbb{R} \times \mathcal{F} : f \in D(x)\}$. It contains the graph of the constant measurable map $\mathbb{R} \ni x \mapsto \mathrm{id}_{\mathbb{R}_+}$.

- Regarding the one-stage cost function $c : D \times \mathbb{R} \to \mathbb{R}$ we also consider two cases:

  Cumulative loss: $c(x, f, x') = -x'$.

  Incremental loss: $c(x, f, x') = -(x' - x) = x - x'$.

  There is no terminal cost.

Under a finite planning horizon, one could of course formulate a non-stationary version of the model. Mathematically, there is no difficulty so we omit this for notational convenience. Requiring that the aggregate losses are independent and fulfill some integrability condition is standard in actuarial science. Often, the premium income is assumed to be deterministic. Here, we allow for some uncertainty or fluctuation but in practice one will at least know an upper bound (complete and timely payment by all policyholders). The assumption $\pi_R(Y) < \infty$, meaning that the risk can be fully ceded at each stage, is natural for a model with a passive reinsurer. The budget constraint implies that reinsurance cannot be purchased on credit but a temporarily negative capital is allowed.

Lemma 3.3. *a) All retained loss functions $f \in \mathcal{F}$ are Lipschitz continuous with constant $L \leq 1$.*

b) $\mathcal{F}$ 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 metric

$$m(f, g) = \sum_{j=1}^{\infty} 2^{-j} \frac{\max_{0 \leq t \leq j} |f(t) - g(t)|}{1 + \max_{0 \leq t \leq j} |f(t) - g(t)|}.$$

c) The functional $\pi_R : \mathcal{F} \to \mathbb{R}_+,\ f \mapsto \pi_R(f)$ is lower semicontinuous.

d) The transition function T is upper semicontinuous and hence measurable.

e) $D(x)$ is a compact subset of $\mathcal{F}$ for all $x \in \mathbb{R}$.

f) The set-valued mapping $\mathbb{R} \ni x \to D(x)$ is upper semicontinuous.

g) *The one-stage cost* $D \ni (x, f) \mapsto c(x, f, T(x, f, y, z))$ *is lower semicontinuous for every* $(y, z) \in \mathbb{R}_+^2$.

Proof. a) Let $f \in \mathcal{F}$. Since $\mathrm{id}_{\mathbb{R}_+} - f$ is increasing, it holds for $0 \le x \le y$ that $x - f(x) \le y - f(y)$. Rearranging and using that f is increasing, too, yields

$$|f(x) - f(y)| = f(y) - f(x) \le y - x = |x - y|.$$

b) Let $\{f_k\}_{k\in\mathbb{N}}$ be a convergent sequence in $\mathcal{F}$ with limit f. Then, $f_k(x) \to f(x)$ for all $x \in \mathbb{R}_+$. Now it is easily checked that $f \in \mathcal{F}$, i.e. $\mathcal{F}$ is closed. Moreover, $\mathcal{F}$ is relatively compact by the Arzelà-Ascoli Theorem A.32 together with Remark A.33 since $f \le \mathrm{id}_{\mathbb{R}_+}$ for all $f \in \mathcal{F}$ and the functions in $\mathcal{F}$ have a common Lipschitz constant by part a). Hence, $(\mathcal{F}, m)$ is a compact metric space and as such also complete and separable (Aliprantis and Border; 2006, 3.26, 3.28). I.e. $(\mathcal{F}, \mathcal{B}(\mathcal{F}))$ is a Borel space.

c) Let $\{f_k\}_{k\in\mathbb{N}}$ be a sequence in $\mathcal{F}$ such that $f_k \to f \in \mathcal{F}$. Especially, it holds $f_k(x) \to f(x)$ for all $x \in \mathbb{R}_+$ and $Y - f_k(Y) \to Y - f(Y)$ $\mathbb{P}$-a.s. Since $Y - f_k(Y) \le Y \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ for all $k \in \mathbb{N}$, the Fatou property of π_R implies

$$\liminf_{k\to\infty} \pi_R(f_k) = \liminf_{k\to\infty} \pi_R(Y - f_k(Y)) \ge \pi_R(Y - f(Y)) = \pi_R(f).$$

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

$$\begin{aligned} |f_k(y_k) - f(y)| &= |f_k(y_k) - f_k(y) + f_k(y) - f(y)| \\ &\le |f_k(y_k) - f_k(y)| + |f_k(y) - f(y)| \\ &\le |y_k - y| + |f_k(y) - f(y)| \to 0 \qquad \text{as } k \to \infty. \end{aligned}$$

e) Due to b), we only have to consider the budget-constrained case. Since $\mathcal{F}$ is compact it suffices to show that $D(x)$ is closed. Now, $D(x) = \{f \in \mathcal{F} : \pi_R(f) \le (x)^+\}$ is closed as a sublevel set of the lower semicontinuous function $\pi_R : \mathcal{F} \to \mathbb{R}_+$, cf. Lemma A.2.

f) In case of no budget constraint this follows directly from Lemma A.15 b). If there is a budget constraint, we have to show that D is closed to obtain the assertion from part a) of the same lemma. $D(0)$ is closed by part e). From the lower semicontinuity of $\pi_R : \mathcal{F} \to \mathbb{R}_+$ it follows that the epigraph

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

is closed. Thus, $D = \{(x, f) : (f, x) \in \mathrm{epi}(\pi_R)\} \cup (\mathbb{R}_- \times D(0))$ is closed, too.

g) In the total loss case, we have $c(x, f, T(x, f, y, z)) = -T(x, f, y, z)$ which is lower semicontinuous in (x, f) by part d). In the incremental loss case, $c(x, f, T(x, f, y, z)) = x - T(x, f, y, z)$ is lower semicontinuous in (x, f) as a sum of lower semicontinuous functions. □

Hence, all assumptions of the abstract cost model of Section 3.1, especially the Continuity and Compactness Properties 3.1, are satisfied by the dynamic reinsurance model.

3.2.2. Dividends

A model similar to the one of Section 3.2.1 that additionally incorporates dividend payments was introduced by Chen and Assa (2019). However, their results contain a fundamental error in Section 4.2 disregarding the dynamic nature of the optimization problem. We give here a slightly modified version of their model which represents another special case of the abstract cost model. The solution to the optimization problem is studied in Section 4.4.1.

The insurer is again endowed with an initial capital $x \in \mathbb{R}_+$. He incurs aggregate claims Y_{n+1} at the end of each period $[n, n+1)$, $n \in \mathbb{N}_0$, and receives a deterministic premium income $z \in \mathbb{R}_+$ for the next period. Thus, the insurer's uncontrolled surplus process is given recursively by

$$X_0 = x, \qquad X_{n+1} = X_n - Y_{n+1} + z.$$

As in Section 3.2.1, the insurance company can underwrite a reinsurance treaty f_n at the beginning of each period. Moreover, it can now pay a dividend $a_n \in \mathbb{R}_+$ to its shareholders at the beginning of each period. Hence, the dynamic of the controlled surplus process is given by

$$X_0 = x, \qquad X_{n+1} = X_n + z - a_n - f_n(Y_{n+1}) - \pi_R\big(Y_{n+1} - f_n(Y_{n+1})\big).$$

The corresponding stationary Markov Decision Model has the following components:

- The state space is the real line $\mathbb{R}$ with Borel σ-algebra $\mathcal{B}(\mathbb{R})$.
- The action space is $\mathbb{R}_+ \times \mathcal{F}$ with Borel σ-algebra $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathcal{F})$. The first component represents the dividend and the second one the retained loss function.
- The disturbance space is $\mathbb{R}_+$ with Borel σ-algebra $\mathcal{B}(\mathbb{R}_+)$ and the disturbances are $(Y_n)_{n\in\mathbb{N}}$. It is assumed that the claims $(Y_n)_{n\in\mathbb{N}}$ are non-negative random variables defined on a common atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$ satisfying

$$Y_1, Y_2, \ldots \overset{\text{iid}}{\sim} Y \in L^p(\Omega, \mathcal{A}, \mathbb{P})$$

  for some $p \in [1, \infty)$.
- The transition function $T : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathcal{F} \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$T(x, a, f, y) = x + z - a - f(y) - \pi_R(f).$$

Here, $z \in \mathbb{R}_+$ is a constant representing the premium income and $\pi_R : L^p_+ \to \bar{\mathbb{R}}$ is a law-invariant, monotone and normalized premium principle having the Fatou property. Due to the identical distribution of the claims we use again the shorthand notation $\pi_R(f) = \pi_R(Y - f_n(Y))$, $f \in \mathcal{F}$.

- The admissible actions in state $x \in \mathbb{R}$ are

$$D(x) = \Big\{(a, f) \in \mathbb{R}_+ \times \mathcal{F} : \ a \leq x^+, \ \rho(f(Y)) \leq x^+ + z - a - \pi_R(f)\Big\}.$$

The normalized monetary risk measure $\rho : L^p \to \bar{\mathbb{R}}$ is required to have the Fatou property. It is assumed that there is a retained loss function $\hat{f} \in \mathcal{F}$ such that

$$\rho(\hat{f}(Y)) + \pi(\hat{f}) \leq z. \tag{3.4}$$

Therefore, the set of admissible state-action combinations $D = \{(x, a, f) \in \mathbb{R} \times \mathbb{R}_+ \times \mathcal{F} : (a, f) \in D(x)\}$ contains the graph of the constant measurable map $\mathbb{R} \ni x \mapsto (0, \hat{f})$.

- The continuous one-stage cost function $c : D \times \mathbb{R} \to \mathbb{R}$ is given by $c(x, a, f, x') = -a$, i.e. a profit for the shareholders is regarded as a negative cost.

- Under a finite planning horizon, one has the continuous terminal cost function $c_N(x) = -x^+$ meaning that all remaining capital is distributed as a dividend.

The constraint comprises two conditions. Firstly, dividends can only be paid if the insurer has a positive capital. Secondly, the capital requirement for the retained risk calculated by the risk measure ρ must not exceed the insurer's capital at the end of the respective period excluding possible claims. This is an appropriate requirement since the purpose of the solvency capital is to buffer claims arriving at the end of the period. In case of a non-positive capital, no dividend can be paid and the solvency condition is reduced to $\rho(f(Y)) \leq z - \pi_R(f)$. There is at least one reinsurance treaty $\hat{f}$ which satisfies this condition due to (3.4). This assumption means that the premium income together with the possibility to use reinsurance suffices to bear the risk of the claims. In the special case $\hat{f} = \mathrm{id}_{\mathbb{R}_+}$ (full retention), the premium income alone suffices due to the normalization of π_R.

Due to Lemma 3.3 d) it follows directly that also the transition function of the dividends model is upper semicontinuous and hence measurable. The one-stage cost $D \ni (x, a) \mapsto c(x, a, T(x, a, y)) = -a$ is continuous for every $y \in \mathbb{R}_+$. Together with the following result we can conclude that all assumptions of the abstract cost model of Section 3.1, especially the Continuity and Compactness Properties 3.1, are satisfied.

Lemma 3.4. *The set-valued mapping $R \ni x \mapsto D(x)$ is compact-valued and upper semi-continuous.*

Proof. The map $\mathcal{F} \ni f \mapsto \rho(f(Y)) + \pi_R(f)$ is lower semicontinuos by the same arguments as in the proof Lemma 3.3 c). Thus, $\phi : \mathbb{R}_+ \times \mathcal{F} \mapsto \mathbb{R}$, $(a, f) \mapsto \rho(f(Y)) + \pi_R(f) + a - z$ is lower semicontinuous as a sum of lower semicontinuous functions and by Lemma A.2 c)

$$\operatorname{epi}\phi = \{(a, f, x) \in \mathbb{R}_+ \times \mathcal{F} \times \mathbb{R} : \phi(a, f) \le x\}$$

is closed. This is the graph of the set-valued mapping

$$\mathbb{R} \ni x \mapsto D_1(x) = \{(a, f) \in \mathbb{R}_+ \times \mathcal{F} : \phi(a, f) \le x\}, \qquad x \in \mathbb{R}.$$

Obviously, the set-valued mapping

$$\mathbb{R} \ni x \mapsto D_2(x) = \{a \in \mathbb{R}_+ : a \le x\}$$

has compact values, a closed graph and is upper semicontinuous. Hence,

$$\mathbb{R} \ni x \mapsto D_3(x) = D_2(x) \times \mathcal{F}$$

has the same properties. The first two are clear and upper semicontinuity follows from Proposition A.15 a). The graph of

$$\mathbb{R} \ni x \mapsto D_4(x) = D_1(x) \cap D_3(x) = \{(a, f) \in \mathbb{R}_+ \times \mathcal{F} : a \le x,\ \phi(a, f) \le x\}$$

is the intersection of the graphs of $D_1(\cdot)$ and $D_2(\cdot)$ and hence closed, too. Moreover, $D_4(\cdot)$ is upper semicontinuous by Lemma A.14 and compact-valued since closed subsets of compact sets are compact. Finally, $D(\cdot)$ is compact-valued since $D_4(\cdot)$ is and it is upper semicontinuous as the composition of $D_4(\cdot)$ and the upper semicontinuous single-valued set-valued map $\mathbb{R} \ni x \mapsto \{x^+\}$, cf. Lemma A.17. □

Chapter 4

Distributionally Robust Expected Total Cost Minimization

Minimizing the expected total cost has evolved into the standard optimality criterion for Markov Decision Processes. Under this optimality criterion, models similar to the one introduced in Chapter 3.1 have been studied already in the 1960s and 1970s by Blackwell (1965), Hinderer (1970) and Bertsekas and Shreve (1978), only to name a few major contributions.

The novel feature here is that the transition law is no longer assumed to be fully known. In the literature this is referred to as *ambiguity* whereas *uncertainty* relates to random quantities with known distribution. In many applications, the transition law has to be estimated from historical data and is therefore subject to statistical errors. One way of dealing with this ambiguity is the *worst-case approach*, where the controller selects a policy which is optimal with respect to the most adverse transition law in each scenario. This setting can also be interpreted as a dynamic Stackelberg game with the controller as mover and nature as follower.

The worst-case approach is empirically justified by the so-called *Ellsberg Paradox*. The experiment suggested by Ellsberg (1961) has shown that agents tend to be ambiguity averse. Epstein and Schneider (2003) investigated the question whether ambiguity aversion can be incorporated in an axiomatic model of intertemporal utility. The representation of the preferences turned out to be some worst case expected utility, i.e. the minimal expected utility over an appropriate set of probability measures. This set of probability measures needs to satisfy some rectangularity condition for the utility to have a recursive structure making it time consistent.

The rectangularity property has been taken up by Iyengar (2005) as a key assumption for being able to derive a Bellman equation for a distributionally robust MDP with countable state and action spaces. Contemporaneously, Nilim and El Ghaoui (2005) reached similar findings, however, limited to finite state and action spaces. In the following, we will generalize the results of Iyengar (2005) to a model with general Borel spaces. In order to deal with the arising measurability issues, we borrow from the dynamic game setup in González-Trejo et al. (2002) and Jaśkiewicz and Nowak (2011). The major difference between our contribution and these two works is the design of the distributional ambiguity. We replace the topology of convergence in distribution on the ambiguity set by the weak* topology $\sigma(L^q, L^p)$ in order to obtain connections to recursive risk measures in Section 5.3. Moreover, we rigorously derive a Bellman equation.

4.1. Finite Planning Horizon

We consider the non-stationary version of the abstract cost model of Section 3.1 under a finite planning horizon $N \in \mathbb{N}$. Let $p \in [1, \infty)$ with conjugate index $q \in (1, \infty]$. Due to the independence of the disturbances, we may without loss of generality assume that the probability space has a product structure

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{n=1}^{\mathbb{N}} (\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$$

with $Z_n(\bar{\omega}) = Z_n(\omega_n)$ only depending on component ω_n of $\bar{\omega} = (\omega_1, \dots, \omega_N) \in \Omega$ for $n = 1, \dots, N$. When writing shorthand $Z_n(\omega)$ for $Z_n(\omega_n)$ we mean by ω the component of $\bar{\omega} = (\omega_1, \dots, \omega_N)$ that Z_n actually depends on.

One may think of $(\Omega, \mathcal{A}, \mathbb{P})$ as the canonical construction, i.e.

$$(\Omega_n, \mathcal{A}_n, \mathbb{P}_n) = \left(\mathcal{Z}, \mathfrak{Z}, \mathbb{P}^{Z_n}\right) \qquad \text{and} \qquad Z_n(\bar{\omega}) = \omega_n, \quad \bar{\omega} = (\omega_1, \dots, \omega_N) \in \Omega$$

for all $n = 1, \dots, N$. In the sequel, we will require $\mathbb{P}_n$ to be separable (see Appendix B.2 for a definition). Additionally, we will assume for some results that $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ is atomless in order to support a generalized distributional transform. Hence, a canonical construction entails constraints on the choice of the disturbance space, cf. Appendix B.2.

Let $n \in \{0, \dots, N-1\}$ be a stage of the decision process. Due to the product structure of $(\Omega, \mathcal{A}, \mathbb{P})$, the representation of the transition kernel in (3.1) simplifies to

$$Q_n(B|x, a) = \int \mathbb{1}_B(T_n(x, a, Z_{n+1}(\omega)))\mathbb{P}_{n+1}(\mathrm{d}\,\omega), \quad B \in \mathcal{B}(E),\ (x, a) \in D_n. \tag{4.1}$$

We denote by $\mathcal{M}_1(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ the set of probability measures on $(\Omega_n, \mathcal{A}_n)$ which are absolutely continuous with respect to $\mathbb{P}_n$ and define

$$\mathcal{M}_1^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n) = \left\{\mathbb{Q} \in \mathcal{M}_1(\Omega_n, \mathcal{A}_n, \mathbb{P}_n) : \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_n} \in L^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)\right\}.$$

Henceforth, we fix a non-empty subset $\mathcal{Q}_n \subseteq \mathcal{M}_1^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ which is referred to as *ambiguity set* at stage n. Due to absolute continuity, we can identify $\mathcal{Q}_n$ with the set of corresponding densities w.r.t. $\mathbb{P}_n$

$$\mathcal{Q}_n^d = \left\{ \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_n} \in L^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n) : \mathbb{Q} \in \mathcal{Q}_n \right\}.$$

Accordingly, we view $\mathcal{Q}_n$ as a subset of $L^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ and endow it with the trace topolgy of the weak* topolgy $\sigma(L^q, L^p)$ on $L^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$. The weak* topology in turn induces a Borel σ-algebra on $\mathcal{Q}_n$ making it a measurable space.

Lemma 4.1. *Let the ambiguity set be norm-bounded and the probability measure $\mathbb{P}_n$ on $(\Omega_n, \mathcal{A}_n)$ be separable. Then $\mathcal{Q}_n$ endowed with the weak* topology $\sigma(L^q, L^p)$ is a separable metrizable space. If $\mathcal{Q}_n$ is additionally weak* closed, it is even a compact Borel space.*

Proof. Recall that we identify $\mathcal{Q}_n$ with the set of the corresponding densities $\mathcal{Q}_n^d$. The closure $\overline{\mathcal{Q}}_n^d$ of $\mathcal{Q}_n^d$ remains norm bounded. This can be seen as follows: Let $X \in \overline{\mathcal{Q}}_n^d$. Then there exists a net $\{X_\alpha\}_{\alpha \in I} \subseteq \mathcal{Q}_n^d$ such that $X_\alpha \stackrel{w^*}{\to} X$. Hence,

$$\mathbb{E}[X_\alpha Y] \to \mathbb{E}[XY] \qquad \text{for all } Y \in L^p(\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \text{ with } \|Y\|_{L^p} = 1.$$

By Hölder's inequality we have for all $\alpha \in I$

$$|\mathbb{E}[X_\alpha Y]| \leq \mathbb{E}|X_\alpha Y| \leq \|X_\alpha\|_{L^q} \|Y\|_{L^p} = \|X_\alpha\|_{L^q} \leq K.$$

Thus, $|\mathbb{E}[XY]| \leq K$. Finally, due to duality it follows

$$\|X\|_{L^q} = \sup_{\|Y\|_{L^p}=1} |\mathbb{E}[XY]| \leq K.$$

The separability of the probability measure $\mathbb{P}_n$ makes $L^p(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ a separable Banach space, cf. Lemma B.6. Consequently, the weak* topology is metrizable on the norm bounded set $\overline{\mathcal{Q}}_n^d$ (Morrison; 2001, p. 157). The trace topology on the subset $\mathcal{Q}_n^d \subseteq \overline{\mathcal{Q}}_n^d$ coincides with the topology induced by the restriction of the metric (Ó Searcóid; 2007, 4.4.1), i.e. $\mathcal{Q}_n^d$ is metrizable, too.

Since $\overline{\mathcal{Q}}_n^d$ is norm bounded and weak* closed, the Theorem of Banach-Alaoglu (Aliprantis and Border; 2006, 6.21) yields that it is weak* compact. As a compact metrizable space $\overline{\mathcal{Q}}_n^d$ is complete (Aliprantis and Border; 2006, 3.28) and also separable (Aliprantis and Border; 2006, 3.26). Hence, $\overline{\mathcal{Q}}_n^d$ is a Borel Space. The set of densities $\mathcal{Q}_n^d$ is also separable as a subspace of a separable metrizable space (Aliprantis and Border; 2006, 3.5). □

In our abstract cost model, we allow for any norm-bounded ambiguity set $\mathcal{Q}_n \subseteq \mathcal{M}_1^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$. For applications, a meaningful way of choosing $\mathcal{Q}_n$ (within a norm bound) is to take all probability measures in $\mathcal{M}_1^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ which are in some sense close to $\mathbb{P}_n$. In an insurance context, Birghila and Pflug (2019) recently suggested taking either the convex hull of a finite number of probability measures or a neighborhood of the reference probability measure w.r.t. the *(contorted) Wasserstein distance.* The latter approach may be extended to any metric for probability measures. In our setting, that requires absolute continuity, the *Kullback–Leibler divergence*

$$D_{\mathrm{KL}}(\mathbb{Q}\|\mathbb{P}_n) = \int \log\left(\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_n}\right)\mathrm{d}\,\mathbb{Q}, \qquad \mathbb{Q} \in \mathcal{M}_1^q(\Omega_n, \mathcal{A}_n, \mathbb{P}_n),$$

is a natural choice.

For $n = 0, \dots, N-1$ we denote by $\mathcal{H}_n^* = D_0 \times \cdots \times D_n$ the set of *extended feasible histories* of the decision process up to time n. A generic element of $\mathcal{H}_n^*$ has the form $h_n^* = (x_0, a_0, \dots, x_n, a_n)$. The controller only knows that the transition kernel (4.1) at each stage is defined by some $\mathbb{Q} \in \mathcal{Q}_{n+1}$ instead of $\mathbb{P}_{n+1}$ but not which one exactly. From the perspective of a dynamic game against nature this means that nature reacts to the controller's action a_n in scenario $h_n \in \mathcal{H}_n$ with a decision rule $\gamma_n : \mathcal{H}_n^* \to \mathcal{Q}_{n+1}$. A *policy of nature* is a sequence of such decision rules $\gamma = (\gamma_0, \dots, \gamma_{N-1})$. Let Γ be the set of all policies of nature. Since nature is an unobserved theoretical opponent of the controller, her actions are not considered to be part of the history of the decision process. A *Markov decision rule of nature* at time n is a measurable mapping $\gamma_n : D_n \to \mathcal{Q}_{n+1}$ and a *Markov policy of nature* is a sequence $\gamma = (\gamma_0, \dots, \gamma_{N-1})$ of such decision rules. The set of Markov policies of nature is denoted by Γ^M.

Lemma 4.2. *For $n = 0, \dots, N-1$ a decision rule $\gamma_n : \mathcal{H}_n^* \to \mathcal{Q}_{n+1}$ induces a stochastic kernel from H_n^* to Ω_{n+1} by*

$$\gamma_n(B|h_n^*) = \gamma_n(h_n^*)(B), \qquad B \in \mathcal{A}_{n+1},\ h_n^* \in \mathcal{H}_n^*.$$

Proof. By definition, $\gamma_n(\cdot|h_n^*)$ is a probability measure for every $h_n^* \in \mathcal{H}_n^*$. Now fix $B \in \mathcal{A}_{n+1}$. The map $\delta : \mathcal{Q}_{n+1} \to [0,1]$, $\delta(\mathbb{Q}) = \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_B]$ is weak* continuous since $\mathbb{1}_B \in L^p(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$ and hence Borel measurable. Therefore,

$$\mathcal{H}_n^* \ni h_n^* \mapsto \gamma_n(B|h_n^*) = \delta \circ \gamma_n(h_n^*)$$

is measurable as a composition of measurable maps. □

In the sequel, it will be clear from the context where we refer to γ_n as a decision rule or as a stochastic kernel. Lemma 4.2 is a well-known result, of which even the converse holds for probability measures on Borel spaces, cf. Proposition 7.26 in Bertsekas and Shreve (1978). However, here this is not applicable since firstly $(\Omega_{n+1}, \mathcal{A}_{n+1})$ can be any measurable space and secondly the Borel σ-algebra on $\mathcal{Q}_{n+1}$ is induced by the weak* topology $\sigma(L^q, L^p)$ and not the topology of convergence in distribution.

The probability measure $\gamma_n(\cdot|h_n^*)$, which is unknown for the controller, now takes the role of $\mathbb{P}_{n+1}$ in defining the transition kernel of the decision process in (4.1). Let

$$Q_n^\gamma(B|h_n^*) = \int \mathbb{1}_B\big(T_n(x_n, a_n, Z_{n+1}(\omega))\big)\gamma_n(\mathrm{d}\,\omega|h_n^*), \qquad B \in \mathcal{B}(E),\ h_n^* \in \mathcal{H}_n^*.$$

As in the case without ambiguity, the Theorem of Ionescu-Tulcea (Klenke; 2014, 14.32) yields that each starting point $x \in E$ and pair of policies of the controller and nature $(\pi, \gamma) \in \Pi^R \times \Gamma$ induce a unique law of motion

$$\mathbb{Q}_x^{\pi\gamma} = \delta_x \otimes \pi_0 \otimes Q_0^\gamma \otimes \pi_1 \otimes Q_1^\gamma \otimes \dots \tag{4.2}$$

on $\overline{\mathcal{H}}_\infty$ satisfying

$$\begin{aligned} \mathbb{Q}_x^{\pi\gamma}(X_0 \in B) &= \delta_x(B), \\ \mathbb{Q}_x^{\pi\gamma}(A_n \in C|H_n = h_n) &= \pi_n(C|h_n), \\ \mathbb{Q}_x^{\pi\gamma}(X_{n+1} \in B|H_n^* = h_n^*) &= Q_n^\gamma(B|h_n^*) \end{aligned}$$

for all $B \in \mathcal{B}(E)$ and $C \in \mathcal{B}(A)$. In the usual way, we denote with $\mathbb{E}_x^{\pi\gamma}$ the expectation operator with respect to $\mathbb{Q}_x^{\pi\gamma}$ and with $\mathbb{E}_{nh_n}^{\pi\gamma}$ or $\mathbb{E}_{nx}^{\pi\gamma}$ the respective conditional expectation given $H_n = h_n$ or $X_n = x$.

The value of a policy pair $(\pi, \gamma) \in \Pi^R \times \Gamma$ at time $n = 0, \dots, N$ is defined as

$$\begin{aligned} V_{N\pi\gamma}(h_N) &= c_N(x_N), & h_N &\in \mathcal{H}_N, \\ V_{n\pi\gamma}(h_n) &= \mathbb{E}_{nh_n}^{\pi\gamma}\left[\sum_{k=n}^{N-1} c_k(X_k, A_k, X_{k+1}) + c_N(X_N)\right], & h_n &\in \mathcal{H}_n. \end{aligned} \tag{4.3}$$

Since the controller is unaware which probability measure in the ambiguity set determines the transition law in each scenario, it is prudential to minimize the expected cost under the assumption to be confronted with the most adverse probability measure. The value functions are thus given by

$$V_n(h_n) = \inf_{\pi \in \Pi^R} \sup_{\gamma \in \Gamma} V_{n\pi\gamma}(h_n), \qquad h_n \in \mathcal{H}_n,$$

and this section's optimization objective is

$$V_0(x) = \inf_{\pi \in \Pi^R} \sup_{\gamma \in \Gamma} V_{0\pi\gamma}(x), \qquad x \in E. \tag{4.4}$$

In game-theoretic terminology this is the *upper value of a dynamic zero-sum game.* If nature were to act first, i.e. if infimum and supremum were interchanged, one would obtain the game's *lower value.* If the two values agree and the infimum and supremum are attained, the games has a *Nash equilibrium*, see also Section 4.3.1.

Iyengar (2005) does not model nature to make active decisions, but instead defines the set of all possible laws of motion. When each law of motion is of the form (4.2), he calls the set *rectangular*. Shapiro (2016) devotes an entire paper to the rectangularity property. Our approach with active decisions of nature, based on González-Trejo et al. (2002) and Jaśkiewicz and Nowak (2011), is needed to construct stochastic kernels as in Lemma 4.2 with probability measures from a given ambiguity set. When state and action spaces are countable as in Iyengar (2005), the technical problem of measurability does not arise and one can directly construct an ambiguous law of motion by simply multiplying conditional probabilities.

Our model feature that there is no ambiguity in the transition functions is justified in many applications. Typically, transition functions describe a technical process or economic calculation (e.g. the calculation of the insurer's surplus in Section 3.2) which is known ex-ante and does not have to be estimated. The same applies to the cost function.

In order to have well-defined value functions, we need some integrability condition. Together with all other assumptions of this section, it is listed here.

Assumption 4.3. (i) The model data has the Continuity and Compactness Properties 3.1 with the transition function T_n being continuous in (x,a) for all $n = 0, \dots, N-1$ (case 1).

(ii) There exist $\alpha, \underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$ and measurable functions $\underline{b} : E \to (-\infty, -\underline{\epsilon}]$, $\bar{b} : E \to [\bar{\epsilon}, \infty)$ such that it holds for all $n = 0, \dots, N-1$, $\mathbb{Q} \in \mathcal{Q}_{n+1}$ and $(x,a) \in D_n$

$$\mathbb{E}^{\mathbb{Q}}\left[-c_n^-(x,a,T_n(x,a,Z_{n+1}))\right] \geq \underline{b}(x), \qquad \mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T_n(x,a,Z_{n+1}))\right] \geq \alpha \underline{b}(x),$$
$$\mathbb{E}^{\mathbb{Q}}\left[c_n^+(x,a,T_n(x,a,Z_{n+1}))\right] \leq \bar{b}(x), \qquad \mathbb{E}^{\mathbb{Q}}\left[\bar{b}(T_n(x,a,Z_{n+1}))\right] \leq \alpha \bar{b}(x).$$

Furthermore, it holds $\underline{b}(x) \leq c_N(x) \leq \bar{b}(x)$ for all $x \in E$.

(iii) We define $b : E \to [1, \infty)$, $b(x) = \bar{b}(x) - \underline{b}(x)$. For all $n = 0, \dots, N-1$ and $(\bar{x}, \bar{a}) \in D_n$ there exists an $\epsilon > 0$ and measurable functions $\Theta_{n,1}^{\bar{x},\bar{a}}, \Theta_{n,2}^{\bar{x},\bar{a}} : \mathcal{Z} \to \mathbb{R}_+$ such that $\Theta_{n,1}^{\bar{x},\bar{a}}(Z_{n+1}), \Theta_{n,2}^{\bar{x},\bar{a}}(Z_{n+1}) \in L^p(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$ and

$$|c_n(x,a,T_n(x,a,z))| \leq \Theta_{n,1}^{\bar{x},\bar{a}}(z), \qquad b(T_n(x,a,z)) \leq \Theta_{n,2}^{\bar{x},\bar{a}}(z)$$

for all $z \in \mathcal{Z}$ and $(x,a) \in B_\epsilon(\bar{x}, \bar{a}) \cap D_n$. Here, $B_\epsilon(\bar{x}, \bar{a})$ is the closed ball around $(\bar{x}, \bar{a})$ w.r.t. an arbitrary product metric on $E \times A$.

(iv) The probability measure $\mathbb{P}_n$ on $(\Omega_n, \mathcal{A}_n)$ is separable for all $n = 1, \dots, N$.

(v) The ambiguity sets $\mathcal{Q}_n$ are norm bounded, i.e. there exists $K \in [1, \infty)$ such that

$$\mathbb{E}\left|\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_n}\right|^q \leq K$$

for all $\mathbb{Q} \in \mathcal{Q}_n$ and $n = 1, \dots, N$.

The next remark summarizes some notes on the model assumptions.

Remark 4.4. a) $\underline{b}, \bar{b}$ are called *lower* and *upper bounding function*, respectively, while b is referred to as *bounding function*. As the absolute value is the sum of positive and negative part, b satisfies

$$\mathbb{E}^{\mathbb{Q}}\left[|c_n(x,a,T_n(x,a,Z_{n+1}))|\right] \le b(x) \qquad \text{and} \qquad \mathbb{E}^{\mathbb{Q}}\left[|b(T_n(x,a,Z_{n+1}))|\right] \le \alpha b(x)$$

for all $n = 0, \dots, N-1$, $\mathbb{Q} \in \mathcal{Q}_{n+1}$ and $(x,a) \in D_n$.

b) Assumptions 4.3 (ii) and (iii) are satisfied if the cost functions are bounded.

c) If $p = 1$ and hence $q = \infty$, it is technically sufficient if part (ii) of Assumption 4.3 holds under the reference probability measure $\mathbb{P}_n$. Using Hölder's inequality and part (v) we get for every $\mathbb{Q} \in \mathcal{Q}_{n+1}$

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[-c_n^-(x,a,T_n(x,a,Z_{n+1}))\right] &\ge \mathbb{E}\left[-c_n^-(x,a,T_n(x,a,Z_{n+1}))\right] \operatorname{ess\,sup} \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_{n+1}} \\ &\ge K\underline{b}(x), \\ \mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T_n(x,a,Z_{n+1}))\right] &\ge \mathbb{E}\left[\underline{b}(T_n(x,a,Z_{n+1}))\right] \operatorname{ess\,sup} \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_{n+1}} \\ &\ge \alpha K \underline{b}(x)\end{aligned}$$

and analogous results for the upper bounding function. I.e. one simply has to replace α by $K\alpha$. However, the factor $K\alpha$ may be unnecessarily crude. Since its magnitude matters under an infinite planning horizon (Section 4.2), we allow for a better estimate.

d) Separability of a finite measure is defined in Appendix B.2. For probability measures a countably generated σ-algebra is sufficient (Lemma B.5). The Borel σ-algebra of a second countable topological space, e.g. a separable metric space, has this property.

The next lemma shows that due to Assumption 4.3 (ii) the value (4.3) of a policy pair $(\pi, \gamma) \in \Pi^R \times \Gamma$ is well-defined at all stages $n = 0, \dots, N$. One can see that the existence of either a lower or an upper bounding function is sufficient for the policy value to be well-defined since the integral exists if either the negative or positive part of the integrand is integrable. However, for the existence of an optimal policy pair we will need the integral to exist with finite value and therefore require both a lower and an upper bounding function.

Lemma 4.5. *Under Assumption 4.3 it holds for all policy pairs $(\pi, \gamma) \in \Pi^R \times \Gamma$, time points $n = 0, \dots, N$ and histories $h_n \in \mathcal{H}_n$*

(i)

$$V_{n\pi\gamma}(h_n) \ge \mathbb{E}^{\pi\gamma}_{nh_n}\left[\sum_{k=n}^{N-1} -c_k^-(X_k, A_k, X_{k+1}) - c_N^-(X_N)\right] \ge \sum_{k=n}^{N} \alpha^{k-n} \underline{b}(x_n).$$

(ii)

$$V_{n\pi\gamma}(h_n) \le \mathbb{E}^{\pi\gamma}_{nh_n}\left[\sum_{k=n}^{N-1} c_k^+(X_k, A_k, X_{k+1}) + c_N^+(X_N)\right] \le \sum_{k=n}^{N} \alpha^{k-n} \bar{b}(x_n).$$

(iii)

$$|V_{n\pi\gamma}(h_n)| \le \mathbb{E}^{\pi\gamma}_{nh_n}\left[\sum_{k=n}^{N-1} |c_k(X_k, A_k, X_{k+1})| + |c_N(X_N)|\right] \le \sum_{k=n}^{N} \alpha^{k-n} b(x_n).$$

Proof. We only prove (i). Part (ii) is analogous and part (iii) follows from combining the first two parts. The first inequality is obvious. Regarding the second one we use that

$$\mathbb{E}^{\pi\gamma}_{nh_n}\left[\sum_{k=n}^{N-1} -c_k^-(X_k, A_k, X_{k+1}) - c_N^-(X_N)\right] = \sum_{k=n}^{N-1} \mathbb{E}^{\pi\gamma}_{nh_n}\left[-c_k^-(X_k, A_k, X_{k+1})\right] + \mathbb{E}^{\pi\gamma}_{Nh_N}\left[-c_N^-(X_N)\right]$$

and consider the summands individually. We have $\mathbb{E}^{\pi\gamma}_{Nh_N}\left[-c_N^-(X_N)\right] \ge \mathbb{E}^{\pi\gamma}_{Nh_N}\left[\underline{b}(X_N)\right]$ by Assumption 4.3 (ii). Since γ_k is a mapping to $\mathcal{Q}_{k+1}$ it follows from the first inequality of Assumption 4.3 (ii) that

$$\begin{aligned}
&\mathbb{E}^{\pi\gamma}_{nh_n}\left[-c_k^-(X_k, A_k, X_{k+1})\right] \\
&= \int \mathbb{E}^{\pi\gamma}_{kh_k}\left[-c_k^-(X_k, A_k, X_{k+1})\right] \mathbb{Q}^{\pi\gamma}_x(\mathrm{d}\, h_k | H_n = h_n) \\
&= \iiint -c_k^-\left(x_k, a_k, T_k(x_k, a_k, Z_{k+1}(\omega))\right) \gamma_k(\mathrm{d}\,\omega | h_k^*)\, \pi_k(\mathrm{d}\, a_k | h_k)\, \mathbb{Q}^{\pi\gamma}_x(\mathrm{d}\, h_k | H_n = h_n) \\
&\ge \int \underline{b}(x_k)\, \mathbb{Q}^{\pi\gamma}_x(\mathrm{d}\, h_k | H_n = h_n) \\
&= \mathbb{E}^{\pi\gamma}_{nh_n}\left[\underline{b}(X_k)\right]
\end{aligned}$$

for $k = n, \dots, N-1$. Now, the second inequality of Assumption 4.3 (ii) yields for $k \ge n+1$

$$\begin{aligned}
&\mathbb{E}^{\pi\gamma}_{nh_n}\left[\underline{b}(X_k)\right] \\
&= \int \mathbb{E}_{k-1h_{k-1}}\left[\underline{b}(X_k)\right] \mathbb{Q}^{\pi\gamma}_x(\mathrm{d}\, h_{k-1} | H_n = h_n) \\
&= \iiint \underline{b}(T_{k-1}(x_{k-1}, a_{k-1}, Z_k(\omega)))\, \gamma_{k-1}(\mathrm{d}\omega | h_{k-1}^*)\, \pi_{k-1}(\mathrm{d}a_{k-1} | h_{k-1})\, \mathbb{Q}^{\pi\gamma}_x(\mathrm{d}h_{k-1} | H_n = h_n) \\
&\ge \alpha \int \underline{b}(x_{k-1})\, \mathbb{Q}^{\pi\gamma}_x(\mathrm{d}\, h_{k-1} | H_n = h_n) \\
&= \alpha \mathbb{E}^{\pi\gamma}_{nh_n}\left[\underline{b}(X_{k-1})\right].
\end{aligned}$$

Iterating this argument, one obtains

$$\mathbb{E}^{\pi\gamma}_{nh_n}\left[-c_N^-(X_N)\right] \ge \alpha^{N-n} \underline{b}(x_n) \qquad \text{and} \qquad \mathbb{E}^{\pi\gamma}_{nh_n}\left[-c_k^-(X_k, A_k, X_{k+1})\right] \ge \alpha^{k-n} \underline{b}(x_n).$$

Finally, summation over k yields

$$\mathbb{E}^{\pi\gamma}_{nh_n}\left[\sum_{k=n}^{N-1} -c_k^-(X_k, A_k, X_{k+1}) - c_N^-(X_N)\right] \ge \sum_{k=n}^{N} \alpha^{k-n} \underline{b}(x_n)$$

as claimed. □

Having ensured that the policy values are well-defined, we can now proceed to deriving a value iteration.

Proposition 4.6. *Under Assumption 4.3 the value of a policy pair* $(\pi, \gamma) \in \Pi^R \times \Gamma$ *can be calculated recursively for* $n = 0, \dots, N$ *and* $h_n \in \mathcal{H}_n$ *as*

$$\begin{aligned} V_{N\pi\gamma}(h_N) &= c_N(x_N), \\ V_{n\pi\gamma}(h_n) &= \iint c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ &\qquad + V_{n+1\pi\gamma}(h_n^*, T(x_n, a_n, Z_{n+1}(\omega)))\, \gamma_n(\mathrm{d}\,\omega | h_n^*)\, \pi_n(\mathrm{d}\,a_n | h_n). \end{aligned}$$

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 at time n

$$\begin{aligned} & V_{n\pi\gamma}(h_n) \\ &= \mathbb{E}^{\pi\gamma}_{nh_n}\left[\sum_{k=n}^{N-1} c_k(X_k, A_k, X_{k+1}) + c_N(X_N)\right] \\ &= \mathbb{E}^{\pi\gamma}_{nh_n}\left[c_n(X_n, A_n, X_{n+1}) + \sum_{k=n+1}^{N-1} c_k(X_k, A_k, X_{k+1}) + c_N(X_N)\right] \\ &= \mathbb{E}^{\pi\gamma}_{nh_n}\left[c_n(X_n, A_n, X_{n+1}) + \mathbb{E}^{\pi\gamma}_{n+1h_nA_nX_{n+1}}\left[\sum_{k=n+1}^{N-1} c_k(X_k, A_k, X_{k+1}) + c_N(X_N)\right]\right] \\ &= \iint c_n(x_n, a_n, x_{n+1}) \\ &\qquad + \mathbb{E}^{\pi\gamma}_{n+1h_na_nx_{n+1}}\left[\sum_{k=n+1}^{N-1} c_k(X_k, A_k, X_{k+1}) + c_N(X_N)\right] Q^\gamma_n(\mathrm{d}\,x_{n+1}|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n) \\ &= \iint c_n(x_n, a_n, x_{n+1}) + V_{n+1\pi\gamma}(h_{n+1})\ Q^\gamma_n(\mathrm{d}\,x_{n+1}|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n) \\ &= \iint c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ &\qquad + V_{n+1\pi\gamma}(h_n^*, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \gamma_n(\mathrm{d}\,\omega|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n) \end{aligned}$$

for all $h_n \in \mathcal{H}_n$. □

With the bounding function b we define the function space

$$\mathbb{B}_b = \{v : E \to \mathbb{R} \mid v \text{ measurable with } \lambda \in \mathbb{R}_+ \text{ s.t. } |v(x)| \le \lambda\, b(x) \text{ for all } x \in E\}.$$

Endowing $\mathbb{B}_b$ with the weighted supremum norm

$$\|v\|_b = \sup_{x \in E} \frac{|v(x)|}{b(x)}$$

makes $(\mathbb{B}_b, \|\cdot\|_b)$ a Banach space, cf. Proposition 7.2.1 in Hernández-Lerma and Lasserre (1999). The following consequence of Assumption 4.3 (iii) is needed in several proofs.

Lemma 4.7. *Let $v \in \mathbb{B}_b$ and $n \in \{0, \dots, N-1\}$. Under Assumption 4.3 (iii) each sequence of random variables*

$$C_k = c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})) + v(T_n(x_k, a_k, Z_{n+1}))$$

induced by a convergent sequence $\{(x_k, a_k)\}_{k\in\mathbb{N}}$ in D_n has an L^p-bound $\bar{C}$, i.e. $|C_k| \le \bar{C} \in L^p(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$ for all $k \in \mathbb{N}$.

Proof. There exists a constant $\lambda \in \mathbb{R}_+$ such that $|v| \le \lambda b$. Since D_n is closed by Lemma A.16, the limit point (x_0, a_0) of $\{(x_k, a_k)\}_{k\in\mathbb{N}}$ lies in D_n. Let $\epsilon > 0$ be the constant from Assumption 4.3 (iii) corresponding to (x_0, a_0). Since the sequence is convergent, there exists $m \in \mathbb{N}$ such that $(x_k, a_k) \in B_\epsilon(x_0, a_0) \cap D_n$ for all $k > m$. For the finite number of points $(x_0, a_0), (x_1, a_1), \dots, (x_m, a_m)$ there exist bounding functions $\Theta_{n,1}^{x_i,a_i}, \Theta_{n,2}^{x_i,a_i}$ by Assumption 4.3 (iii). Thus, the random variable

$$\bar{C} = \max_{i=0,\dots,m} \left(\Theta_{n,1}^{x_i,a_i}(Z_{n+1}) + \lambda \Theta_{n,2}^{x_i,a_i}(Z_{n+1}) \right)$$

is an L^p-bound as desired. □

Now, we evaluate a policy of the controller under the worst-case scenario regarding nature's reaction. We define the *robust value of a policy* $\pi \in \Pi^R$ at time $n = 0, \dots, N$ as

$$V_{n\pi}(h_n) = \sup_{\gamma\in\Gamma} V_{n\pi\gamma}(h_n), \qquad h_n \in \mathcal{H}_n.$$

To minimize this quantity is the controller's optimization objective. For the robust policy value, a value iteration holds, too. With regard to a policy of nature this is a Bellman equation given a fixed policy of the controller.

Theorem 4.8. *Let Assumption 4.3 be satisfied.*

a) The robust value of a policy $\pi \in \Pi^R$ is a measurable function of $h_n \in \mathcal{H}_n$ for $n = 0, \dots, N$. It can be calculated recursively as

$$\begin{aligned} V_{N\pi}(h_N) &= c_N(x_N), \\ V_{n\pi}(h_n) &= \int \sup_{\mathbb{Q}\in\mathcal{Q}_{n+1}} \int c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ &\qquad + V_{n+1\pi}(h_n^*, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \mathbb{Q}(\mathrm{d}\,\omega)\, \pi_n(\mathrm{d}\, a_n|h_n). \end{aligned}$$

b) If the ambiguity set $\mathcal{Q}_{n+1}$ is weak closed, there exists a maximizing decision rule γ_n^* of nature at time $n = 0, \dots, N-1$, i.e.*

$$V_{n\pi}(h_n) = \iint c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega)))$$

$$+ V_{n+1\pi}(h_n^*, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \gamma_n^*(\mathrm{d}\,\omega|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n).$$

Each sequence of such decision rules $\gamma^* = (\gamma_1^*, \dots, \gamma_{N-1}^*) \in \Gamma$ *is an optimal response of nature to the controller's policy in the sense that*

$$V_{n\pi} = V_{n\pi\gamma^*}, \qquad n = 0, \dots, N-1.$$

Proof. a) The proof is by backward induction. At time N there is nothing to show. Now assume the assertion holds at time $n+1$, i.e. that $V_{n+1\pi}$ is measurable and that for every $\epsilon > 0$ there exists an ϵ-optimal strategy $\hat{\gamma} = (\hat{\gamma}_{n+1}, \dots, \hat{\gamma}_{N-1})$ of nature. By Proposition 4.6 we have at time n

$$\begin{aligned}
&V_{n\pi}(h_n) \\
&= \sup_{\gamma\in\Gamma} V_{n\pi\gamma}(h_n) \\
&= \sup_{\gamma\in\Gamma} \iint c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\
&\qquad\qquad + V_{n+1\pi\gamma}(h_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \gamma_n(\mathrm{d}\,\omega|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n) \\
&\le \sup_{\gamma\in\Gamma} \iint c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\
&\qquad\qquad + V_{n+1\pi}(h_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \gamma_n(\mathrm{d}\,\omega|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n).
\end{aligned}$$

Given action $a_n \in D_n(x_n)$ the maximization only depends on $\gamma_n(\cdot|h_n^*) \in \mathcal{Q}_{n+1}$. Assuming measurability one can estimate the integrand by

$$\begin{aligned}
\le \int \sup_{\mathbb{Q}\in\mathcal{Q}_{n+1}} \int & c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\
&+ V_{n+1\pi}(h_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \mathbb{Q}(\mathrm{d}\,\omega)\, \pi_n(\mathrm{d}\,a_n|h_n). \qquad (4.5)
\end{aligned}$$

Let $\epsilon > 0$ be arbitrary. Given the existence of a measurable $\frac{\epsilon}{2}$-maximizer $\hat{\gamma}_n : \mathcal{H}_n^* \to \mathcal{Q}_{n+1}$ we have the inequality

$$\begin{aligned}
\le \iint & c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\
&+ V_{n+1\pi}(h_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \hat{\gamma}_n(\mathrm{d}\,\omega|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n) + \frac{\epsilon}{2}. \qquad (4.6)
\end{aligned}$$

By the induction hypothesis, this is bounded by

$$\begin{aligned}
\le \iint & c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\
&+ V_{n+1\pi\hat{\gamma}}(h_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \hat{\gamma}_n(\mathrm{d}\,\omega|h_n^*)\, \pi_n(\mathrm{d}\,a_n|h_n) + \epsilon.
\end{aligned}$$

Again by Proposition 4.6, it equals

$$= V_{n\pi\hat{\gamma}}(h_n) + \epsilon$$
$$\leq V_{n\pi}(h_n) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, equality holds. It remains to show the measurability of the outer integrand at (4.5) and the existence of an $\frac{\epsilon}{2}$-maximizer at (4.6). Our aim is to apply the general result by Rieder (1978) on optimal measurable selection stated in Theorem A.22. To that end, we first show that the function

$$f(h_n^*, \mathbb{Q}) = \int c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1})) + V_{n+1\pi}(h_n^*, T_n(x_n, a_n, Z_{n+1}))\, \mathrm{d}\,\mathbb{Q},$$

$h_n^* \in \mathcal{H}_n^*,\ \mathbb{Q} \in \mathcal{Q}_{n+1}$, is jointly measurable. The integrand is a measurable function of (h_n^*, ω) as a composition of measurable functions and in $L^p(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$ for every $h_n^* \in \mathcal{H}_n^*$ by Lemma 4.5 (iii). Hence, Fubini's Theorem B.2 yields that $h_n^* \mapsto f(h_n^*, \mathbb{Q})$ is measurable for every $\mathbb{Q} \in \mathcal{Q}_{n+1}$ and by the definition of the weak* topology $\mathbb{Q} \mapsto f(h_n^*, \mathbb{Q})$ is continuous for every $h_n^* \in \mathcal{H}_n^*$. I.e. f is a Carathéodory function. Since $\mathcal{Q}_{n+1}$ is a separable metrizable space (see Lemma 4.1), Lemma 4.51 in Aliprantis and Border (2006) yields that f is jointly measurable. Consequently,

$$\{(h_n^*, \mathbb{Q}) \in \mathcal{H}_n^* \times \mathcal{Q}_{n+1} : f(h_n^*, \mathbb{Q}) \geq \eta\} \in \{S \times Q : S \in \mathcal{B}(\mathcal{H}_n^*),\ Q \subseteq \mathcal{Q}_{n+1}\}.$$

for every $\eta \in \mathbb{R}$. The right hand side is a selection class as defined in Appendix A.3. Obviously, it holds

$$\mathcal{H}_n^* \times \mathcal{Q}_{n+1} \in \{S \times Q : S \in \mathcal{B}(\mathcal{H}_n^*),\ Q \subseteq \mathcal{Q}_{n+1}\}.$$

Now, Theorem A.22 yields that

$$\mathcal{H}_n^* \ni h_n^* \mapsto \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} f(h_n^*, \mathbb{Q})$$

is measurable and for every $\epsilon > 0$ there exists an ϵ-maximizer $\gamma_n : \mathcal{H}_n^* \to \mathcal{Q}_{n+1}$.

b) We have to show that there exists not only a ϵ-maximizer at (4.6) but a maximizer. This follows from Theorem A.23. The additional requirements are that $\mathcal{Q}_{n+1}$ is a separable metrizable space, which holds by Lemma 4.1, and that the set

$$\{\mathbb{Q} \in \mathcal{Q}_{n+1} : f(h_n^*, \mathbb{Q}) \geq \eta\}$$

is compact for every $\eta \in \mathbb{R}$ and $h_n^* \in \mathcal{H}_n^*$. By assumption, $\mathcal{Q}_{n+1}$ is weakly closed and therefore compact by Lemma 4.1. In part a) we have seen that $\mathbb{Q} \mapsto f(h_n^*, \mathbb{Q})$ is continuous for every $h_n^* \in \mathcal{H}_n^*$. Hence, $\{\mathbb{Q} \in \mathcal{Q}_{n+1} : f(h_n^*, \mathbb{Q}) \geq \eta\}$ is closed as the preimage of a closed set. Since closed subsets of compact sets are compact, the proof is complete. □

So far we only considered the case that the ambiguity set may depend on the time index but not on the state of the decision process. This covers many applications, e.g. the connection to risk measures in Section 5.3. Moreover, we can allow any norm bounded ambiguity sets as long as it is independent of the state using the optimal selection theorem by Rieder (1978) in Theorem 4.8. If the ambiguity set is weak* closed, the following generalization is possible.

Corollary 4.9. *For $n = 0, \dots, N-1$ let $\mathcal{Q}_{n+1}$ be weak* closed and*

$$D_n \ni (x,a) \mapsto \mathcal{Q}_{n+1}(x,a) \subseteq \mathcal{Q}_{n+1}$$

be a non-empty- and closed-valued set-valued mapping giving the potential probability measures at time n in state $x \in E$ if the controller chooses $a \in D_n(x)$. Then the assertion of Theorem 4.8 b) still holds.

Proof. We have to show the existence of a measurable maximizer at (4.6). The rest of the Theorem's proof is not affected. Since $\mathcal{Q}_{n+1}$ is weak* closed, it is a compact Borel space by Lemma 4.1. Consequently, the set-valued mapping $\mathcal{Q}_{n+1}(\cdot)$ is compact-valued, as closed subsets of compact sets are compact. In the proof of the theorem it has been shown that the function $f(h_n^*, \mathbb{Q})$ is jointly measurable and continuous in $\mathbb{Q}$. Hence, Proposition A.24 yields the existence of a measurable maximizer. □

In fact, the set-valued mapping $\mathcal{Q}_{n+1}(\cdot)$ may depend on the entire extended history $h_n^* \in \mathcal{H}_n^*$ but then we cannot expect the optimal policy of nature to be Markovian given a Markov policy of the controller, cf. Corollary 4.11 below.

State-dependent ambiguity sets are a possibility to make the distributionally robust optimality criterion less conservative. E.g. they allow to incorporate learning about the unknown disturbance distribution. We refer the reader to Bielecki et al. (2019) for an interesting example where the ambiguity sets are recursive confidence regions for an unknown parameter of the disturbance distribution.

Let us now consider specifically deterministic Markov policies $\pi \in \Pi^M$ of the controller. The subspace

$$\mathbb{B} = \{v \in \mathbb{B}_b : \ v \text{ lower semicontinuous}\}$$

of $(\mathbb{B}_b, \|\cdot\|_b)$ turns out to be the set of potential value functions under such policies. $(\mathbb{B}, \|\cdot\|_b)$ is a complete metric space since the subset of lower semicontinuous functions is closed in $(\mathbb{B}_b, \|\cdot\|_b)$ by Lemma A.10. We define the following operators on $\mathbb{B}_b$ and especially on $\mathbb{B}$.

Definition 4.10. For $v \in \mathbb{B}_b$ and Markov decision rules $d : E \to A$, $\gamma : D_n \to \mathcal{Q}_{n+1}$ let

$$L_n v(x,a,\mathbb{Q}) = \int c_n(x,a,T_n(x,a,Z_{n+1})) + v(T_n(x,a,Z_{n+1})) \,\mathrm{d}\,\mathbb{Q}, \quad (x,a,\mathbb{Q}) \in D_n \times \mathcal{Q}_{n+1},$$

$$\hat{L}_n v(x,a) = \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} L_n v(x,a,\mathbb{Q}), \qquad (x,a) \in D_n,$$

$$\begin{aligned}
\mathcal{T}_{nd\gamma}v(x) &= L_n v(x, d(x), \gamma(x, d(x))), && x \in E, \\
\mathcal{T}_{nd}v(x) &= \hat{L}_n v(x, d(x)), && x \in E, \\
\mathcal{T}_n v(x) &= \inf_{a \in D_n(x)} \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} L_n v(x, a, \mathbb{Q}), && x \in E.
\end{aligned}$$

Note that the operators are monotone in v. Under Markov policies $\pi = (d_0, \dots, d_{N-1}) \in \Pi^M$ of the controller and $\gamma = (\gamma_0, \dots, \gamma_{N-1}) \in \Gamma^M$ of nature, the value iteration can be expressed with the operators. In order to distinguish from the history-dependent case, we denote policy values here with J. Setting $J_{N\pi\gamma}(x) = c_N(x),\ x \in E$, we obtain for $n = 0, \dots, N-1$ and $x \in E$

$$\begin{aligned}
J_{n\pi\gamma}(x) &= \int c_n(x, d_n(x), T_n(x, d_n(x), Z_{n+1}(\omega))) \\
&\qquad + J_{n+1\pi\gamma}(T_n(x, d_n(x), Z_{n+1}(\omega)))\, \gamma_n(\mathrm{d}\,\omega | x, d_n(x)) \\
&= \mathcal{T}_{nd_n\gamma_n} J_{n+1\pi\gamma}(x).
\end{aligned}$$

We define the robust value of Markov policy $\pi \in \Pi^M$ of the controller as

$$J_{n\pi}(x) = \sup_{\gamma \in \Gamma^M} J_{n\pi\gamma}(x), \qquad x \in E.$$

For the robust value of a Markov policy of the controller, a robust value iteration as in Theorem 4.8 holds, too.

Corollary 4.11. *Let $\pi \in \Pi^M$. It holds for $n = 0, \dots, N$ that $J_{n\pi}(x_n) = V_{n\pi}(h_n),\ h_n \in \mathcal{H}_n$. I.e., we have the robust value iteration*

$$\begin{aligned}
J_{n\pi}(x) &= \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} \int c_n(x, d_n(x), T_n(x, d_n(x), Z_{n+1})) + J_{n+1\pi}(T_n(x, d_n(x), Z_{n+1}))\, \mathrm{d}\,\mathbb{Q} \\
&= \mathcal{T}_{nd_n} J_{n+1\pi}(x).
\end{aligned}$$

Moreover, there exists a Markovian ϵ-optimal policy of nature and if the ambiguity sets $\mathcal{Q}_{n+1}$ are all weak closed, even a Markovian optimal policy.*

Proof. For $n = N$ the assertion is trivial. Assuming it holds at time $n+1$, it follows at time n from Theorem 4.8 that

$$\begin{aligned}
V_{n\pi}(h_n) &= \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} \int c_n(x, d_n(x), T_n(x, d_n(x), Z_{n+1})) + J_{n+1\pi}(T_n(x, d_n(x), Z_{n+1}))\, \mathrm{d}\,\mathbb{Q} \\
&= J_{n\pi}(x_n).
\end{aligned}$$

To show the last equality, one replaces $\mathcal{H}_n^*$ by D_n and verifies the existence of an (ϵ-) optimal Markov policy of nature by the same arguments as in the proof of Theorem 4.8. □

Let us further define for $n = 0, \dots, N$ the Markov value function

$$J_n(x) = \inf_{\pi \in \Pi^M} \sup_{\gamma \in \Gamma^M} J_{n\pi\gamma}(x), \qquad x \in E.$$

The next result shows that V_n satisfies a Bellman equation and proves that an optimal policy of the controller exists and is Markov.

Theorem 4.12. *Let Assumption 4.3 be satisfied.*

a) For $n = 0, \dots, N-1$, it suffices to consider deterministic Markov policies both for the controller and nature, i.e. $V_n(h_n) = J_n(x_n)$ for all $h_n \in \mathcal{H}_n$. The value function J_n lies in $\mathbb{B}$ and satisfies the Bellman equation

$$J_N(x) = c_N(x),$$
$$J_n(x) = \mathcal{T}_n J_{n+1}(x), \qquad x \in E.$$

Furthermore, for $n = 0, \dots, N-1$ there exist Markov decision rules d_n^* with $\mathcal{T}_{nd_n^*} J_{n+1} = \mathcal{T}_n J_{n+1}$ and every sequence of such minimizers constitutes an optimal policy $\pi^* = (d_0^*, \dots, d_{N-1}^*) \in \Pi^M$ of the controller.

b) If the one-stage ambiguity set $\mathcal{Q}_{n+1}$ is weak closed, there exists a Markov decision rule γ_n^* of nature with $J_n = \mathcal{T}_{d_n^* \gamma_n^*} J_{n+1}$ and every sequence of such maximizers induces an optimal policy $\gamma^* = (\gamma_0^*, \dots, \gamma_{N-1}^*) \in \Gamma^M$ of nature satisfying $J_n = J_{n\pi^*\gamma^*}$.*

Proof. a) We proceed by backward induction. At time N we have $V_N = J_N = c_N$ which is in $\mathbb{B}$ due to semicontinuity and Assumption 4.3 (ii). Now assume the assertion holds at time $n+1$. Using the robust value iteration (Theorem 4.8), one obtains at time n:

$$\begin{aligned} V_n(h_n) &= \inf_{\pi \in \Pi^R} \sup_{\gamma \in \Gamma} V_{n\pi\gamma}(h_n) \\ &= \inf_{\pi \in \Pi^R} V_{n\pi}(h_n) \\ &= \inf_{\pi \in \Pi^R} \int \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} \int c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ &\qquad\qquad + V_{n+1\pi}(h_n^*, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \mathbb{Q}(\mathrm{d}\,\omega)\, \pi_n(\mathrm{d}\,a_n | h_n). \end{aligned}$$

By the induction hypothesis, V_{n+1} is lower semicontinuous and especially measurable. Hence, we can estimate

$$\begin{aligned} &\geq \inf_{\pi \in \Pi^R} \int \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} \int c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ &\qquad\qquad + V_{n+1}(h_n^*, T_n(x_n, a_n, Z_{n+1}(\omega)))\, \mathbb{Q}(\mathrm{d}\,\omega)\, \pi_n(\mathrm{d}\,a_n | h_n). \end{aligned}$$

This equals by the induction hypothesis

$$= \inf_{\pi \in \Pi^R} \int \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} \int c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ + J_{n+1}(T_n(x_n, a_n, Z_{n+1}(\omega))) \, \mathbb{Q}(\mathrm{d}\,\omega) \, \pi_n(\mathrm{d}\,a_n|h_n).$$

The outer integral can be estimated by the infimum of the integrand

$$\geq \inf_{a_n \in D_n(x_n)} \sup_{\mathbb{Q} \in \mathcal{Q}_{n+1}} \int c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1}(\omega))) \\ + J_{n+1}(T_n(x_n, a_n, Z_{n+1}(\omega))) \, \mathbb{Q}(\mathrm{d}\,\omega)$$

$$= \mathcal{T}_n J_{n+1}(x_n).$$

Here, objective and constraint depend on the history of the process only through x_n. Thus, given existence of a minimizing Markov decision rule d_n^*, one obtains the identity

$$= \mathcal{T}_{n d_n^*} J_{n+1}(x_n). \tag{4.7}$$

Again by the induction hypothesis, there exists an optimal Markov policy $\pi^* = (d_{n+1}^*, \dots, d_{N-1}^*) \in \Pi^M$ such that

$$= \mathcal{T}_{n d_n^*} J_{n+1 \pi^*}(x_n),$$

which equals by Corollary 4.11

$$\begin{aligned} &= J_{n\pi^*}(x_n) \\ &\geq J_n(x_n) \\ &\geq V_n(h_n). \end{aligned}$$

It remains to show the existence of a minimizing Markov decision rule d_n^* at (4.7) and that $J_n \in \mathbb{B}$. We want to apply Proposition A.25. The set-valued mapping $E \ni x \mapsto D_n(x)$ is compact-valued and upper semicontinuous. Next, we show that $D_n \ni (x, a) \mapsto \hat{L}_n v(x, a)$ is lower semicontinuous for every $v \in \mathbb{B}$. Let $\{(x_k, a_k)\}_{k \in \mathbb{N}}$ be a convergent sequence in D_n with limit $(x^*, a^*) \in D_n$. By Lemma A.4 a) the function $D_n \ni (x, a) \mapsto c_n(x, a, T_n(x, a, Z_{n+1}(\omega))) + v(T_n(x, a, Z_{n+1}(\omega)))$ is lower semicontinuous for every $\omega \in \Omega_{n+1}$. Consequently,

$$\begin{aligned} &\liminf_{k \to \infty} c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1}(\omega))) + v(T_n(x_k, a_k, Z_{n+1}(\omega))) \\ &\geq \; c_n(x^*, a^*, T_n(x^*, a^*, Z_{n+1}(\omega))) + v(T_n(x^*, a^*, Z_{n+1}(\omega))), \qquad \omega \in \Omega. \end{aligned} \tag{4.8}$$

The sequence of random variables $\{C_k\}_{n\in\mathbb{N}}$ given by

$$C_k(\omega) = c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1}(\omega))) + v(T_n(x_k, a_k, Z_{n+1}(\omega))), \quad \omega \in \Omega_{n+1}$$

is bounded by some $\bar{C} \in L^p(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$ due to Lemma 4.7. Now, Fatou's Lemma B.1 yields for every $\mathbb{Q} \in \mathcal{Q}_{n+1}$

$$\begin{aligned}
\liminf_{k\to\infty} L_n v(x_k, a_k, \mathbb{Q}) &= \liminf_{k\to\infty} \mathbb{E}^{\mathbb{Q}}\Big[c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})) \\
&\qquad\qquad + v(T_n(x_k, a_k, Z_{n+1}))\Big] \\
&\geq \mathbb{E}^{\mathbb{Q}}\Big[\liminf_{k\to\infty} c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})) \\
&\qquad\qquad + v(T_n(x_k, a_k, Z_{n+1}))\Big] \\
&\geq \mathbb{E}^{\mathbb{Q}}\left[c_n(x^*, a^*, T_n(x^*, a^*, Z_{n+1})) + v(T_n(x^*, a^*, Z_{n+1}))\right] \\
&= L_n v(x^*, a^*, \mathbb{Q}),
\end{aligned}$$

where the last inequality is by (4.8). Thus, the function $D_n \ni (x, a) \mapsto L_n v(x, a, \mathbb{Q})$ is lower semicontinuous for every $\mathbb{Q} \in \mathcal{Q}_{n+1}$ and consequently $D_n \ni (x, a) \mapsto \hat{L}_n v(x, a)$ is lower semicontinuous as a supremum of lower semicontinuous functions, cf. Corollary A.3. Now, Proposition A.25 yields the existence of a minimizing Markov decision rule d_n^* at (4.7) and that $J_n = \mathcal{T}_n J_{n+1}$ is lower semicontinuous. Furthermore, J_n bounded by λb for some $\lambda \in \mathbb{R}_+$ due to Lemma 4.5. Thus $J_n \in \mathbb{B}$.

b) This follows from Theorem 4.8 b). □

4.2. Infinite Planning Horizon

In this section, we consider the distributionally robust cost minimization problem under an infinite planning horizon. This is a reasonable approach if the terminal period is unknown. It can also be seen as an approximation of a model with large but finite planning horizon. Solving the infinite horizon problem will turn out to be easier since it admits a stationary optimal policy.

In the following, we assume the abstract cost model to be stationary and the terminal cost to be zero, i.e. D, T, $\mathcal{Q}$ do not depend on n, the disturbances are identically distributed, the one-stage cost functions are of the form $c_n = \beta^n c$ with some discount factor $\beta \in (0, 1]$ and $c_N \equiv 0$. Let Z be a representative of the disturbance distribution. Due to stationarity, the probability space is given by

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{n=1}^{\infty} (\Omega_1, \mathcal{A}_1, \mathbb{P}_1).$$

The model with infinite planning horizon is derived as a limit of the one with finite horizon. So besides a stationary version of Assumption 4.3 we have to assume that the discount factor β satisfies $\alpha\beta < 1$ to ensure convergence of the value functions when the planning horizon tends to infinity. For clarity, all assumptions of this section are summarized below.

Assumption 4.13. (i) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being continuous (case 1).

(ii) There exist $\alpha, \underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$ and measurable functions $\underline{b} : E \to (-\infty, -\underline{\epsilon}]$, $\bar{b} : E \to [\bar{\epsilon}, \infty)$ such that it holds for all $\mathbb{Q} \in \mathcal{Q}$ and $(x, a) \in D$

$$\mathbb{E}^{\mathbb{Q}}\left[-c^-(x,a,T(x,a,Z))\right] \geq \underline{b}(x), \qquad \mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T(x,a,Z))\right] \geq \alpha\underline{b}(x),$$
$$\mathbb{E}^{\mathbb{Q}}\left[c^+(x,a,T(x,a,Z))\right] \leq \bar{b}(x), \qquad \mathbb{E}^{\mathbb{Q}}\left[\bar{b}(T(x,a,Z))\right] \leq \alpha\bar{b}(x).$$

(iii) We define $b : E \to [1, \infty)$, $b(x) = \bar{b}(x) - \underline{b}(x)$. For all $(\bar{x}, \bar{a}) \in D$ there exists an $\epsilon > 0$ and measurable functions $\Theta_1^{\bar{x},\bar{a}}, \Theta_2^{\bar{x},\bar{a}} : \mathcal{Z} \to \mathbb{R}_+$ such that $\Theta_1^{\bar{x},\bar{a}}(Z), \Theta_2^{\bar{x},\bar{a}}(Z) \in L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ and

$$|c(x,a,T(x,a,z))| \leq \Theta_1^{\bar{x},\bar{a}}(z), \qquad b(T(x,a,z)) \leq \Theta_2^{\bar{x},\bar{a}}(z)$$

for all $z \in \mathcal{Z}$ and $(x, a) \in B_\epsilon(\bar{x}, \bar{a}) \cap D$. Here, $B_\epsilon(\bar{x}, \bar{a})$ is the closed ball around $(\bar{x}, \bar{a})$ w.r.t an arbitrary product metric on $E \times A$.

(iv) The probability measure $\mathbb{P}_1$ on $(\Omega_1, \mathcal{A}_1)$ is separable.

(v) The ambiguity set $\mathcal{Q}$ is norm bounded, i.e. there exists $K \in [1, \infty)$ such that

$$\mathbb{E}\left|\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_1}\right|^q \leq K$$

for all $\mathbb{Q} \in \mathcal{Q}$.

(vi) The discount factor β satisfies $\alpha\beta < 1$.

Since the infinite horizon model is constructed as a limit of one with finite horizon, the consideration can be restricted to deterministic Markov policies $\pi = (d_1, d_2, \dots) \in \Pi^M$ of the controller and $\gamma = (\gamma_1, \gamma_2, \dots) \in \Gamma^M$ of nature due to Corollary 4.11 and Theorem 4.12. 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. With regard to the value iteration, this is also called *forward form.* It is only possible under Markov policies in a stationary model. There, the two ways of indexing are equivalent. In a non-stationary model or under a history-depended policy in a stationary model, the distance-to-horizon indexing is not possible and a change of notation is therefore inevitable. The value of a policy pair $(\pi, \gamma) \in \Pi^M \times \Gamma^M$ up to a planning horizon $N \in \mathbb{N}$ now is

$$J_{N\pi\gamma}(x) = \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{N-1} \beta^k c(X_k, d_k(X_k), X_{k+1})\right], \qquad x \in E. \tag{4.9}$$

Remark 4.14. A stationary model is a special case of a non-stationary one. In a non-stationary formulation of the stationary model, the discounting is included in the cost functions which can vary over time. However, this makes it necessary to calibrate the discounting w.r.t. a fixed reference time, usually the initial stage. If the value functions are considered at a later point in time, the non-stationary and stationary version differ by a discounting factor:

$$J_n^{\text{non-stat}}(x) = \beta^n J_{N-n}^{\text{stat}}(x), \qquad x \in E,\ n = 0, \dots, N.$$

The difference has only interpretational implications, for the optimization it is irrelevant.

The reformulation (4.9) makes it necessary to write the value iteration in terms of the *shifted policies* $\vec{\pi} = (d_1, d_2, \dots)$ corresponding to $\pi = (d_0, d_1, \dots) \in \Pi^M$ and $\vec{\gamma} = (\gamma_1, \gamma_2, \dots)$ corresponding to $\gamma = (\gamma_0, \gamma_1, \dots) \in \Gamma^M$:

$$\begin{aligned} J_{N\pi\gamma}(x) &= \int c\big(x, d_0(x), T(x, d_0(x), Z(\omega))\big) + \beta J_{n+1\vec{\pi}\vec{\gamma}}\big(T(x, d_0(x), Z(\omega))\big)\, \gamma_0(\mathrm{d}\,\omega | x, d_0(x)) \\ &= \mathcal{T}_{d_0\gamma_0} J_{N-1\vec{\pi}\vec{\gamma}}(x), \end{aligned} \tag{4.10}$$

$x \in E$. Due to Theorem 4.8 and Corollary 4.11, the robust value $J_{N\pi} = \sup_{\gamma \in \Gamma^M} J_{N\pi\gamma}$ of a policy $\pi \in \Pi^M$ of the controller satisfies a robust value iteration. It has to be expressed in terms of the shifted policy as well:

$$\begin{aligned} J_{N\pi}(x) &= \sup_{\mathbb{Q} \in \mathcal{Q}} \int c\big(x, d_0(x), T(x, d_0(x), Z(\omega))\big) + \beta J_{n+1\vec{\pi}}\big(T(x, d_0(x), Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega) \\ &= \mathcal{T}_{d_0} J_{N-1\vec{\pi}}(x), \end{aligned} \tag{4.11}$$

$x \in E$. The value function $J_N = \inf_{\pi \in \Pi^M} J_{N\pi}$ under planning horizon $N \in \mathbb{N}$ satisfies due to Theorem 4.12 the Bellman equation

$$J_N(x) = \mathcal{T} J_{N-1}(x) = \mathcal{T}^N 0(x), \qquad x \in E. \tag{4.12}$$

The value of a policy pair $(\pi, \gamma) \in \Pi^M \times \Gamma^M$ under an infinite planning horizon is defined as

$$J_{\infty\pi\gamma}(x) = \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{\infty} \beta^k c(X_k, d_k(X_k), X_{k+1})\right], \qquad x \in E, \tag{4.13}$$

and the corresponding robust value of a policy $\pi \in \Pi^M$ of the controller as

$$J_{\infty\pi}(x) = \sup_{\gamma \in \Gamma^M} J_{\infty\pi\gamma}(x), \qquad x \in E.$$

Hence, this section's optimality criterion is

$$J_\infty(x) = \inf_{\pi \in \Pi^M} J_{\infty\pi}(x), \qquad x \in E. \tag{4.14}$$

Lemma 4.15. *Under Assumption 4.13, the sequences* $\{J_{N\pi\gamma}\}_{N\in\mathbb{N}}$*,* $\{J_{N\pi}\}_{N\in\mathbb{N}}$*,* $\{J_N\}_{N\in\mathbb{N}}$

are weakly increasing. Hence, they converge pointwise for every policy pair $(\pi,\gamma)\in\Pi^M\times\Gamma^M$ *to limits which are bounded by* $\frac{1}{1-\alpha\beta}\underline{b}$ *below and* $\frac{1}{1-\alpha\beta}\bar{b}$ *above. Moreover, it holds*

$$\lim_{N\to\infty} J_{N\pi\gamma} = J_{\infty\pi\gamma}(x), \qquad x\in E.$$

Proof. We have for $1\le m\le N$

$$\begin{aligned}
J_{N\pi\gamma}(x) &= \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{N-1}\beta^k c(X_k,d_k(X_k),X_{k+1})\right]\\
&= \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{m-1}\beta^k c(X_k,d_k(X_k),X_{k+1})\right] + \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=m+1}^{N-1}\beta^k c(X_k,d_k(X_k),X_{k+1})\right]\\
&= J_{m\pi\gamma}(x) + \sum_{k=m+1}^{N-1}\beta^k\mathbb{E}_{0x}^{\pi\gamma}\left[c(X_k,d_k(X_k),X_{k+1})\right]\\
&\ge J_{m\pi\gamma}(x) + \sum_{k=m+1}^{N-1}\beta^k\mathbb{E}_{0x}^{\pi\gamma}\left[-c^-(X_k,d_k(X_k),X_{k+1})\right]\\
&\ge J_{m\pi\gamma}(x) + \underline{b}(x)\sum_{k=m+1}^{N-1}(\alpha\beta)^k\\
&\ge J_{m\pi\gamma}(x) + \delta_m(x)
\end{aligned}\tag{4.15}$$

where the second inequality follows as in the proof of Lemma 4.5 and

$$\delta_m:\mathbb{R}\to(-\infty,0],\quad \delta_m(x)=\underline{b}(x)\sum_{k=m}^{\infty}(\alpha\beta)^k$$

is a non-positive function with $\lim_{m\to\infty}\delta_m(x)=0$ for all $x\in E$. Hence, the sequence of functions $\{J_{N\pi\gamma}\}_{N\in\mathbb{N}}$ is weakly increasing. Taking the supremum over γ (and the infimum over π) on both sides of (4.15), yields that the sequences $\{J_{N\pi}\}_{N\in\mathbb{N}}$ and $\{J_N\}_{N\in\mathbb{N}}$ are weakly increasing, too. By Lemma A.9 a) all three sequences are convergent.

To due Lemma 4.5 (ii), we can apply Theorem B.3 which yields

$$\begin{aligned}
J_{\infty\pi\gamma}(x) &= \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{\infty}\beta^k c(X_k,d_k(X_k),X_{k+1})\right]\\
&= \lim_{N\to\infty} E_{0x}^{\pi\gamma}\left[\sum_{k=0}^{N-1}\beta^k c(X_k,d_k(X_k),X_{k+1})\right]\\
&= \lim_{N\to\infty} J_{N\pi\gamma}(x).
\end{aligned}$$

Observing the discounting and zero terminal cost, it follows from Lemma 4.5 that

$$\sum_{k=0}^{N-1}(\alpha\beta)^k\underline{b}(x)\le J_{N\pi\gamma}(x)\le\sum_{k=0}^{N-1}(\alpha\beta)^k\bar{b}(x).$$

Taking the limit $N \to \infty$ yields

$$\frac{1}{1-\alpha\beta}\underline{b}(x) \leq J_{\infty\pi\gamma}(x) \leq \frac{1}{1-\alpha\beta}\bar{b}(x).$$

For the other limits the bounds hold, too. □

The fact that $\{J_{N\pi\gamma}\}_{N\in\mathbb{N}}$ is weakly increasing is exploited to show the convergence of $\{J_{N\pi}\}_{N\in\mathbb{N}}$ and $\{J_N\}_{N\in\mathbb{N}}$. The convergence of $\{J_{N\pi\gamma}\}_{N\in\mathbb{N}}$ itself can directly be inferred from Theorem B.3. The pointwise limits

$$J_\pi(x) = \lim_{N\to\infty} J_{N\pi}(x) \qquad \text{and} \qquad J(x) = \lim_{N\to\infty} J_N(x), \qquad x \in E,$$

are referred to as *limit robust policy value* of $\pi \in \Pi^M$ and *limit value function*, respectively.

Remark 4.16. The robust policy values and value functions have the following relations.

a) It holds for any policy pair $(\pi,\gamma) \in \Pi^M \times \Gamma^M$ that $J_{N\pi\gamma} \leq J_{N\pi}$. By taking the limit $N \to \infty$ it follows $J_{\infty\pi\gamma} \leq J_\pi$ and finally by taking the supremum over $\gamma \in \Gamma^M$

$$J_{\infty\pi}(x) \leq J_\pi(x), \qquad x \in E.$$

b) It holds for any policy $\pi \in \Pi$ that $J_N \leq J_{N\pi}$. Taking limits yields

$$J(x) \leq J_\pi(x), \qquad x \in E.$$

Lemma 4.17. *Given Assumption 4.13, the Bellman operator $\mathcal{T}$ is a contraction on $\mathbb{B}$ with modulus $\alpha\beta \in (0,1)$.*

Proof. Let $v \in \mathbb{B}$. It has been established in the proof of Theorem 4.12 that $\mathcal{T}v$ is lower semicontinuous. Furthermore,

$$\begin{aligned}
|\mathcal{T}v(x)| &= \left| \inf_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c(x,a,T(x,a,Z)) + \beta v(T(x,a,Z))\right]\right| \\
&\leq \inf_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[|c(x,a,T(x,a,Z))|\right] + \beta\mathbb{E}^{\mathbb{Q}}\left[|v(T(x,a,Z))|\right] \\
&\leq \inf_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[|c(x,a,T(x,a,Z)))|\right] + \beta\mathbb{E}^{\mathbb{Q}}\left[b(T(x,a,Z))\right] \\
&\leq (1+\alpha\beta)b(x),
\end{aligned}$$

where the last inequality is by Remark 4.4. Hence, the operator $\mathcal{T}$ is an endofunction on $\mathbb{B}$ and it remains to verify the Lipschitz constant $\alpha\beta$. It holds for $v_1, v_2 \in \mathbb{B}$

$$\begin{aligned}
\mathcal{T}v_1(x) - \mathcal{T}v_2(x) \leq \sup_{a\in D(x)} \Big(&\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c(x,a,T(x,a,Z)) + \beta v_1(T(x,a,Z))\right] \\
&- \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c(x,a,T(x,a,Z))) + \beta v_2(T(x,a,Z))\right] \Big)
\end{aligned}$$

$$\begin{aligned}
&\le \beta \sup_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[v_1(T(x,a,Z)) - v_2(T(x,a,Z))\right] \\
&\le \beta \|v_1 - v_2\|_b \sup_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[b(T(x,a,Z))\right] \\
&\le \alpha\beta\|v_1 - v_2\|_b b(x).
\end{aligned}$$

The first two inequalities follow from Lemma A.31 and the last one from Remark 4.4. Interchanging the roles of v_1 and v_2 yields

$$|\mathcal{T}v_1(x) - \mathcal{T}v_2(x)| \le \alpha\beta\|v_1 - v_2\|_b b(x).$$

Now, dividing by $b(x)$ and taking the supremum over $x \in E$ on the left hand side completes the proof. □

Under a finite planning horizon $N \in \mathbb{N}$ we have characterized the value function with the Bellman equation (4.12). Theorem 4.18 below shows that this is compatible with the infinite horizon optimality criterion (4.14).

Theorem 4.18. *Let Assumption 4.13 be satisfied.*

a) The limit value function J is the unique fixed point of the Bellman operator $\mathcal{T}$ in $\mathbb{B}$.

b) There exists a Markov decision rule $d^* : E \to A$ of the controller such that

$$\mathcal{T}_{d^*} J(x) = \mathcal{T} J(x), \qquad x \in E.$$

Moreover, for every $\epsilon > 0$ there exists an ϵ-optimal Markov decision rule $\hat{\gamma}_0 : D \to \mathcal{Q}$ of nature such that

$$\mathcal{T}_{d^*\hat{\gamma}_0} J(x) + \epsilon \ge \mathcal{T} J(x), \qquad x \in E.$$

c) If the ambiguity set $\mathcal{Q}$ is weak closed, there exists an optimal Markov decision rule $\gamma_0^* : D \to \mathcal{Q}$ of nature such that*

$$\mathcal{T}_{d^*\gamma_0^*} J(x) = \mathcal{T} J(x), \qquad x \in E.$$

d) Each stationary policy $\pi^* = (d^*, d^*, \dots)$ induced by a Markov decision rule d^* as in part b) is optimal for optimization problem (4.14) and it holds $J_\infty = J$.

e) If the ambiguity set $\mathcal{Q}$ is weak closed, each stationary policy $\gamma^* = (\gamma_0^*, \gamma_0^*, \dots)$ induced by a decision rule γ_0^* as in part c) is an optimal response of nature to π^*, i.e. $J_{\infty\pi^*\gamma^*} = J_\infty$.*

Proof. a) The fact that J is the unique fixed point of the operator $\mathcal{T}$ in $\mathbb{B}$ follows directly from Banach's Fixed Point Theorem using Lemma 4.17.

b) The existence of a minimizing Markov decision rule of the controller and an ϵ-optimal Markov decision rule of nature follow from the respective results in the finite horizon case, cf. Theorem 4.12 a) and Corollary 4.11.

c) This follows analogously from Theorem 4.12 b).

d) Let d^*, $\hat{\gamma}_0$ be Markov decision rules as in part b) and $\pi^* = (d^*, d^*, \dots)$, $\hat{\gamma} = (\hat{\gamma}_0, \hat{\gamma}_0, \dots)$. It has to be shown that

$$J_{\infty\pi^*}(x) = J_\infty(x) = J(x), \qquad x \in E. \tag{4.16}$$

We proceed in two steps. Firstly, we prove that

$$J(x) \geq J_{\pi^*}(x), \qquad x \in E \tag{4.17}$$

and secondly we prove that

$$J(x) \leq J_{\infty\pi}(x), \qquad x \in E, \qquad \text{for all } \pi \in \Pi^M. \tag{4.18}$$

Combining (4.17) and Remark 4.16 a), we get $J \geq J_{\infty\pi^*} \geq J_\infty$. On the other hand, taking the infimum over $\pi \in \Pi^M$ in (4.18) yields $J \leq J_\infty$. Together, these inequalities imply (4.16) and the assertion is proven.

Step 1: We show by induction that for all $N \in \mathbb{N}_0$

$$J(x) \geq J_{N\pi^*}(x) + \frac{(\alpha\beta)^N}{1-\alpha\beta}\underline{b}(x), \qquad x \in E.$$

Then letting $N \to \infty$ yields (4.17). Regarding the base case $N = 0$ consider Lemma 4.5 (i). Taking into account the discounting and zero terminal cost, we have

$$J_N(x) \geq \sum_{k=0}^{N-1} (\alpha\beta)^k \underline{b}(x)$$

Letting $N \to \infty$ yields $J(x) \geq \frac{1}{1-\alpha\beta}\underline{b}(x)$, i.e. the claim holds for $N = 0$. For $N \geq 1$ it follows from the induction hypothesis

$$\begin{aligned}
J(x) &= \mathcal{T}_{d^*} J(x) \\
&= \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c\big(x, d^*(x), T(x, d^*(x), Z)\big) + \beta J\big(T(x, d^*(x), Z)\big)\right] \\
&\geq \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\Bigg[c\big(x, d^*(x), T(x, d^*(x), Z)\big) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big) \\
&\qquad\qquad + \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\underline{b}\big(T(x, d^*(x), Z)\big)\Bigg] \\
&= \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\Bigg[c\big(x, d^*(x), T(x, d^*(x), Z)\big) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big)\Bigg] \\
&\qquad + \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\mathbb{E}^{\mathbb{Q}}\Bigg[\underline{b}\big(T(x, d^*(x), Z)\big)\Bigg]
\end{aligned}$$

$$\begin{aligned}
&\geq \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\Big[c(x, d^*(x), T(x, d^*(x), Z)) + \beta J_{N-1\pi^*}(T(x, d^*(x), Z))\Big] \\
&\qquad + \frac{(\alpha\beta)^N}{1-\alpha\beta}\underline{b}(x) \\
&= J_{N\pi^*}(x) + \frac{(\alpha\beta)^N}{1-\alpha\beta}\underline{b}(x).
\end{aligned}$$

Note that the last inequality is by Assumption 4.13 (ii) and the last equality by the robust value iteration (4.11).

Step 2: Let $\pi = (d_0, d_1, \dots) \in \Pi^M$ be arbitrary. We show by induction that for all $N \in \mathbb{N}_0$

$$J(x) \leq J_{N\pi\hat{\gamma}}(x) + \frac{\epsilon}{1-\beta} + \frac{(\alpha\beta)^N}{1-\alpha\beta}\bar{b}(x), \qquad x \in E.$$

Then letting $N \to \infty$ yields $J \leq J_{\infty\pi\hat{\gamma}} + \frac{\epsilon}{1-\beta}$. Since $\epsilon > 0$ is arbitrarily small, it follows $J \leq J_{\infty\pi}$, i.e. (4.18) holds. The base case $N = 0$ follows analogously from Lemma 4.5 (ii). For $N \geq 1$ we have

$$\begin{aligned}
J(x) &= \mathcal{T} J(x) \\
&\leq \mathcal{T}_{d_0} J(x) \\
&\leq \mathcal{T}_{d_0\hat{\gamma}_0} J(x) + \epsilon \\
&\leq \mathcal{T}_{d_0\hat{\gamma}_0}\left(J_{N-1\vec{\pi}\hat{\gamma}}(x) + \frac{\epsilon}{1-\beta} + \frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\bar{b}(x)\right) + \epsilon \\
&= \int c(x, d_0(x), T(x, d_0(x), Z(\omega))) + \beta J_{N-1\vec{\pi}\hat{\gamma}}(T(x, d_0(x), Z(\omega))) \\
&\qquad + \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\bar{b}(T(x, d_0(x), Z(\omega)))\, \hat{\gamma}_0(\mathrm{d}\,\omega|x, d_0(x)) + \left(1 + \frac{\beta}{1-\beta}\right)\epsilon \\
&= J_{N\pi\hat{\gamma}}(x) + \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\int \bar{b}(T(x, d_0(x), Z(\omega)))\, \hat{\gamma}_0(\mathrm{d}\,\omega|x, d_0(x)) + \frac{\epsilon}{1-\beta} \\
&\leq J_{N\pi\hat{\gamma}}(x) + \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[\bar{b}(T(x, d_0(x), Z))\right] + \frac{\epsilon}{1-\beta} \\
&\leq J_{N\pi\hat{\gamma}}(x) + \frac{(\alpha\beta)^N}{1-\alpha\beta}\bar{b}(x) + \frac{\epsilon}{1-\beta}.
\end{aligned}$$

The second inequality is by Theorem 4.8 a) and the third one by the induction hypothesis. There, we also used that $\pi \in \Pi^M$ is arbitrary, so it is no problem to switch to the shifted policy $\vec{\pi}$. The third equality is by the value iteration (4.10) and the last inequality by Assumption 4.13 (ii).

e) Replacing the ϵ-optimal decision rule $\hat{\gamma}_0$ by the optimal one γ_0^* in step 2 of part d) yields $J \leq J_{\infty\pi\gamma^*}$ for all $\pi \in \Pi^M$, so especially $J \leq J_{\infty\pi^*\gamma^*}$. Combining this with

(4.16), we get

$$J \le J_{\infty\pi^*\gamma^*} \le J_{\infty\pi^*} = J_\infty = J,$$

which concludes the proof. □

Iyengar (2005, 3.3) observed for his model with countable state and action spaces that if the controller chooses a stationary policy under an infinite planning horizon, it is optimal for nature to react with a stationary policy. Part e) of Theorem 4.18 shows that this holds for more general state and action spaces, too. An ambiguity model, where nature has to apply the same disturbance distribution each time a state action combination is revisited, was termed *static*.

4.3. Real Line as State Space

The abstract cost model has been introduced in Section 3.1 with a general Borel space as state space. In order to solve the distributionally robust cost minimization problem in Sections 4.1 and 4.2 we needed a continuous transition function despite having a semicontinuous model, cf. the proof of Theorem 4.12 together with Lemma A.17 a). 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. In some applications, see e.g. Section 4.4.1, this relaxation of the continuity assumption is relevant. Furthermore, a real state space can be exploited to address the distributionally robust cost minimization problem with more specific techniques.

To ease the notational burden, we consider the stationary model with no terminal cost under both finite and infinite horizon in this section. All results can be transferred to a non-stationary setting by mere notational changes if the planning horizon is finite. We make the following assumptions in this section.

Assumption 4.19. (i) The state space is the real line $E = \mathbb{R}$.

(ii) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being lower semicontinuous (case 2).

(iii) The model data has the following monotonicity properties:

(iii a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is decreasing.

(iii b) The transition function T is increasing in x.

(iii c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is increasing for all (a, z).

(iv) Assumptions 4.13 (ii) to (vi) hold.

Requiring that the one-stage cost function c is increasing both in x and x' is sufficient for Assumption 4.19 (iii c) to hold since the transition function is increasing in x. Besides, if c is increasing in x', it is sufficient for Continuity and Compactness Properties 3.1 (iii) that c is lower semicontinuous due to Lemma A.4 b). With the real line as state space, a simple separation condition is sufficient for Assumption 4.13 (iii).

Corollary 4.20. *Let there be upper semicontinuous functions* $\vartheta_1, \vartheta_2 : D \to \mathbb{R}_+$ *and measurable functions* $\Theta_1, \Theta_2 : \mathcal{Z} \to \mathbb{R}_+$ *which fulfill* $\Theta_1(Z), \Theta_2(Z) \in L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ *and*

$$|c(x, a, T(x, a, z))| \le \vartheta_1(x, a) + \Theta_1(z), \qquad b(T(x, a, z)) \le \vartheta_2(x, a) + \Theta_2(z)$$

for every $(x, a, z) \in D \times \mathcal{Z}$*. Then Assumption 4.13 (iii) is satisfied.*

Proof. Let $(\bar{x}, \bar{a}) \in D$. We can choose $\epsilon > 0$ arbitrarily. The set $S = [\bar{x} - \epsilon, \bar{x} + \epsilon] \times D(\bar{x} - \epsilon)$ is compact w.r.t. the product topology by the Tychonoff Product Theorem (Aliprantis and Border; 2006, 2.61). Moreover, $B_\epsilon(\bar{x}, \bar{a}) \cap D \subseteq S$ since the set-valued mapping $D(\cdot)$ is decreasing. Due to upper semicontinuity there exist $(x_i, a_i) \in S$ such that $\vartheta_i(x_i, a_i) = \sup_{(x,a)\in S} \vartheta_i(x, a)$, $i = 1, 2$. Hence, one can define

$$\Theta_i^{\bar{x},\bar{a}}(\cdot) = \vartheta_i(x_i, a_i) + \Theta_i(\cdot), \qquad i = 1, 2$$

and Assumption 4.13 (iii) is satisfied. □

The question is how replacing Assumption 4.13 (i) by Assumption 4.19 (i) to (iii) affects the validity of all previous results. 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 4.12 and 4.18. All other statements are unaffected.

Proposition 4.21. *The assertions of Theorems 4.12 and 4.18 hold under Assumption 4.19, too. Moreover, the value functions* J_n *and* J *are increasing. The set of potential value functions can therefore be replaced by*

$$\mathbb{B} = \{v \in \mathbb{B}_b : \ v \textit{ lower semicontinuous and increasing}\}.$$

Proof. The subset of increasing functions in $\{v \in \mathbb{B}_b : \ v \text{ lower semicontinuous}\}$ is closed w.r.t. pointwise convergence, so especially w.r.t. $\|\cdot\|_b$. Hence, $(\mathbb{B}, \|\cdot\|_b)$ is a complete metric space as a closed subset of complete metric space.

The proof of Theorem 4.18 uses the continuity of T only indirectly through Theorem 4.12. Thus, we only have to validate the assertion of the latter. There, the continuity of T is used to show that $D \ni (x, a) \mapsto Lv(x, a)$ is lower semicontinuous for every $v \in \mathbb{B}$. Due to the monotonicity assumptions, the integrand

$$D \ni (x, a) \mapsto c\big(x, a, T(x, a, Z(\omega))\big) + \beta v\big(T(x, a, Z(\omega))\big)$$

is lower semicontinuous for every $\omega \in \Omega_1$ by part b) of Lemma A.4 (instead of part a) which is used in the proof of Theorem 4.12). Now, the lower semicontinuity of $D \ni (x, a) \mapsto Lv(x, a)$ and the existence of a minimizing decision rule follow as in the proof of Theorem 4.12. The fact that $\mathcal{T}v$ is increasing for every $v \in \mathbb{B}$ follows from Lemma A.19. □

The monotonicity requirements in Assumption 4.19 (iii) are only one option. The following alternative is relevant i.a. for the dynamic reinsurance models introduced in Section 3.2.

Corollary 4.22. *Assumptions 4.19 (ii) and (iii) can be replaced by*

(ii') The model data has the Continuity and Compactness Properties 3.1 with the transition function T being upper semicontinuous (case 3).

(iii') The model data has the following monotonicity properties:

(iii' a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is increasing.

(iii' b) The transition function T is increasing in x.

(iii' c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is decreasing for all (a, z).

Then, the assertions of Theorems 4.12 and 4.18 still hold. Moreover, the value functions J_n and J are decreasing and the set of potential value functions is

$$\mathbb{B} = \{v \in \mathbb{B}_b : \ v \text{ lower semicontinuous and decreasing}\}.$$

Proof. One argues analogously to the proof of Proposition 4.21. In order to show that $D \ni (x, a) \mapsto Lv(x, a)$ is lower semicontinuous for every $v \in \mathbb{B}$, one uses Remark A.5 to verify that the integrand

$$D \ni (x, a) \mapsto c\big(x, a, T(x, a, Z(\omega))\big) + \beta v\big(T(x, a, Z(\omega))\big)$$

is lower semicontinuous for every $\omega \in \Omega_1$. □

Requiring that the one-stage cost function c is decreasing both in x and x' is sufficient for (iii' c) to hold since the transition function is increasing in x. Besides, if c is decreasing in x', it is sufficient for Continuity and Compactness Assumption 3.1 (iii) that c is lower semicontinuous due to Remark A.5.

In the following Section 4.3.1, we use a minimax approach as an alternative way to solve the Bellman equation of the distributionally robust cost minimization problem and to study its game-theoretical properties. Subsequently in Section 4.3.2, we consider special choices of the ambiguity set which are advantageous for solving the optimization problem.

4.3.1. Minimax Approach and Game Theory

Compared to a risk-neutral Markov Decision Model, the Bellman equation of the robust model

$$\begin{aligned} J_N(x) &= 0, \\ J_n(x) &= \inf_{a \in D(x)} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}} \Big[c\big(x, a, T(x, a, Z)\big) + \beta J_{n+1}\big(T(x, a, Z)\big) \Big], \qquad x \in \mathbb{R}, \end{aligned} \tag{4.19}$$

has the additional complication that a supremum over possibly uncountably many expectations needs to be calculated. This can be a quite challenging task. Therefore, it may be advantageous to interchange the infimum and supremum in (4.19). For instance, in concrete applications it may be possible to infer structural properties of the optimal actions independently from the probability measure $\mathbb{Q}$ after the interchange. Based on the minimax theorem by Sion (1958), cf. Appendix A.4, this section presents a criterion under which the interchange of infimum and supremum is possible.

Lemma 4.23. *Let A be a subset of a vector space, the admissible state-action-combinations D be a convex set, the transition function T be convex in (x,a) and the composition $D \ni (x,a) \mapsto c(x,a,T(x,a,z))$ be a convex function for every $z \in \mathcal{Z}$. Then the value functions J_n and the limit value function J are convex.*

Proof. The proof is by backward induction. J_N is convex as a constant function. Now assume that J_{n+1} is convex. Recall that J_{n+1} is increasing (Proposition 4.21). Hence, for every $\omega \in \Omega$ the function

$$D \ni (x,a) \mapsto c(x,a,T(x,a,Z(\omega))) + \beta J_{n+1}(T(x,a,Z(\omega)))$$

is convex as the second summand is a composition of an increasing convex with a convex function. By the linearity of expectation,

$$D \ni (x,a) \mapsto \mathbb{E}^{\mathbb{Q}}\Big[c(x,a,T(x,a,Z)) + \beta J_{n+1}(T(x,a,Z))\Big] \tag{4.20}$$

is convex for every $\mathbb{Q} \in \mathcal{Q}$. As the pointwise supremum of a collection of convex functions is convex, we obtain convexity of $D \ni (x,a) \mapsto \hat{L}J_{n+1}(x,a)$. Now, Proposition 2.4.18 in Bäuerle and Rieder (2011) yields the assertion. □

The assumptions of Lemma 4.23 are subsequently referred to as *convex model*.

Theorem 4.24. *In a convex model we have for all $n = 0,\dots,N-1$*

$$J_n(x) = \inf_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} LJ_{n+1}(x,a,\mathbb{Q}) = \sup_{\mathbb{Q}\in\mathcal{Q}} \inf_{a\in D(x)} LJ_{n+1}(x,a,\mathbb{Q}), \qquad x \in \mathbb{R}.$$

Proof. Let $x \in \mathbb{R}$ be fixed and define $f : D(x) \times \mathcal{Q} \to \mathbb{R}$,

$$f(a,\mathbb{Q}) = Lv(x,a,\mathbb{Q}) = \mathbb{E}^{\mathbb{Q}}\Big[c(x,a,T(x,a,Z)) + \beta J_{n+1}(T(x,a,Z))\Big].$$

The function f is convex in a by (4.20) and linear in $\mathbb{Q}$, i.e. especially concave. Furthermore, the set $D(x)$ is compact and it has been shown in the proof of Theorem 4.12 that f is lower semicontinuous in a. Hence, the assertion follows from Theorem A.27 a). □

Remark 4.25. The interchange of infimum and supremum in Theorem 4.24 is based on Sion's Minimax Theorem A.27, which requires convexity of the function

$$a \mapsto \int c(x,a,T(x,a,Z(\omega))) + \beta J_{n+1}(T(x,a,Z(\omega)))\, \mathbb{Q}(\mathrm{d}\,\omega) \tag{4.21}$$

for every $(x, \mathbb{Q}) \in \mathbb{R} \times \mathcal{Q}$. This can be guaranteed by a convex model (cf. Lemma 4.23) which means that several components of the decision model need to have some convexity property. However, these assumptions are quite restrictive. Resorting to randomized actions is a standard approach to convexify (or more precisely linearize) the function (4.21) without assumptions on the model components. Let $\mathcal{P}(D(x))$ be the set of all probability measures on $D(x)$. Then it follows from Sion's Theorem A.27 that

$$\inf_{\mu\in\mathcal{P}(D(x))} \sup_{\mathbb{Q}\in\mathcal{Q}} \iint c\big(x,a,T(x,a,Z(\omega))\big) + \beta J_{n+1}\big(T(x,a,Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega)\, \mu(\mathrm{d}\,a) \tag{4.22}$$

$$= \sup_{\mathbb{Q}\in\mathcal{Q}} \inf_{\mu\in\mathcal{P}(D(x))} \iint c\big(x,a,T(x,a,Z(\omega))\big) + \beta J_{n+1}\big(T(x,a,Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega)\, \mu(\mathrm{d}\,a)$$

$$= \sup_{\mathbb{Q}\in\mathcal{Q}} \inf_{a\in D(x)} \int c\big(x,a,T(x,a,Z(\omega))\big) + \beta J_{n+1}\big(T(x,a,Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega). \tag{4.23}$$

The last equality holds since $a \mapsto c\big(x,a,T(x,a,Z)\big) + \beta J_{n+1}\big(T(x,a,Z)\big)$ is lower semicontinuous (cf. the proof of Theorem 4.12) and $D(x)$ is compact. This appears to be a very elegant solution for the interchange problem but unfortunately the Bellman equation of the distributionally robust cost minimization problem (4.4) under a randomized action of the controller is given by

$$J_n(x) = \inf_{\mu\in\mathcal{P}(D(x))} \int \sup_{\mathbb{Q}\in\mathcal{Q}} \int c\big(x,a,T(x,a,Z(\omega))\big) + \beta J_{n+1}\big(T(x,a,Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega)\, \mu(\mathrm{d}\,a) \tag{4.24}$$

$$= \inf_{a\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} \int c\big(x,a,T(x,a,Z(\omega))\big) + \beta J_{n+1}\big(T(x,a,Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega),$$

cf. Theorems 4.8 and 4.12. Equation (4.22) does in general not equal (4.24). Recall that in our model nature is allowed to react to any realization of the controllers action. This was crucial to obtain a robust value iteration in Theorem 4.8. In contrast to that, (4.22) means that nature maximizes only knowing the distribution of the controller's action. In order to formally see that (4.22) $\neq$ (4.24) consider the simple static counter example $N = 1$, $E = \mathbb{R}$, $A = [0,1]$, $D = \mathbb{R} \times A$, $Z \sim \mathrm{Bin}(1,p)$, $p \in [0,1] = \mathcal{Q}$, $T(x,a,z) = -(a-z)^2$ and $c(x,a,x') = x'$. It is readily checked that Assumption 4.19 is satisfied. Especially, one has constant bounding functions. In this example (4.24) equals

$$\begin{aligned}\inf_{a\in[0,1]} \sup_{p\in[0,1]} \mathbb{E}^p\big[c(x,a,T(x,a,Z))\big] &= \inf_{a\in[0,1]} \sup_{p\in[0,1]} -(1-p)a^2 - p(a-1)^2 \\ &= -\sup_{a\in[0,1]} \inf_{p\in[0,1]} (1-p)a^2 + p(a-1)^2 \\ &= -\sup_{a\in[0,1]} \min\{a^2,(1-a)^2\} = -\frac{1}{4}.\end{aligned}$$

If controller chooses $\mu \sim \mathcal{U}(0,1)$, then (4.22) must be lower or equal than

$$\sup_{p\in[0,1]} \int_0^1 -(1-p)a^2 - p(a-1)^2 \,\mathrm{d}\, a = \sup_{p\in[0,1]} -\frac{1}{3}(1-p) - \frac{1}{3}p = -\frac{1}{3}.$$

In fact, by solving (4.23), one sees that (4.22) equals

$$\sup_{p\in[0,1]} \inf_{a\in[0,1]} -(1-p)a^2 - p(a-1)^2 = -\inf_{p\in[0,1]} \max\{1-p, p\} = -\frac{1}{2}.$$

The approach to interchange infimum and supremum through a linearization with randomized actions is used by Bäuerle and Rieder (2019) for a problem similar to (4.22). It works when nature only observes the distribution and not the realization of the controller's action. In this case, it matters whether the controller can use randomized decisions. Only after the interchange of infimum and supremum, i.e. when he can react to any state of nature, he can resort to deterministic decisions without increasing his cost.

As mentioned before, the distributionally robust cost minimization model can be interpreted as a dynamic game with nature as the controllers opponent. Since nature chooses her action after the controller, observing his action but not being restricted by it, there is a (weak) *second-mover advantage* by construction of the game. The fact that infimum and supremum in the Bellman equation can be interchanged means that the second-mover advantage vanishes in the special case of a convex model.

Let additionally the ambiguity set $\mathcal{Q}$ be weak* closed. This is e.g. the case if $\mathcal{Q}$ is induced by the dual representation of a proper coherent risk measure with the Fatou property, cf. Proposition 2.21. Now, the conditions of Theorem A.27 b) are fulfilled, too, since the ambiguity set is weak* compact by Lemma 4.1 and by Lemma 4.7 we have that $c(x,a,T(x,a,Z)) + \beta J_{n+1}(T(x,a,Z))$ is in L^p. Thus, $\mathbb{Q} \mapsto LJ_{n+1}(x,a,\mathbb{Q})$ is weak* continuous for every $(x,a) \in D$. Remark A.28 yields that $(a,\mathbb{Q}) \mapsto LJ_{n+1}(x,a,\mathbb{Q})$ satisfies the minimax equality

$$\min_{a\in D(x)} \max_{\mathbb{Q}\in\mathcal{Q}} LJ_{n+1}(x,a,\mathbb{Q}) = \max_{\mathbb{Q}\in\mathcal{Q}} \min_{a\in D(x)} LJ_{n+1}(x,a,\mathbb{Q})$$

and Lemma implies A.30 that for every $x \in \mathbb{R}$ the function has a saddle point $(a^*, \mathbb{Q}^*)$, i.e.

$$LJ_{n+1}(x,a^*,\mathbb{Q}) \le LJ_{n+1}(x,a^*,\mathbb{Q}^*) \le LJ_{n+1}(x,a,\mathbb{Q}^*)$$

for all $a \in D(x)$ and $\mathbb{Q} \in \mathcal{Q}$. Such a saddle point constitutes a *Nash equilibrium* in the subgame scenario $X_n = x$. We will show that Nash equilibria exist not only in one-stage subgames but also globally.

Before, let us introduce a modification of the game against nature where nature instead of the controller moves first, i.e. $\sup_\gamma \inf_\pi \; V_{n\pi\gamma}$. Given a policy of nature, the controller faces an arbitrary but fixed probability measure in each scenario $X_n = x$. Thus, the inner optimization problem is a risk-neutral MDP and it follows from standard theory (cf.

e.g. Hernández-Lerma and Lasserre; 1996) that is suffices for the controller to consider deterministic Markov policies. Therefore, we can directly use a forward (or distance to horizon) indexation. Clearly, the controller's optimal policy will depend on the policy that nature has chosen before. It will turn out to be a pointwise dependence on the actions of nature. To clarify this and for comparability with the original game (4.4), where the controller moves first, we distinguish the following types of Markov strategies of the controller

$$\begin{aligned}\Pi(\mathbb{R}) = \Pi^M &= \{\pi = (d_0, d_1, \dots) |\, d_n : \mathbb{R} \to A \text{ measurable, } d_n(x) \in D(x),\, x \in \mathbb{R}\}\\ \Pi(\mathbb{R}, \mathcal{Q}) &= \{\pi = (d_0, d_1, \dots) |\, d_n : \mathbb{R} \times \mathcal{Q} \to A \text{ measurable, } d_n(x, \mathbb{Q}) \in D(x),\, x \in \mathbb{R}\}\end{aligned}$$

and of nature

$$\begin{aligned}\Gamma(\mathbb{R}) &= \{\gamma = (\gamma_0, \gamma_1, \dots) |\, \gamma_n : \mathbb{R} \to \mathcal{Q} \text{ measurable}\}\\ \Gamma(\mathbb{R}, A) = \Gamma^M &= \{\gamma = (\gamma_0, \gamma_1, \dots) |\, \gamma_n : \mathbb{R} \times A \to \mathcal{Q} \text{ measurable}\}\,.\end{aligned}$$

The sets of corresponding stationary strategies will be denoted by a superscript S. The value $J_{N\pi\gamma}$ of a pair of Markov policies $(\gamma, \pi) \in \Gamma(\mathbb{R}) \times \Pi(\mathbb{R}, \mathcal{Q})$ is defined as in (4.9) with forward indexation. The bounds in Lemma 4.5 and the value iteration (4.10) apply since the proofs do not use properties of the policies. The game under consideration is

$$\widetilde{J}_N(x) = \sup_{\gamma \in \Gamma(\mathbb{R})} \inf_{\pi \in \Pi(\mathbb{R}, \mathcal{Q})} J_{N\pi\gamma}(x), \qquad x \in \mathbb{R},\ N \in \mathbb{N}_0. \tag{4.25}$$

For clarity, we mark all quantities of the game where nature moves first which differ from the respective quantity of the original game with a tilde. The *value of a policy of nature* $\gamma \in \Gamma(\mathbb{R})$ at time $N \in \mathbb{N}_0$ is defined as

$$\widetilde{J}_{N\gamma}(x) = \inf_{\pi \in \Pi(\mathbb{R}, \mathcal{Q})} J_{N\pi\gamma}(x), \qquad x \in \mathbb{R}.$$

The Bellman operator on $\mathbb{B}$ can be introduced in the usual way:

$$\begin{aligned}\widetilde{\mathcal{T}} v(x) &= \sup_{\mathbb{Q} \in \mathcal{Q}} \inf_{a \in D(x)} Lv(x, a, \mathbb{Q})\\ &= \sup_{\mathbb{Q} \in \mathcal{Q}} \inf_{a \in D(x)} \int c(x, a, T(x, a, Z(\omega))) + \beta v(T(x, a, Z(\omega)))\ \mathbb{Q}(\mathrm{d}\,\omega), \qquad x \in \mathbb{R}.\end{aligned}$$

The infinite horizon value $J_{\infty\pi\gamma}$ of a policy pair $(\gamma, \pi) \in \Gamma(\mathbb{R}) \times \Pi(\mathbb{R}, \mathcal{Q})$ is defined as in (4.13). Consequently, the pointwise convergence is ensured by Lemma 4.15. Accordingly, one defines the *infinite horizon value of a policy of nature* $\gamma \in \Gamma(\mathbb{R})$ as

$$\widetilde{J}_{\infty\gamma}(x) = \inf_{\pi \in \Pi(\mathbb{R}, \mathcal{Q})} J_{\infty\pi\gamma}(x), \qquad x \in \mathbb{R},$$

and the value function under an infinite planning horizon as

$$\widetilde{J}_\infty(x) = \sup_{\gamma\in\Gamma(\mathbb{R})} \widetilde{J}_{\infty\gamma}(x), \qquad x \in \mathbb{R}. \tag{4.26}$$

The limit value function is also defined in the usual way

$$\widetilde{J}(x) = \lim_{N\to\infty} \widetilde{J}_N(x) = \lim_{N\to\infty} \widetilde{\mathcal{T}}^N 0(x), \qquad x \in \mathbb{R},$$

as the limit of the finite horizon value function.

Theorem 4.26. *Let Assumption 4.19 be satisfied, the ambiguity set $\mathcal{Q}$ be weak* closed and the model be convex.*

a) For $N \in \mathbb{N}_0$ the value function $\widetilde{J}_N$ lies in $\mathbb{B}$ and satisfies the Bellman equation

$$\begin{aligned}\widetilde{J}_0(x) &= 0,\\ \widetilde{J}_N(x) &= \widetilde{\mathcal{T}}\widetilde{J}_{N-1}(x), \qquad x \in \mathbb{R}.\end{aligned}$$

There exist optimal decision rules $\tilde{\gamma}_0 : \mathbb{R} \to \mathcal{Q}$ of nature and $\tilde{d}_0 : \mathbb{R} \times \mathcal{Q} \to A$ of the controller such that $\widetilde{J}_N(x) = \mathcal{T}_{\tilde{d}_0\tilde{\gamma}_0}\widetilde{J}_{N-1}(x)$. All sequences of optimal decision rules induce an optimal policy pair $\tilde{\gamma} = (\tilde{\gamma}_0, \dots, \tilde{\gamma}_{N-1}) \in \Gamma(\mathbb{R})$ and $\tilde{\pi} = (\tilde{d}_0, \dots, \tilde{d}_{N-1}) \in \Pi(\mathbb{R}, \mathcal{Q})$ satisfying $\widetilde{J}_N = J_{N\tilde{\pi}\tilde{\gamma}}$.

b) For $N \in \mathbb{N}_0$ it holds $\widetilde{J}_N = J_N$.

c) The Bellman operator $\widetilde{\mathcal{T}}$ is a contraction on $\mathbb{B}$ with modulus $\alpha\beta \in (0,1)$ and the limit value function $\widetilde{J}$ its unique fixed point in $\mathbb{B}$. It equals J.

d) There exist decision rules $\tilde{\gamma}_0 : \mathbb{R} \to \mathcal{Q}$ of nature and $\tilde{d} : \mathbb{R} \times \mathcal{Q} \to A$ of the controller such that $\widetilde{J} = \mathcal{T}_{\tilde{d}\tilde{\gamma}_0}\widetilde{J}$. Each pair of stationary policies $\tilde{\gamma} = (\tilde{\gamma}_0, \tilde{\gamma}_0, \dots) \in \Gamma^S(\mathbb{R})$ and $\tilde{\pi} = (\tilde{d}, \tilde{d}, \dots) \in \Pi^S(\mathbb{R}, \mathcal{Q})$ induced by such decision rules is optimal for the infinite horizon optimization problem (4.26), i.e. $\widetilde{J}_\infty = J_{\infty\tilde{\pi}\tilde{\gamma}}$. Furthermore, it holds $\widetilde{J}_\infty = \widetilde{J}$.

Proof. a,b) We have for $N \in \mathbb{N}_0$ and $x \in \mathbb{R}$

$$\begin{aligned}J_N(x) = \inf_{\pi\in\Pi(\mathbb{R})} \sup_{\gamma\in\Gamma(\mathbb{R},A)} J_{N\pi\gamma}(x) &\geq \inf_{\pi\in\Pi(\mathbb{R})} \sup_{\gamma\in\Gamma(\mathbb{R})} J_{N\pi\gamma}(x)\\ &\geq \sup_{\gamma\in\Gamma(\mathbb{R})} \inf_{\pi\in\Pi(\mathbb{R})} J_{N\pi\gamma}(x)\\ &\geq \sup_{\gamma\in\Gamma(\mathbb{R})} \inf_{\pi\in\Pi(\mathbb{R},\mathbb{Q})} J_{N\pi\gamma}(x) = \widetilde{J}_N(x).\end{aligned} \tag{4.27}$$

Note that the second inequality holds generally for the interchange of infimum and supremum. Let $\pi^* = (d_0^*, \dots, d_{N-1}^*) \in \Pi(\mathbb{R})$ and $\gamma^* = (\gamma_0^*, \dots, \gamma_{N-1}^*) \in \Gamma(\mathbb{R}, A)$ be optimal strategies for the original game (4.4). The existence is guaranteed by Theorem 4.12. Then $\tilde{\gamma} = (\tilde{\gamma}_0, \dots, \tilde{\gamma}_{N-1})$ defined by $\tilde{\gamma}_n = \gamma_n^*(\cdot, d_n^*(\cdot))$ lies in $\Gamma(\mathbb{R})$ since the decision rules are well-defined as compositions of measurable maps.

By (forward) induction we prove a) and b) and that $\tilde{\gamma}$ constitutes an optimal policy of nature in (4.25). For $N = 0$ there is nothing to show. Now assume the assertion holds at time $N-1$. With the forward form of the value iteration (4.10) one obtains

$$\begin{aligned}\widetilde{J}_{N\tilde{\gamma}}(x) &= \inf_{\pi\in\Pi(\mathbb{R},\mathcal{Q})} J_{N\pi\tilde{\gamma}}(x)\\ &= \inf_{\pi\in\Pi(\mathbb{R},\mathcal{Q})} \int c\big(x, d_0(x,\tilde{\gamma}_0(x)), T(x, d_0(x,\tilde{\gamma}_0(x)), Z(\omega))\big)\\ &\qquad\qquad + \beta \widetilde{J}_{N-1\vec{\pi}\vec{\tilde{\gamma}}}\big(T(x, d_0(x,\tilde{\gamma}_0(x)), Z(\omega))\big)\, \tilde{\gamma}_0(\mathrm{d}\,\omega|x).\end{aligned}$$

By the induction hypothesis $\widetilde{J}_{N-1\vec{\tilde{\gamma}}} = \widetilde{J}_{N-1} = J_{N-1}$ is measurable as a lower semicontinuous function. Hence, we can estimate

$$\begin{aligned}&\geq \inf_{d_0} \int c\big(x, d_0(x,\tilde{\gamma}_0(x)), T(x, d_0(x,\tilde{\gamma}_0(x)), Z(\omega))\big)\\ &\qquad + \beta \widetilde{J}_{N-1\vec{\tilde{\gamma}}}\big(T(x, d_0(x,\tilde{\gamma}_0(x)), Z(\omega))\big)\, \tilde{\gamma}_0(\mathrm{d}\,\omega|x).\end{aligned}$$

The minimization only depends on $d_0(x, \gamma_0(x)) \in D(x)$, i.e.

$$\begin{aligned}&= \inf_{a\in D(x)} \int c\big(x, a, T(x,a,Z(\omega))\big) + \beta \widetilde{J}_{N-1\vec{\tilde{\gamma}}}\big(T(x,a,Z(\omega))\big)\, \tilde{\gamma}_0(\mathrm{d}\,\omega|x)\\ &= \int \inf_{a\in D(x)} \int c\big(x, a, T(x,a,Z(\omega))\big)\\ &\qquad\qquad + \beta \widetilde{J}_{N-1\vec{\tilde{\gamma}}}\big(T(x,a,Z(\omega))\big)\, \mathbb{Q}(\mathrm{d}\,\omega)\, \delta_{\tilde{\gamma}_0(x)}(\mathrm{d}\,\mathbb{Q}).\end{aligned}$$

Given existence of a minimizing decision rule $\tilde{d}_0 : \mathbb{R}\times\mathbb{Q} \to A$ one ontains the identity

$$= \widetilde{\mathcal{T}}_{\tilde{d}_0\gamma_0} \widetilde{J}_{N-1\vec{\tilde{\gamma}}}(x). \tag{4.28}$$

Again by the induction hypothesis, there is an optimal policy $\vec{\tilde{\pi}} = (\tilde{d}_1, \ldots, \tilde{d}_{N-1}) \in \Pi(\mathbb{R}, \mathcal{Q})$ such that

$$= \widetilde{\mathcal{T}}_{\tilde{d}_0\tilde{\gamma}_0} J_{N-1\vec{\tilde{\pi}}\vec{\tilde{\gamma}}}(x),$$

which equals by the forward form of the value iteration (4.10)

$$\begin{aligned}&= J_{N\tilde{\pi}\tilde{\gamma}}(x)\\ &\geq \widetilde{J}_{N\tilde{\gamma}}(x).\end{aligned}$$

Next, we verify the existence of a minimizing decision rule $\tilde{d}_0$ at (4.28). To that end, we show that the function

$$\mathbb{R}\times A\times \mathcal{Q} \ni (x,a,\mathbb{Q}) \mapsto Lv(x,a,\mathbb{Q}) = \int c(x,a,T(x,a,z))+\beta v(T(x,a,Z(\omega)))\mathbb{Q}(\mathrm{d}\,\omega)$$

is jointly lower semicontinuous for any $v \in \mathbb{B}$. Recall that $\mathcal{Q}$ endowed with the weak* topology is a compact Borel space due to Lemma 4.1. Let $\{(x_n, a_n, \mathbb{Q}_n)\}_{n\in\mathbb{N}}$ be a convergent sequence in $\mathbb{R} \times A \times \mathcal{Q}$ with limit $(x^*, a^*, \mathbb{Q}^*)$. The increasing sequence of random variables $\{C_n\}_{n\in\mathbb{N}}$ given by

$$C_n(\omega) = \inf_{k\geq n} c(x_k, a_k, T(x_k, a_k, Z(\omega))) + \beta v(T(x_k, a_k, Z(\omega))), \qquad \omega \in \Omega_1$$

has an absolute bound in $L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ according to Lemma 4.7 and is therefore convergent. By Lemma A.4 b) the function $D \ni (x,a) \mapsto c(x,a,T(x,a,Z(\omega))) + v(T(x,a,Z(\omega)))$ is lower semicontinuous for every $\omega \in \Omega_1$. Consequently,

$$C^*(\omega) = \lim_{n\to\infty} C_n(\omega) \geq c(x^*, a^*, T(x^*, a^*, Z(\omega))) + \beta v(T(x^*, a^*, Z(\omega))), \quad \omega \in \Omega_1.$$

By dominated convergence we get

$$C_n \xrightarrow{L^p} C^* \geq c(x^*, a^*, T(x^*, a^*, Z)) + \beta v(T(x^*, a^*, Z))).$$

Since $\mathcal{Q}$ is norm bounded, Corollary 6.40 in Aliprantis and Border (2006) yields that the duality

$$(X, \mathbb{Q}) \mapsto \mathbb{E}^{\mathbb{Q}}[X]$$

restricted to $L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \times \mathcal{Q}$ is jointly continuous, where $L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ is considered with the norm topology and $\mathcal{Q}$ with the weak* topology. Thus, we get

$$\begin{aligned}
\liminf_{n\to\infty} Lv(x_n, a_n, \mathbb{Q}_n) &= \liminf_{n\to\infty} \mathbb{E}^{\mathbb{Q}_n}\Big[c(x_n, a_n, T(x_n, a_n, Z)) + \beta v(T(x_n, a_n, Z))\Big] \\
&\geq \liminf_{n\to\infty} \mathbb{E}^{\mathbb{Q}_n}\Big[\inf_{k\geq n} c(x_k, a_k, T(x_k, a_k, Z)) + \beta v(T(x_k, a_k, Z))\Big] \\
&= \lim_{n\to\infty} \mathbb{E}^{\mathbb{Q}_n}[C_n] \\
&= \mathbb{E}^{\mathbb{Q}^*}[C^*] \\
&\geq \mathbb{E}^{\mathbb{Q}^*}\Big[c(x^*, a^*, T(x^*, a^*, Z)) + \beta v(T(x^*, a^*, Z))\Big] \\
&= Lv(x^*, a^*, \mathbb{Q}^*),
\end{aligned}$$

which establishes the joint lower semicontinuity of $Lv(\cdot)$. Note that $\widetilde{J}_{N-1\vec{\gamma}} \in \mathbb{B}$ and $(x, \mathbb{Q}) \mapsto D(x)$ is a compact-valued and upper semicontinuous. Hence, it follows from Theorem A.25 that there exists a a minimizing decision rule $\tilde{d}_0 : \mathbb{R} \times \mathcal{Q} \to A$ at (4.28) and that

$$\mathbb{R} \times \mathcal{Q} \ni (x, \mathbb{Q}) \mapsto \inf_{a\in D(x)} L\widetilde{J}_{N-1\vec{\gamma}}(x, a, \mathbb{Q}) = L\widetilde{J}_{N-1\vec{\gamma}}(x, \tilde{d}_0(x, \mathbb{Q}), \mathbb{Q})$$

is lower semicontinuous.

By the induction hypothesis we have $\widetilde{J}_{N-1} = J_{N-1}$. From Theorem 4.24 and subsequent remarks together with Proposition 4.21 and Theorem 4.12 it follows that $(d_0^*(x), \tilde{\gamma}_0(x)) = (d_0^*(x), \gamma_0^*(x, d_0^*(x)))$ is a saddle point of $(a, \mathbb{Q}) \mapsto L\widetilde{J}_{N-1}(x, a, \mathbb{Q})$ for every $x \in \mathbb{R}$, i.e.

$$J_N(x) = L\widetilde{J}_{N-1}(x, d_0^*(x), \tilde{\gamma}_0(x)) \leq L\widetilde{J}_{N-1}(x, a, \tilde{\gamma}_0(x)), \qquad \text{for all } a \in D(x).$$

Hence, $J_N(x) \leq \inf_{a\in D(x)} L\widetilde{J}_{N-1\vec{\tilde{\gamma}}}(x, a, \tilde{\gamma}_0(x)) = \widetilde{J}_{N\tilde{\gamma}}(x)$ for all $x \in \mathbb{R}$. Due to (4.27) it follows $J_N = \tilde{J}_N = \widetilde{J}_{N\tilde{\gamma}}$, i.e. the policy $\tilde{\gamma}$ must be optimal for nature. Especially, $\widetilde{J}_{N\tilde{\gamma}}$ is in $\mathbb{B}$.

Since $J_N = \tilde{J}_N$, the joint Bellman equation for the controller and nature $\widetilde{J}_N = \widetilde{\mathcal{T}}\widetilde{J}_{N-1}$ follows from Theorem 4.24.

c) Due to Theorem 4.24 the Bellman Operators $\mathcal{T}$ and $\widetilde{\mathcal{T}}$ coincide. Thus, the assertion follows from Theorem 4.18 a).

d) The existence of decision rules $\tilde{\gamma}_0$ and $\tilde{d}$ such that $\widetilde{J} = \mathcal{T}_{\tilde{d}\tilde{\gamma}_0}\widetilde{J}$ is due to part a). Finally, the existence of a stationary optimal policy pair follows analogously to the proof of Theorem 4.18. □

As a direct consequence, we get the existence of Nash equilibria on policy level.

Corollary 4.27. *Consider a convex model with weak* closed ambiguity set $\mathcal{Q}$ and Assumption 4.19 fulfilled.*

a) Let the planning horizon $N \in \mathbb{N}_0$ be finite. Then it holds for $x \in \mathbb{R}$

$$J_N(x) = \min_{\pi\in\Pi(\mathbb{R})} \max_{\gamma\in\Gamma(\mathbb{R},A)} J_{N\pi\gamma}(x) = \max_{\gamma\in\Gamma(\mathbb{R})} \min_{\pi\in\Pi(\mathbb{R},\mathcal{Q})} J_{N\pi\gamma}(x) = \tilde{J}_N(x).$$

Consequently, it even holds

$$J_N(x) = \min_{\pi\in\Pi(\mathbb{R})} \max_{\gamma\in\Gamma(\mathbb{R})} J_{N\pi\gamma}(x) = \max_{\gamma\in\Gamma(\mathbb{R})} \min_{\pi\in\Pi(\mathbb{R})} J_{N\pi\gamma}(x).$$

b) The statement of part a) holds for an infinite planning horizon, too. However, Markov strategies can be replaced by stationary ones.

Proof. a) Theorem 4.26 implies equality in (4.27), i.e.

$$\begin{aligned} J_N(x) &= \min_{\pi\in\Pi(\mathbb{R})} \max_{\gamma\in\Gamma(\mathbb{R},A)} J_{N\pi\gamma}(x) \\ &= \inf_{\pi\in\Pi(\mathbb{R})} \sup_{\gamma\in\Gamma(\mathbb{R})} J_{N\pi\gamma}(x) \\ &= \sup_{\gamma\in\Gamma(\mathbb{R})} \inf_{\pi\in\Pi(\mathbb{R})} J_{N\pi\gamma}(x) \end{aligned}$$

$$= \max_{\gamma \in \Gamma(\mathbb{R})} \min_{\pi \in \Pi(\mathbb{R}, \mathbb{Q})} J_{N\pi\gamma}(x) = \widetilde{J}_N(x). \tag{4.29}$$

It remains to find optimal policies for the second and third line of (4.29). Let

$$\begin{aligned} &\pi^* = (d_0^*, \dots, d_{N-1}^*) \in \Pi(\mathbb{R}), \qquad &&\gamma^* = (\gamma_0^*, \dots, \gamma_{N-1}^*) \in \Gamma(\mathbb{R}, A) \\ \text{and} \qquad &\tilde{\gamma} = (\tilde{\gamma}_0, \dots, \tilde{\gamma}_{N-1}) \in \Gamma(\mathbb{R}) &&\tilde{\pi} = (\tilde{d}_0, \dots, \tilde{d}_{N-1}) \in \Pi(\mathbb{R}, \mathcal{Q}) \end{aligned}$$

be optimal strategies for the first and forth line of (4.29), respectively, which exist by Proposition 4.21 and Theorem 4.26. Then

$$\inf_{\pi \in \Pi(\mathbb{R})} \sup_{\gamma \in \Gamma(\mathbb{R})} J_{N\pi\gamma} = \sup_{\gamma \in \Gamma(\mathbb{R})} \inf_{\pi \in \Pi(\mathbb{R})} J_{N\pi\gamma}$$

is attained by the admissible strategy pair $(\hat{\pi}, \hat{\gamma}) \in \Pi(\mathbb{R}) \times \Gamma(\mathbb{R})$ which can be defined by $\hat{d}_n = d_n^*$ and $\hat{\gamma}_n = \gamma_n^*(\cdot, d_n^*(\cdot))$ or alternatively by $\hat{d}_n = \tilde{d}_n(\cdot, \tilde{\gamma}_n(\cdot))$ and $\hat{\gamma}_n = \tilde{\gamma}_n$ for $n = 0, \dots, N-1$.

b) One can apply the same arguments as in a). □

The game-theoretic interpretation of Corollary 4.27 is that in a convex model our original game (4.4), where the controller moves first, yields the same optimal expected cost as a corresponding game (4.25) where the controller reacts to a move of nature. Moreover, it states that the corresponding game where the controller and nature move simultaneously and unaware of the other's action has a Nash equilibrium in deterministic Markov (or stationary in case of an infinite planning horizon) strategies. The expected cost in this equilibrium is also equal to the optimal expected cost of our original game.

Since the assertion of Corollary 4.27 holds for every point in time, the simultaneous move game does not only have a Nash equilibrium but a *subgame perfect equilibrium* in the sense of Selten (1975), i.e. there is a strategy pair inducing a Nash equilibrium in any subgame that starts at some time point n conditional on an arbitrary admissible history h_n. More specifically, we have a *(stationary) Markov perfect equilibrium* as defined by Maskin and Tirole (1988).

The fact that in a convex model our game against nature is equivalent to a simultaneous move game is closely related to the concept of *s-rectangularity* of ambiguity sets which was introduced by Wiesemann et al. (2012) in the context of robust Markov Decision Processes. While a rectangular ambiguity set in the sense of Iyengar (2005) allows nature to assign at each point in time to each state-action combination $(x, a) \in D$ the most adverse disturbance distribution, an s-rectangular ambiguity set allows only for a statewise change of the disturbance distribution at each stage.

4.3.2. Special Ambiguity Sets

In this section, we consider some special choices for the ambiguity set $\mathcal{Q}$ which simplify solving the Markov Decision Problems (4.4) and (4.14) or allow for structural statements about the solution.

Convex hull. It does not change the optimal value of the optimization problems if a given ambiguity set $\mathcal{Q}$ is replaced by its convex hull $\operatorname{conv}(\mathcal{Q})$ or its closed convex hull $\overline{\operatorname{conv}}(\mathcal{Q})$, where the closure is with respect to the weak* topology. To demonstrate this, it suffices to compare the corresponding Bellman equations.

Lemma 4.28. *Let $\mathcal{Q}$ be any norm bounded ambiguity set. Then it holds for all $v \in \mathbb{B}$ and $x \in \mathbb{R}$*

$$\inf_{a \in D(x)} \sup_{\mathbb{Q} \in \mathcal{Q}} Lv(x, a, \mathbb{Q}) = \inf_{a \in D(x)} \sup_{\mathbb{Q} \in \operatorname{conv}(\mathcal{Q})} Lv(x, a, \mathbb{Q}) = \inf_{a \in D(x)} \sup_{\mathbb{Q} \in \overline{\operatorname{conv}}(\mathcal{Q})} Lv(x, a, \mathbb{Q}).$$

Proof. Fix $(x, a) \in D$. The function $\mathbb{Q} \mapsto Lv(x, a, \mathbb{Q})$ is linear. Thus, for a generic element $\mathbb{Q} = \sum_{i=1}^{n} \lambda_i \mathbb{Q}_i \in \operatorname{conv}(\mathcal{Q})$ we have

$$Lv\left(x, a, \sum_{i=1}^{n} \lambda_i \mathbb{Q}_i\right) = \sum_{i=1}^{n} \lambda_i Lv(x, a, \mathbb{Q}_i) \leq \max_{i=1,\dots,n} Lv(x, a, \mathbb{Q}_i),$$

i.e. there can be no improvement of the supremum on the convex hull. With the same arguments as in the proof of Lemma 4.1, we find that $\overline{\operatorname{conv}}(\mathcal{Q})$ is metrizable and therefore coincides with the limit points of sequences in $\operatorname{conv}(\mathcal{Q})$. Since $\mathbb{Q} \mapsto Lv(x, a, \mathbb{Q})$ is weak* continuous (cf. proof of Theorem 4.8), the supremum cannot be improved on the closure either. □

From Theorem 2.21 and Remark 2.22 we know that ambiguity sets induced by the dual representation of a coherent risk measure with the Fatou property coincide with their closed convex hull. Nonetheless, Lemma 4.28 has a useful application. In the context of optimal (re)insurance, Birghila and Pflug (2019) suggested constructing an ambiguity set as the convex hull of a finite number of probability measures $\mathcal{Q} = \operatorname{conv}\{\mathbb{Q}_1, \dots, \mathbb{Q}_m\}$. E.g. this is an obvious choice when different "scenarios" for the disturbance distribution are derived from expert opinions. With such an ambiguity set, the Bellman equation simplifies to

$$J_n(x) = \inf_{a \in D(x)} \max_{i=1,\dots,m} \mathbb{E}^{\mathbb{Q}_i}\Big[c\big(x, a, T(x, a, Z)\big) + \beta J_{n+1}\big(T(x, a, Z)\big)\Big], \quad x \in \mathbb{R}.$$

From a computational perspective, this is an advantageous situation. Note that the conclusions hold for general state spaces, too.

Integral stochastic orders on $\mathcal{Q}$. Following an idea of Müller (1997), one can define integral stochastic orders on the ambiguity $\mathcal{Q}$ set by

$$\mathbb{Q}_1 \leq_{\mathbb{B},x,a} \mathbb{Q}_2 \;:\Longleftrightarrow\; \int c(x, a, T(x, a, Z(\omega))) + \beta v(T(x, a, Z(\omega)))\mathbb{Q}_1(\mathrm{d}\,\omega)$$

$$\leq \int c(x,a,T(x,a,Z(\omega))) + \beta v(T(x,a,Z(\omega)))\mathbb{Q}_2(\mathrm{d}\,\omega) \quad \text{for all } v \in \mathbb{B}$$

where $(x,a) \in D$ is fixed and

$$\mathbb{Q}_1 \leq_{\mathbb{B}} \mathbb{Q}_2 \;:\Longleftrightarrow\; \mathbb{Q}_1 \leq_{\mathbb{B},x,a} \mathbb{Q}_2 \quad \text{for all } (x,a) \in D.$$

If there exists a maximal element with respect to one of these stochastic orders, this probability measure is an optimal action for nature (in the respective scenario).

Lemma 4.29. *a) If there exists a maximal element $\mathbb{Q}_{x,a} \in \mathcal{Q}$ w.r.t. $\leq_{\mathbb{B},x,a}$ for every $(x,a) \in D$, then $\gamma = (\gamma_0, \gamma_1, \dots)$ with $\gamma_n(x,a) = \mathbb{Q}_{x,a}$ defines a stationary optimal policy of nature in both* (4.4) *and* (4.14)*.*

b) If there exists a maximal element $\mathbb{Q}^* \in \mathcal{Q}$ w.r.t. $\leq_{\mathbb{B}}$, then $\gamma = (\gamma_0, \gamma_1, \dots)$ with $\gamma_n \equiv \mathbb{Q}^*$ defines a constant optimal action of nature. That is, (4.4) *and* (4.14) *can be reformulated to risk-neutral MDP under the probability measure $\mathbb{Q}^*$.*

Proof. a) Fix $(x,a) \in D$ and let $n \in \{0,\dots,N-1\}$ and $\mathbb{Q} \in \mathcal{Q}$ be arbitrary. Since $J_{n+1} \in \mathbb{B}$ by Theorem 4.12, $\mathbb{Q} \leq_{\mathbb{B},x,a} \mathbb{Q}_{x,a}$ implies

$$\begin{aligned} LJ_{n+1}(x,a,\mathbb{Q}) &= \int c(x,a,T(x,a,Z(\omega))) + \beta v(T(x,a,Z(\omega)))\mathbb{Q}(\mathrm{d}\,\omega) \\ &\leq \int c(x,a,(T(x,a,Z(\omega))) + \beta v(T(x,a,Z(\omega)))\mathbb{Q}_{x,a}(\mathrm{d}\,\omega) \\ &= LJ_{n+1}(x,a,\mathbb{Q}_{x,a}). \end{aligned}$$

Hence, $Q_{x,a}$ is a maximizing action of nature and the selection $\gamma_n(x,a) = \mathbb{Q}_{x,a}$ is measurable by Theorem 4.8. Letting $N \to \infty$ yields the assertion for the infinite horizon case.

b) This follows directly from a). □

In fact, Lemma 4.29 holds for any state space. But it is only a reformulation of what is an optimal action for nature. However, under Assumption 4.19 it has practical relevance when a simpler sufficient condition for the integral stochastic order $\leq_{\mathbb{B}}$ is fulfilled. We give three exemplary criteria:

1. Let the one-stage cost function c be increasing in x'. Further, let $\mathcal{Z}$ be a partially ordered space, e.g. $\mathcal{Z} = \mathbb{R}^m$, and assume that the transition function is increasing in z. Then the functions

$$\mathcal{Z} \ni z \mapsto c(x,a,T(x,a,z)) + \beta v(T(x,a,z)), \qquad v \in \mathbb{B},\ (x,a) \in D \tag{4.30}$$

are increasing. Thus, $\mathbb{Q}_1 \leq_{\mathbb{B}} \mathbb{Q}_2$ is implied by the usual stochastic order of the disturbance distributions $\mathbb{Q}_1^Z \leq_{st} \mathbb{Q}_2^Z$ and a maximal element of $\mathcal{Q}$ w.r.t. $\leq_{st}$ allows the same conclusion as in Lemma 4.29 b).

2. Let the one-stage cost function c be increasing in x', let $\mathcal{Z}$ be a real vector space, assume a convex model (cf. Lemma 4.23) and let the transition function T additionally be convex in z. For T this means that it is (jointly) convex in (x, a) and convex in z, i.e. componentwise convex as a function with two arguments. Simply requiring T to be convex is of course sufficient.
   Now, Lemma 4.23 yields that the functions (4.30) are convex as compositions of increasing convex and convex mappings. Consequently, $\leq_{\mathbb{B}}$ is implied by the convex order $\leq_{cx}$ of the disturbance distributions $\mathbb{Q}^Z$.

3. Combining the requirements of 1. and 2., the sufficient condition can be weakened to the increasing convex order $\leq_{icx}$.

Convex order on the set of densities. Since the probability measures in $\mathcal{Q}$ are absolutely continuous with respect to the reference probability measure $\mathbb{P}_1$, we can alternatively consider the set of densities

$$\mathcal{Q}^d = \left\{ \frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_1} \in L^q(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) : \mathbb{Q} \in \mathcal{Q} \right\}.$$

In general, one has to take care both of the marginal distribution of the density and the dependence structure with the random cost when searching for a maximizing density of the Bellman equation

$$\inf_{a \in D(x)} \sup_{Y \in \mathcal{Q}^d} \mathbb{E}\Big[\Big(c(x, a, T(x, a, Z)) + \beta J_{n+1}(T(x, a, Z))\Big) Y\Big].$$

However, if $\mathcal{Q}^d$ is sufficiently rich, the maximization reduces to comparing marginal distributions.

Definition 4.30. The set of densities $\mathcal{Q}^d$ is called *law invariant* if for $Y_1 \in \mathcal{Q}^d$ every $Y_2 \in L^q(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ with $Y_2 \sim Y_1$ is in $\mathcal{Q}^d$, too.

Lemma 4.31. *Let Assumption 4.19 be satisfied and the one-stage cost function c be increasing in x'. If $\mathcal{Q}^d$ is law invariant, the supremum*

$$\sup_{Y \in \mathcal{Q}^d} \mathbb{E}\Big[\Big(c(x, a, T(x, a, Z)) + \beta v(T(x, a, Z))\Big) Y\Big], \qquad (x, a) \in D,\ v \in \mathbb{B},$$

is not changed by restricting the maximization to densities which are comonotonic to the random variable $T(x, a, Z)$.

Proof. For random vectors (X_1, X_2) and (Y_1, Y_2) with the same marginals it follows from the upper Fréchet-Hoeffding bound that

$$\mathbb{E}[X_1 X_2] \leq \mathbb{E}[Y_1 Y_2]$$

if (Y_1, Y_2) is comonotonic and the expectations exist, cf. Müller and Stoyan (2002, 3.1.1, 3.8.2). By Lemma 4.7 the random variable $c(x, a, T(x, a, Z)) + \beta v(T(x, a, Z))$ is in

$L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ for all $(x, a) \in D$. Thus the expectation

$$\mathbb{E}\Big[\Big(c(x, a, T(x, a, Z)) + \beta v(T(x, a, Z))\Big)Y\Big]$$

exists for all $Y \in L^q(\mathcal{Z}, \mathfrak{Z}, \mathbb{P}^Z)$. I.e. due to the law invariance of $\mathcal{Q}^d$ we can find for every $Y \in \mathcal{Q}^d$ some $Y' \in \mathcal{Q}^d$ comonotonic to $c(x, a, T(x, a, Z)) + \beta v(T(x, a, Z))$ such that $Y' \sim Y$ and

$$\mathbb{E}\Big[\Big(c(x, a, T(x, a, Z)) + \beta v(T(x, a, Z))\Big)Y\Big] \leq \mathbb{E}\Big[\Big(c(x, a, T(x, a, Z)) + \beta v(T(x, a, Z))\Big)Y'\Big].$$

Since, the function $\mathbb{R} \ni x' \mapsto c(x, a, x') + \beta v(x')$ is increasing, this is the same as requiring comonotonicity to $T(x, a, Z)$. □

For the comparison of marginal distributions one would naturally think of stochastic orders. Here, the convex order yields a sufficient criterion for optimality.

Lemma 4.32. *Let Assumption 4.19 be satisfied, the one-stage cost function c be increasing in x', $\mathcal{Q}^d$ be law invariant and suppose there exists a maximal element Y^* of $\mathcal{Q}^d$ w.r.t. the convex order $\leq_{cx}$.*

a) Then

$$\rho_\phi(X) = \sup_{Y \in \mathcal{Q}^d} \mathbb{E}[XY], \qquad X \in L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1),$$

defines a spectral risk measure with spectrum $\phi(u) = q^+_{Y^*}(u),\ u \in [0, 1]$. In this case, $\gamma = (\gamma_0, \gamma_1, \dots)$ with $\gamma_n(x, a) = \phi(U_{T(x,a,Z)})$ is a stationary optimal strategy of nature in both (4.4) and (4.14). Here, $q^+_{Y^*}$ denotes the upper quantile function of Y^* and $U_{T(x,a,Z)}$ the distributional transform of $T(x, a, Z)$.

b) If additionally the disturbance space is the real line $\mathcal{Z} = \mathbb{R}$ and the transition function T is increasing and lower semicontinuous in z, $\gamma = (\gamma_0, \gamma_1, \dots)$ with $\gamma_n \equiv \phi(U_Z)$ defines a constant optimal action of nature. That is, (4.4) and (4.14) can be reformulated to a risk-neutral MDP with probability measures $\mathrm{d}\,\mathbb{Q} = \phi(U_Z)\,\mathrm{d}\,\mathbb{P}_1$.

Proof. a) It holds $Y^* = q^+_{Y^*}(U_{Y^*})$ $\mathbb{P}$-a.s. by Lemma B.10. Therefore,

$$\mathcal{Q}^d \subseteq \{Y \in L^q(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) :\ Y \leq_{cx} \phi(U),\ U \sim \mathcal{U}(0, 1)\},$$

and the random variables $\phi(U),\ U \sim \mathcal{U}(0, 1)$ are contained in both sets due to law invariance. By Proposition 2.24 ρ_ϕ indeed defines a spectral risk measure and $\phi(\tilde{U})$ is an optimal action of nature at time n given (x, a), where $\tilde{U}$ is the distributional transform of the random variable

$$c(x, a, T(x, a, Z)) + \beta J_{n+1}(T(x, a, Z)).$$

Since the function $\mathbb{R} \ni x' \mapsto c(x, a, x') + \beta v(x')$ is increasing and lower semicontinuous, i.e. left continuous by Lemma A.6, it follows from Lemma B.11 that $\tilde{U} = U_{T(x,a,Z)}$.

b) Under the additional assumptions we have again by Lemma B.11 that $U_{T(x,a,Z)} = U_Z$. □

Recall that the probability space under consideration is the product space

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{k=1}^{\infty} (\Omega_1, \mathcal{A}_1, \mathbb{P}_1).$$

Under the assumptions of Lemma 4.32 b) we can replace the probability measure $\mathbb{P}$ by

$$\widehat{\mathbb{Q}} = \bigotimes_{k=1}^{\infty} \mathbb{Q}^*, \qquad \mathrm{d}\,\mathbb{Q}^* = \phi(U_Z)\,\mathrm{d}\,\mathbb{P}_1$$

and the optimization problems (4.4) and (4.14) can be equivalently written as

$$\inf_{\pi \in \Pi^M} \mathbb{E}^{\widehat{\mathbb{Q}}} \left[\sum_{k=0}^{N-1} \beta^k c(X_k, d_k(X_k), X_{k+1}) \right], \tag{4.31}$$

where $N \in \mathbb{N} \cup \{\infty\}$. With the reversed argumentation of Lemma 4.32, a robust formulation of (4.31) is given by

$$\inf_{\pi \in \Pi^M} \sup_{\mathbb{Q} \in \mathfrak{Q}} \mathbb{E}^{\mathbb{Q}} \left[\sum_{k=0}^{N-1} \beta^k c(X_k, d_k(X_k), X_{k+1}) \right] \tag{4.32}$$

where

$$\mathfrak{Q} = \left\{ \bigotimes_{k=1}^{\infty} \mathbb{Q}_k : \ \mathrm{d}\,\mathbb{Q}_k = Y_k \,\mathrm{d}\,\mathbb{P}_1, \ Y_k \in L^q(\Omega_1, \mathcal{A}_1, \mathbb{P}_1), \ Y_k \leq_{cx} \phi(U), \ U \sim \mathcal{U}(0,1) \right\}.$$

The $Y_k, \ k \in \mathbb{N}$, are indeed densities by Remark 2.25. Now, (4.32) can be interpreted as the minimization of the coherent risk measure

$$\tilde{\rho}(X) = \sup_{\mathbb{Q} \in \mathfrak{Q}} \mathbb{E}^{\mathbb{Q}}[X], \qquad X \in L^p(\Omega, \mathcal{A}, \mathbb{P}) \tag{4.33}$$

on the product space $(\Omega, \mathcal{A}, \mathbb{P})$ applied to the discounted total cost.

4.4. Applications

In this section, we study the distributionally robust maximization of the expected dividend payment for an insurance company in the dynamic reinsurance model of Section 3.2.2 as a concrete actuarial application. Moreover, we apply the distributionally robust optimality criterion to the class of stochastic linear-quadratic problems.

4.4.1. Robust Dividend Maximization of an Insurance Company

Chen and Assa (2019) introduced a dynamic reinsurance model with the maximization of the discounted lifetime dividends of the insurance company as optimality criterion. This optimality criterion can be interpreted in two ways: Either as the insurer's actual objective. In this case, one is especially interested in the optimal reinsurance and dividend policy. Alternatively, this optimality criterion is used for the valuation of the insurer's portfolio (Dividend Discount Model). In this case, one is only interested in the value of the objective function.

The results of Chen and Assa (2019) contain a fundamental error in Section 4.2 disregarding the dynamic nature of the optimization problem. In Section 3.2.2 we gave a slightly modified version of their model, which can be seen as a special case of our abstract cost model. Hence, the results of Chapter 4 can be applied to correctly ensure the existence of an optimal reinsurance and dividend policy. Risk-neutral (or unambiguous) dividend maximization as in Chen and Assa (2019) is the special case when the ambiguity set is a singleton.

Under a finite planning horizon, the value of a policy pair $(\pi, \gamma) \in \Pi^R \times \Gamma$ of the insurer and nature at time $n = 0, \dots, N-1$ is defined as

$$\begin{aligned} V_{N\pi\gamma}(h_N) &= -x_N^+, & h_N &\in \mathcal{H}_N, \\ V_{n\pi\gamma}(h_n) &= \mathbb{E}_{nh_n}^{\pi\gamma}\left[-\sum_{k=n}^{N-1} \beta^{k-n} A_k - \beta^{N-n} X_N^+\right], & h_n &\in \mathcal{H}_n, \end{aligned}$$

since we want to treat the dividends as negative costs. The corresponding value functions are

$$V_n(h_n) = \inf_{\pi \in \Pi^R} \sup_{\gamma \in \Gamma} V_{n\pi\gamma}(h_n), \qquad h_n \in \mathcal{H}_n,$$

and the optimization objective is to determine the robust maximal discounted dividend

$$V_0(x) = \inf_{\pi \in \Pi^R} \sup_{\gamma \in \Gamma} V_{0\pi\gamma}(x), \qquad x \in \mathbb{R}. \tag{4.34}$$

Due to the real state space we want to apply Corollary 4.22 for solving the finite horizon optimization problem. Let us verify the assumptions. The numbering is as in the corollary.

(i) The state space is the real line $E = \mathbb{R}$.

(ii') The Continuity and Compactness Properties 3.1 with upper semicontinuous transition function have been verified in Section 3.2.2.

(iii') Monotonicity properties:

(iii' a) The set-valued mapping

$$\mathbb{R} \ni x \mapsto D(x) = \Big\{(a, f) \in \mathbb{R}_+ \times \mathcal{F} : \ a \le x^+, \ \rho(f(Y)) \le x^+ + z - a - \pi_R(f)\Big\}$$

is increasing.

(iii' b) The transition function $T : \mathbb{R} \times \mathcal{F} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, $T(x, f, y, z) = x + z - a - f(y) - \pi_R(f)$ is increasing in x.

(iii' c) The terminal cost function $c_N(x) = -x^+$ is decreasing and the composition $\mathbb{R} \ni x \mapsto c(x, a, f, T(x, a, f, y)) = -a$ is decreasing for all (a, f, y).

(iv) Obviously, $\bar{b} \equiv 0$ is an upper bounding function for every $\alpha \geq 1$. We now show that

$$\underline{b}(x) = -\frac{x^+}{1-\beta} - \frac{z}{(1-\beta)^2}, \qquad x \in \mathbb{R},$$

is a lower bounding function, where $\alpha = \frac{1-(1-\beta)^2}{\beta}$. We have for all $(x, a, f) \in D$ and $\mathbb{Q} \in \mathcal{Q}$:

$$\begin{aligned}
-c_N^-(x) &= -x^+ \geq \underline{b}(x), \\
-c^-(x, a, f, T(x, a, f, Y))) &= -a \geq -x^+ \geq \underline{b}(x), \\
\mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T(x, a, f, Y))\right] &= -\frac{z}{(1-\beta)^2} - \frac{1}{1-\beta}\mathbb{E}^{\mathbb{Q}}\left[(x + z - a - f(Y) - \pi_R(f))^+\right] \\
&\geq -\frac{z}{(1-\beta)^2} - \frac{1}{1-\beta}\mathbb{E}^{\mathbb{Q}}\left[(x+z)^+\right] \\
&\geq -z\left(\frac{1}{1-\beta} + \frac{1}{(1-\beta)^2}\right) - \frac{x^+}{1-\beta} \\
&= -x^+\frac{1}{\beta}\left(\frac{1}{1-\beta} - 1\right) - z\frac{1}{\beta}\left(\frac{1}{(1-\beta)^2} - 1\right) \\
&= \frac{1}{\beta}\left(\underline{b}(x) + x^+ + z\right) \\
&\geq \frac{1}{\beta}\left(\underline{b}(x) + (1-\beta)x^+ + z\right) \\
&= \frac{1-(1-\beta)^2}{\beta}\underline{b}(x).
\end{aligned}$$

The second equality holds since

$$\frac{1}{1-\beta} = 1 + \frac{\beta}{1-\beta} \qquad \text{and} \qquad 1 + \frac{\beta}{1-\beta} + \frac{\beta}{(1-\beta)^2} = \frac{1}{(1-\beta)^2}.$$

The (absolute) bounding function is given by $b = \bar{b} - \underline{b} = -\underline{b}$.

(v) Here, we use the separation condition of Lemma 4.20:

$$|c(x, a, f, T(x, a, f, Y))| = a \leq x^+$$

i.e $\vartheta_1(x) = x^+$, which is continuous, and $\Theta_1(Y) = 0 \in L^p(\Omega, \mathcal{A}, \mathbb{P})$. Furthermore,

$$b(T(x, a, f, Y) = \frac{z}{(1-\beta)^2} + \frac{1}{1-\beta}\Big(x + z - a - f(Y) - \pi_R(f)\Big)^+$$

$$\leq \frac{z}{(1-\beta)^2} + \frac{1}{1-\beta}\,(x+z)^+$$

implying that $\vartheta_2(x) = \frac{z}{(1-\beta)^2} + \frac{1}{1-\beta}\,(x+z)^+$, which is continuous, and $\Theta_2(Y) = 0 \in L^p(\Omega, \mathcal{A}, \mathbb{P})$.

(vi) The probability measure $\mathbb{P}_1$ on $(\Omega_1, \mathcal{A}_1)$ can w.l.o.g. assumed to be separable since $\mathcal{B}(\mathbb{R}_+)$ is countably generated (apply Lemma B.5 and a canonical construction).

(vii) We assume that the ambiguity set $\mathcal{Q}$ is norm bounded, i.e. there exists $K \in [1, \infty)$ such that

$$\mathbb{E}\left|\frac{\mathrm{d}\,\mathbb{Q}}{\mathrm{d}\,\mathbb{P}_1}\right|^q \leq K$$

for all $\mathbb{Q} \in \mathcal{Q}$.

(viii) The discount factor β satisfies $\alpha\beta = 1 - (1-\beta)^2 < 1$ for all $\beta \in (0,1)$.

Hence, Corollary 4.22 implies that it is sufficient for the insurer to minimize over all Markov policies, the value functions lie in

$$\mathbb{B} = \{v : \mathbb{R} \to \mathbb{R} :\ v \text{ lower semicontinuous and decreasing}\}$$

and satisfy the Bellman equation

$$\begin{aligned} J_N(x) &= -x^+, \\ J_n(x) &= \inf_{(a,f)\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} -a + \beta\mathbb{E}^{\mathbb{Q}}\left[J_{n+1}(x+z-a-f(Y)-\pi_R(f))\right], \qquad x \in \mathbb{R}, \end{aligned}$$

for $n = 0, \ldots, N-1$. There exists a Markov Decision rule $d_n^* : \mathbb{R} \to \mathbb{R}_+ \times \mathcal{F}$ minimizing J_{n+1} and every sequence $\pi = (d_0^*, \ldots, d_{N-1}^*) \in \Pi^M$ of such minimizers is a solution to (4.34).

Under an infinite planning horizon there is no terminal cost and it suffices to consider Markov policies due to the respective results in the finite horizon case. The optimization objective is

$$J_\infty(x) = \inf_{\pi\in\Pi^M} \sup_{\gamma\in\Gamma^M} \mathbb{E}^{\pi\gamma}\left[-\sum_{k=0}^{\infty} \beta^k A_k\right].$$

Corollary 4.22 states that the Bellman operator

$$\mathcal{T} : \mathbb{B} \to \mathbb{B},\ \mathcal{T}v(x) = \inf_{(a,f)\in D(x)} \sup_{\mathbb{Q}\in\mathcal{Q}} -a + \beta\mathbb{E}^{\mathbb{Q}}\left[J_{n+1}(x+z-a-f(Y)-\pi_R(f))\right]$$

is a contraction with modulus $1-(1-\beta)^2$ and J_∞ is its unique fixed point. Every stationary policy $\pi = (d^*, d^*, \ldots) \in \Pi^S$ induced by a minimizer d^* of J_∞ is optimal for the insurer under an infinite planning horizon.

4.4.2. Robust Linear-Quadratic Programming

The term linear-quadratic (LQ) problem refers to Markov decision problems with linear transition function and quadratic one-stage cost function. Such models occur i.a. in automatic control of motions, where one wants to keep the object close to the origin. The unambiguous stochastic LQ problem has been studied extensively in the literature. For a particularly detailed account see Bertsekas (2017, 2012). The popularity is due to the nice feature that the value functions retain the quadratic structure of the one-stage cost functions and the optimal decision rules are linear and can be determined analytically.

The state and action space are $E = \mathbb{R}^m$ and $A = \mathbb{R}^d$. There is no constraint. Let $U_1, \dots, U_n$, $V_1, \dots, V_N$ be $\mathbb{R}^{m\times m}$- and $\mathbb{R}^{m\times d}$-valued random matrices, respectively, and $W_1, \dots, W_N$ be random vectors with values in $\mathbb{R}^m$. The random elements $\{Z_n = (U_n, V_n, W_n)\}_{1\le n\le N}$ are independent and the n-th element is defined on $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$. It is supposed that the disturbances $\{Z_n\}_{1\le n\le N}$ have finite $2p$-th moments, $p \ge 1$. The transition function is given by

$$T_n(x, a, Z_{n+1}) = U_{n+1}x + V_{n+1}a + W_{n+1}$$

for $n = 0, \dots, N-1$. Furthermore, let there be deterministic positive semidefinite symmetric matrices $Q_0, \dots, Q_N \in \mathbb{R}^{m\times m}$ and deterministic positive definite symmetric matrices $R_0, \dots, R_{N-1} \in \mathbb{R}^{d\times d}$. The one-stage cost functions are

$$c_n(x, a, x') = x^\top Q_n x + a^\top R_n a$$

and the terminal cost function is $c_N(x) = x^\top Q_N x$. Hence, the optimization problem under consideration is

$$\inf_{\pi\in\Pi^R} \sup_{\gamma\in\Gamma} \mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{N-1} X_k^\top Q_k X_k + A_k^\top R_k A_k + X_N^\top Q_N X_N\right]. \tag{4.35}$$

Policy values and value functions are defined in the usual way.

Since the matrices Q_n and R_n are positive semidefinite, $\underline{b} \equiv 0$ is a lower bounding function and the one-stage costs are at least quasi-integrable. In the sequel, we will determine the value functions and optimal policy by elementary calculations and will show that the value functions are convex and therefore continuous. Hence, we can dispense with an upper bounding function and compactness of the action space.

In contrast to the risk-neutral case, the quadratic structure of the one-stage cost functions is in general not inherited by the value functions under the robust optimality criterion. Then, explicit solutions can no longer be expected. Therefore, we will study special cases of ambiguity where the LQ structure is preserved. Since the Borel σ-algebra of a finite dimensional euclidean space is countably generated, it is no restriction to assume that the probability measures $\mathbb{P}_n$ are separable (canonical construction, Lemma B.5). Further, we assume that for $n = 0, \dots, N-1$

- the ambiguity sets $\mathcal{Q}_{n+1} \subseteq \mathcal{M}_1^q(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$ are norm bounded and weak* closed.
- it holds $\mathbb{E}^{\mathbb{Q}}[W_{n+1}] = 0$ for all $\mathbb{Q} \in \mathcal{Q}_{n+1}$.

I.e. Assumption 4.3 is satisfied apart from upper bounding. Theorems 4.8 and 4.12 use the bounding, continuity and compactness assumptions only to prove the existence of optimal decision rules. Thus, we can employ the Bellman equation and restrict the consideration to Markov policies as long as we are able to prove the existence of optimal decision rules on each stage. We proceed backwards.

At stage N, no action has to be chosen and the value function is $J_N(x) = x^\top Q_N x$.

At stage $N-1$, we have to solve the Bellman equation

$$\begin{aligned} J_{N-1}(x) &= \inf_{a\in A} \sup_{\mathbb{Q}\in\mathcal{Q}_N} c(x,a) + \mathbb{E}^{\mathbb{Q}}\left[J_N(T(x,a,Z_{n+1}))\right] \\ &= \inf_{a\in A} \sup_{\mathbb{Q}\in\mathcal{Q}_N} x^\top Q_{N-1}x + a^\top R_{N-1}a \\ &\qquad\qquad + \mathbb{E}^{\mathbb{Q}}\left[(U_N x + V_N a + W_N)^\top Q_N (U_N x + V_N a + W_N)\right] \\ &= \inf_{a\in A} \sup_{\mathbb{Q}\in\mathcal{Q}_N} x^\top Q_{N-1}x + a^\top R_{N-1}a + \mathbb{E}^{\mathbb{Q}}\Big[x^\top U_N^\top Q_N U_N x + a^\top V_N^\top Q_N V_N a \\ &\qquad\qquad +2x^\top U_N^\top Q_N V_N a + W_N^\top Q_N W_N\Big] \end{aligned} \tag{4.36}$$

For the last equality we used the symmetry of Q_N and that $\mathbb{E}^{\mathbb{Q}}\left[2W_N^\top Q_N(U_N x + V_N a)\right] = 0$ by assumption. Since R_{N-1} and Q_N are positive (semi-)definite, the objective function (4.36) is strictly convex in a. Moreover, it is linear and especially concave in $\mathbb{Q}$. Finally, $\mathcal{Q}_N$ is weak* compact by the Theorem of Banach-Alaoglu (Aliprantis and Border; 2006, 6.21). The objective function (4.36) is continuous in $\mathbb{Q}$ by definition of the weak* topology since the integrand is in $L^p(\Omega_{n+1}, \mathcal{A}_{n+1}, \mathbb{P}_{n+1})$. Thus, the requirements of Sion's Minimax Theorem A.27 b) are satisfied and we can interchange infimum and supremum in (4.36), i.e.

$$\begin{aligned} J_{N-1}(x) &= \sup_{\mathbb{Q}\in\mathcal{Q}_N} \inf_{a\in A} x^\top Q_{N-1}x + a^\top R_{N-1}a + \mathbb{E}^{\mathbb{Q}}\Big[x^\top U_N^\top Q_N U_N x + a^\top V_N^\top Q_N V_N a \\ &\qquad\qquad +2x^\top U_N^\top Q_N V_N a + W_N^\top Q_N W_N\Big] \\ &= \sup_{\mathbb{Q}\in\mathcal{Q}_N} \inf_{a\in A} x^\top Q_{N-1}x + a^\top R_{N-1}a + x^\top \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N U_N]x + a^\top \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]a \\ &\qquad\qquad + 2x^\top \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N V_N]a + \mathbb{E}^{\mathbb{Q}}[W_N^\top Q_N W_N] \end{aligned} \tag{4.37}$$

In oder to solve the inner minimization problem it suffices due to strict convexity and smoothness to determine the unique zero of the gradient of the objective function.

$$\begin{aligned} & 0 = 2R_{N-1}a + 2\mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]a + 2x^\top \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N] \\ \iff\quad & a = -\left(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]\right)^{-1} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N]x. \end{aligned}$$

Note that the matrix $(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N])$ is positive definite and therefore invertible due to the positive (semi-)definiteness of R_{N-1} and Q_N. Setting

$$L_{N-1}^{\mathbb{Q}} = -\Big(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]\Big)^{-1} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N]$$

and inserting in (4.37) gives

$$\begin{aligned}
J_{N-1}(x) &= \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}^{\mathbb{Q}}[W_N^\top Q_N W_N] + x^\top \bigg(Q_{N-1} + \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N U_N] + L_{N-1}^{\mathbb{Q}\top} R_{N-1} L_{N-1}^{\mathbb{Q}} \\
&\qquad + L_{N-1}^{\mathbb{Q}\top} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N] L_{N-1}^{\mathbb{Q}} + 2\mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N V_N] L_{N-1}^{\mathbb{Q}} \bigg) x \\
&= \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}^{\mathbb{Q}}[W_N^\top Q_N W_N] + x^\top \bigg(Q_{N-1} + \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N U_N] \\
&\qquad + \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N V_N] \Big(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]\Big)^{-1} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N] \\
&\qquad - 2\mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N V_N] \Big(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]\Big)^{-1} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N] \bigg) x \\
&= \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}^{\mathbb{Q}}[W_N^\top Q_N W_N] + x^\top \bigg(Q_{N-1} + \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N U_N] \\
&\qquad - \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N V_N] \Big(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]\Big)^{-1} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N] \bigg) x \\
&= \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}^{\mathbb{Q}}[W_N^\top Q_N W_N] + x^\top K_{N-1}^{\mathbb{Q}} x.
\end{aligned} \tag{4.38}$$

The matrix

$$\begin{aligned}
K_{N-1}^{\mathbb{Q}} = Q_{N-1} &+ \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N U_N] \\
&- \mathbb{E}^{\mathbb{Q}}[U_N^\top Q_N V_N] \Big(R_{N-1} + \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N V_N]\Big)^{-1} \mathbb{E}^{\mathbb{Q}}[V_N^\top Q_N U_N]
\end{aligned}$$

is obviously symmetric. Since Q_{N-1} and R_{N-1} are positive semidefinite, the second line of (4.36) shows that J_{N-1} is non-negative. Hence, $K_{N-1}^{\mathbb{Q}}$ is positive semidefinite, i.e. it has the same properties as Q_N. So if there is no ambiguity ($|\mathcal{Q}_N| = 1$), we have solved the stochastic LQ problem at stage $N-1$. The previous stage $N-2$ is analogous, one just has to replace Q_N by K_{N-1} and so on.

The ambiguous case is more intricate. As the supremum of a family of convex functions is convex, we directly have convexity of J_{N-1}. We have noted before that the ambiguity set $\mathcal{Q}_N$ is weak* compact and that the objective function (4.36) is weak* continuous in $\mathbb{Q}$. Since the infimum of a family of continuous functions is at least upper semicontinuous (Corollary A.3 is applicable as the weak* topology is metrizable due to Lemma 4.1), we can guarantee the existence of a maximizing probability measure with Weierstraß' Theorem A.7. But in contrast to the unambiguous case, the quadratic structure of the value function

J_{N-1} will be lost in general. So our technique will fail on any previous stage and there is no reason to expect explicit solutions.

In the sequel, we will therefore present two special cases where the quadratic structure of the value function is preserved under ambiguity. This can only be expected if the maximizing probability measure in (4.38) does not depend on x, implying that after a stage-wise change of measure the problem can be reduced to the unambiguous case.

Case 1. U_N, V_N are deterministic.
Then $K_{N-1} = K^{\mathbb{Q}}_{N-1}$ and the optimal decision rule of the controller $d^*_{N-1}(x) = L_{N-1}x = L^{\mathbb{Q}}_{N-1}x$ do not depend on the probability measure and $\mathbb{Q}^*_N = \operatorname{argmax}_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}^{\mathbb{Q}}[W_N^\top Q_N W_N]$ is an optimal action of nature independently of x. The value function is therefore given by

$$J_{N-1}(x) = \mathbb{E}^{\mathbb{Q}^*}[W_N^\top Q_N W_N] + x^\top K_{N-1}x.$$

Previous stages are analogous, one just has to replace Q_N by K_{N-1} and so on.

Case 2. W_N is deterministic and $m = 1$.
That is, $W_N = 0$ and the state space is the real line but the action space remains arbitrary. Consequently, $K^{\mathbb{Q}}_{N-1}$ is a non-negative real number,

$$J_{N-1}(x) = x^2 \sup_{\mathbb{Q}\in\mathcal{Q}_N} K^{\mathbb{Q}}_{N-1}$$

and the optimal decision rule is $d^*_{N-1}(x) = L^{\mathbb{Q}^*_N}_{N-1}x$ with $\mathbb{Q}^*_N = \operatorname{argmax}_{\mathbb{Q}\in\mathcal{Q}_N} K^{\mathbb{Q}}_{N-1}$. Earlier stages are again analogous.

In most cases, the quadratic structure of the value function will not be preserved under ambiguity. If the ambiguity set is the convex hull of two probability measures, we can at least give an optimal quadratic upper bound for the value function. First, note that by Lemma 4.28 and subsequent remarks, the convex hull has no impact and we can focus on an ambiguity set with two elements $\mathcal{Q}_N = \{\mathbb{Q}_1, \mathbb{Q}_2\}$. We write shorthand $K^i_{N-1} = K^{\mathbb{Q}_i}_{N-1},\ i = 1, 2.$

As a real symmetric matrix, $K^1_{N-1} - K^2_{N-1}$ is orthogonally diagonalizable, i.e there is an orthogonal matrix $P \in \mathbb{R}^{m\times m}$ such that $P^\top(K^1_{N-1} - K^2_{N-1})P = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$. We define $|K^1_{N-1} - K^2_{N-1}| = P \operatorname{diag}(|\lambda_1|, \dots, |\lambda_m|)P^\top$ and

$$\widehat{K}_{N-1} = \max\{K^1_{N-1},\ K^2_{N-1}\} = \frac{1}{2}\Big(K^1_{N-1} + K^2_{N-1} + |K^1_{N-1} - K^2_{N-1}|\Big).$$

It holds for all $x \in \mathbb{R}^m$

$$\begin{aligned} x^\top \widehat{K}_{N-1}x - x^\top K^1_{N-1}x &= \frac{1}{2}x^\top\Big(K^2_{N-1} - K^1_{N-1} + |K^1_{N-1} - K^2_{N-1}|\Big)x \\ &= \frac{1}{2}x^\top P\Big(\operatorname{diag}(|\lambda_1|, \dots, |\lambda_m|) - \operatorname{diag}(\lambda_1, \dots, \lambda_m)\Big)P^\top x \\ &\geq 0 \end{aligned}$$

and analogously $x^\top \widehat{K}_{N-1} x - x^\top K_{N-1}^2 x \geq 0$. The upper bound is optimal in the sense that if $x^\top K_{N-1}^i x \geq x^\top K_{N-1}^j x$ for all $x \in \mathbb{R}^m$, i.e. if $K_{N-1}^i \succeq K_{N-1}^j$ in the Loewner order, then it holds $\widehat{K}_{N-1} = K_{N-1}^i$.

Note that $\widehat{K}_{N-1}$ is symmetric and positive semidefinite. So we can continue with

$$\widehat{J}_{N-1}(x) = \max\left\{\mathbb{E}^{\mathbb{Q}_1}[W_C^\top Q_N W_N], \mathbb{E}^{\mathbb{Q}_2}[W_N^\top Q_N W_N]\right\} + x^\top \widehat{K}_{N-1} x$$

instead of J_{N-1} on the previous stage and recursively obtain an optimal upper bound for the value function. In principle, this procedure works for any ambiguity set which consists of (the convex hull of) a finite number of probability measures. But the upper bound will in general depend on the order in which the pairwise maxima of matrices are taken and might therefore be not optimal.

Remark 4.33. The (robust) LQ problem defines a convex model in the sense of Lemma 4.23 despite having no monotonicity properties. This is possible due to the linear transition function.

CHAPTER 5

Risk-Sensitive Recursive Cost Minimization

A shortcoming of the expected total cost criterion commonly used for MDP is that it cannot take into account risk aversion of the controller. A natural generalization is therefore to replace the expectation by some risk measure. Usually, static risk measures do not have a tower property like conditional expectation. However, this property is crucial to derive a value iteration and solve the optimization problem by means of dynamic programming.

Dynamic risk measures (see Chapter 11 of Föllmer and Schied (2016) for details) often have such a property but typically they rely on conditioning with respect to some filtration. As noted in Ruszczyński (2010), this implies that at each time step the value function may depend on the entire history of the process. Hence, one cannot expect to obtain Markov optimal policies making the problem computationally intractable. Ruszczyński (2010) avoids that by constructing so-called *Markov risk measures*. However, this approach is rather technical, requires coherence, and the dynamic risk measures are only obtained in a dual representation.

A different approach is taken by Bäuerle and Jaśkiewicz (2017, 2018) and Asienkiewicz and Jaśkiewicz (2017). They do not define the optimality criterion based on the total cost but start with the value iteration

$$\begin{aligned} V_{n\pi}(h_n) &= \mathbb{E}_{nh_n}\Big[c_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) + V_{n+1\pi}(H_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi)\Big] \\ &= \mathbb{E}\Big[c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \\ &\qquad + V_{n+1\pi}(h_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1}))\Big] \end{aligned}$$

of the expected total cost criterion and replace the factorization of the conditional expectation by some risk measure $\rho_n : L^p(\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{R}}$. The value iteration then reads

$$\begin{aligned} V_{n\pi}(h_n) = \rho_n\Big(& c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \\ & + V_{n+1\pi}(h_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1}))\Big) \end{aligned}$$

and dynamic programming techniques can be applied by construction. It can be ensured by suitable model components that the value function at each time step does not depend on the history of the process and hence Markov optimal policies can be obtained. The advantage of the *recursive approach* is one can use general static risk measures at each stage like the well-established Value-at-Risk or Expected Shortfall.

Both Bäuerle and Jaśkiewicz (2017, 2018) and Asienkiewicz and Jaśkiewicz (2017) considered specifically the entropic risk measure. This choice originates from the fact that the entropic risk measure is the certainty equivalent of an exponential utility. In the economic literature, the representation of preferences by recursive utility functions has been widely studied with notable contributions by Kreps and Porteus (1978) and Epstein and Zin (1989). A comprehensive presentation can be found in Miao (2014, Ch. 20). The key feature of recursive utilities is that they allow separating intertemporal preferences from risk aversion. An early application to optimal control is the paper of Hansen and Sargent (1995). We extend their results to general law-invariant monetary risk measures with the Fatou property.

Locally at each stage, the recursive approach provides an intuitive and transparent decision criterion. However globally, there is no closed-form expression for the objective function and no obvious interpretation. In many cases that shortcoming can be overcome by reformulating the optimization problem to a distributionally robust MDP, see Section 5.3.

5.1. Finite Planning Horizon

Under a finite planning horizon $N \in \mathbb{N}$, we consider the non-stationary version of the abstract cost model of Section 3.1. In this chapter only deterministic policies $\pi \in \Pi$ of the controller will be considered. The Markov Decision Process therefore has the functional representation (3.3). Here, it is more convenient to index the process and its random history with the policy since we will not explicitly refer to the law of motion. Let $p \in [1, \infty)$ with conjugate index $q \in [1, \infty]$ and let $\rho_0, \dots, \rho_{N-1} : L^p(\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{R}}$ be monetary risk measures. We define the value of a policy $\pi = (d_0, \dots, d_{N-1}) \in \Pi$ at time $n = 0, \dots, N$ given history $h_n \in \mathcal{H}_n$ recursively as

$$\begin{aligned} V_{N\pi}(h_N) &= c_N(x_N) \\ V_{n\pi}(h_n) &= \rho_n\Big(c_n(x_n, d_n(h_n), X^\pi_{n+1}) + V_{n+1\pi}(h_n, d_n(h_n), X^\pi_{n+1})\Big) \qquad (5.1) \\ &= \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \end{aligned}$$

$$+ V_{n+1\pi}\big(h_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})\big)\Big).$$

Slightly abusing the notation, we write X^π_{n+1} in (5.1) instead of the random variable $X^\pi_{n+1}|H^\pi_n = h_n$ with the conditional distribution given the history up to time n. In the special case that the one-stage cost functions c_n do not depend on the next state of the process, the value of a policy simplifies to

$$V_{n\pi}(h_n) = c_n(x_n, d_n(h_n)) + \rho_n\big(V_{n+1\pi}(h_n, d_n(h_n), X^\pi_{n+1})\big), \qquad h_n \in \mathcal{H}_n,$$

for $n = 0, \dots, N-1$ due to the translation invariance of monetary risk measures.

Remark 5.1. For the recursive definition of the policy values to be meaningful, we need to make sure that the risk measures are applied to elements of $L^p(\Omega, \mathcal{A}, \mathbb{P})$. This has two aspects: integrability will be ensured by Assumption 5.2, but first of all $V_{n\pi}$ needs to be a measurable function for all $\pi \in \Pi$ and $n = 0, \dots, N$. For most risk measures with practical relevance, this is fulfilled:

- In the risk-neutral case, i.e. for $\rho = \mathbb{E}$, and also for the entropic risk measure ρ_γ the measurability is obvious.
- For distortion risk measures, the measurability is guaranteed, too. To see this, we proceed backwards. For N there is noting to show and if $V_{n+1\pi}$ is measurable, the function

  $$f(h_n, z) = c_n\big(x_n, d_n(h_n), T_n(x_n, d_n(h_n), z)\big) + V_{n+1\pi}\big(h_n, d_n(h_n), T_n(x_n, d_n(h_n), z\big)$$

  is measurable as a composition of measurable maps. Then, Fubini's Theorem B.2 yields that the survival function of $f(h_n, Z_{n+1})$

  $$S(t|h_n) = \int \mathbb{1}\{f(h_n, Z_{n+1}(\omega)) > t\}\mathbb{P}(\mathrm{d}\,\omega)$$

  is measurable. A distortion function g is increasing and hence measurable. So again by Fubini we obtain the measurability of

  $$V_{n\pi}(h_n) = \rho_g(f(h_n, Z_{n+1})) = \int_0^\infty g(S(t|h_n))\,\mathrm{d}\,t - \int_{-\infty}^0 1 - g(S(t|h_n))\,\mathrm{d}\,t$$

  since the integrands are non-negative and compositions of measurable maps.
- For proper coherent risk measures with the Fatou property one can insert the dual representation of Proposition 2.21. Then, an optimal measurable selection argument as in Theorem 4.8 yields the measurability.

Throughout, it is implicitly assumed that the risk measures are chosen such that all policy values are measurable.

The value functions are given by

$$V_n(h_n) = \inf_{\pi \in \Pi} V_{n\pi}(h_n), \qquad h_n \in \mathcal{H}_n,$$

for $n = 0, \dots, N$ and the controller's optimization objective is

$$V_0(x) = \inf_{\pi \in \Pi} V_{0\pi}(x), \qquad x \in E. \tag{5.2}$$

In order to have well-defined value functions, we need some finiteness conditions as well as some technical conditions for measurability and optimization. All assumptions of this section are listed here.

Assumption 5.2. (i) The model data has the Continuity and Compactness Properties 3.1 with the transition function T_n being continuous in (x, a) for all $n = 0, \dots, N-1$ (case 1).

(ii) There exist $\underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$ and measurable functions $\underline{\mathbf{b}} : E \to (-\infty, -\underline{\epsilon}]$ and $\bar{\mathbf{b}} : E \to [\bar{\epsilon}, \infty)$ such that it holds for all policies $\pi \in \Pi$ and all $n = 0, \dots, N$

$$\underline{\mathbf{b}}(x_n) \leq V_{n\pi}(h_n) \leq \bar{\mathbf{b}}(x_n), \qquad h_n \in \mathcal{H}_n.$$

(iii) We define $\mathbf{b} : E \to [1, \infty)$, $\mathbf{b}(x) = \bar{\mathbf{b}}(x) - \underline{\mathbf{b}}(x)$. For all $n = 0, \dots, N-1$ and $(\bar{x}, \bar{a}) \in D_n$ there exists an $\epsilon > 0$ and measurable functions $\Theta_{n,1}^{\bar{x},\bar{a}}, \Theta_{n,2}^{\bar{x},\bar{a}} : \mathcal{Z} \to \mathbb{R}_+$ such that $\Theta_{n,1}^{\bar{x},\bar{a}}(Z_{n+1}), \Theta_{n,2}^{\bar{x},\bar{a}}(Z_{n+1}) \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and

$$|c_n(x, a, T_n(x, a, z))| \leq \Theta_{n,1}^{\bar{x},\bar{a}}(z), \qquad\qquad \mathbf{b}(T_n(x, a, z)) \leq \Theta_{n,2}^{\bar{x},\bar{a}}(z)$$

for all $z \in \mathcal{Z}$ and $(x, a) \in B_\epsilon(\bar{x}, \bar{a}) \cap D_n$. Here, $B_\epsilon(\bar{x}, \bar{a})$ is the closed ball around $(\bar{x}, \bar{a})$ w.r.t. an arbitrary product metric on $E \times A$.

(iv) The monetary risk measures $\rho_0, \dots, \rho_{N-1} : L^p(\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{R}}$ are law invariant and have the Fatou property.

Remark 5.3. $\underline{\mathbf{b}}, \bar{\mathbf{b}}$ are called *(global) lower* and *upper bounding function*, respectively, while $\mathbf{b}$ is referred to as *(global) bounding function*. Since $\underline{\mathbf{b}}$ is non-positive and $\bar{\mathbf{b}}$ is non-negative it holds

$$\underline{\mathbf{b}}(x_n) \leq -V_{n\pi}^-(h_n) \leq V_{n\pi}(h_n) \leq V_{n\pi}^+(h_n) \leq \bar{\mathbf{b}}(x_n), \qquad h_n \in \mathcal{H}_n,$$

and consequently

$$|V_{n\pi}(h_n)| \leq \mathbf{b}(x_n), \qquad h_n \in \mathcal{H}_n.$$

Bold print is used to distinguish these global bounding functions from corresponding stage-wise bounding functions as in Chapter 4. Such stage-wise bounding functions can be introduced for the risk-sensitive recursive optimality criterion, too, if the risk measures have additional properties.

Lemma 5.4. *Let $\rho_0, \dots, \rho_{N-1}$ be proper coherent risk measures with the Fatou property. If there exist $\underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$, measurable functions $\underline{b} : E \to (-\infty, -\underline{\epsilon}]$, $\bar{b} : E \to [\bar{\epsilon}, \infty)$ and a constant $\alpha \in (0,1)$ such that*

$$\rho_n\big(c_n(x,a,T_n(x,a,Z_{n+1}))\big) \geq \underline{b}(x), \qquad \rho_n\big(-\underline{b}(T_n(x,a,Z_{n+1}))\big) \leq -\alpha\underline{b}(x),$$
$$\rho_n\big(c_n(x,a,T_n(x,a,Z_{n+1}))\big) \leq \bar{b}(x), \qquad \rho_n\big(\bar{b}(T_n(x,a,Z_{n+1}))\big) \leq \alpha\bar{b}(x),$$

for all $n = 0, \dots, N-1$ and $(x,a) \in D_n$ as well as $\underline{b}(x) \leq c_N(x) \leq \bar{b}(x)$ for all $x \in E$, then

$$\underline{\mathbf{b}} = \frac{1}{1-\alpha}\underline{b}, \qquad \bar{\mathbf{b}} = \frac{1}{1-\alpha}\bar{b} \qquad \textit{and} \qquad \mathbf{b} = \frac{1}{1-\alpha}b$$

are global bounding functions satisfying Assumption 5.2 (ii).

Proof. We proceed by backward induction. At time N we have

$$\underline{\mathbf{b}}(x_N) \leq \underline{b}(x_N) \leq c_N(x_N) = V_{N\pi}(h_N) \leq \bar{b}(x_N) \leq \bar{\mathbf{b}}(x_N), \qquad h_N \in \mathcal{H}_N.$$

Assuming the assertion holds for time $n+1$ it follows for time n:

$$\begin{aligned}
V_{n\pi}(h_n) &= \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \\
&\qquad\qquad + V_{n+1\pi}(h_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1}))\Big) \\
&\geq \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \\
&\qquad\qquad + \frac{1}{1-\alpha}\underline{b}(T_n(x_n, d_n(h_n), Z_{n+1}))\Big) \\
&\geq \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1}))\Big) \\
&\qquad - \frac{1}{1-\alpha}\rho_n\Big(-\underline{b}(T_n(x_n, d_n(h_n), Z_{n+1}))\Big) \\
&\geq \underline{b}(x_n) + \frac{\alpha}{1-\alpha}\underline{b}(x_n) \\
&= \underline{\mathbf{b}}(x_n).
\end{aligned}$$

The first inequality is by the induction hypothesis and the monotonicity of ρ_n and the second one is by Lemma 2.23. Additionally, we have used positive homogeneity. Finally, the third inequality is by assumption. Regarding the upper bounding function one can argue similarly using the subadditivity of ρ_n instead of Lemma 2.23:

$$\begin{aligned}
V_{n\pi}(h_n) &= \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \\
&\qquad\qquad + V_{n+1\pi}(h_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1}))\Big) \\
&\leq \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) \\
&\qquad\qquad + \frac{1}{1-\alpha}\bar{b}(T_n(x_n, d_n(h_n), Z_{n+1}))\Big)
\end{aligned}$$

$$\begin{aligned}&\le \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1}))\Big)\\&\quad + \frac{1}{1-\alpha}\rho_n\Big(\bar{b}(T_n(x_n, d_n(h_n), Z_{n+1}))\Big)\\&\le \bar{b}(x_n) + \frac{\alpha}{1-\alpha}\bar{b}(x_n)\\&= \bar{\mathbf{b}}(x_n).\end{aligned}$$ □

Remark 5.5. Regarding the requirements on a stage-wise lower bounding function in Lemma 5.4 it should be noted that $\rho_n\big(-\underline{b}(T_n(x,a,Z_{n+1}))\big) \le -\alpha\underline{b}(x)$ is a stronger assumption than

$$\rho_n\big(\underline{b}(T_n(x,a,Z_{n+1}))\big) \ge \alpha\underline{b}(x). \tag{5.3}$$

Indeed, since $\underline{b} \le 0$ the monotonicity and normalization of ρ_n yields $\rho_n\big(\underline{b}(T_n(x,a,Z_{n+1}))\big) \le 0$. Consequently, we have

$$\begin{aligned}-\rho_n\big(\underline{b}(T_n(x,a,Z_{n+1}))\big) &= \big|\rho_n\big(\underline{b}(T_n(x,a,Z_{n+1}))\big)\big| \le \rho_n\Big(\big|\underline{b}(T_n(x,a,Z_{n+1}))\big|\Big)\\&= \rho_n\Big(-\underline{b}(T_n(x,a,Z_{n+1}))\Big) \le -\alpha\underline{b}(x).\end{aligned}$$

The first inequality is Lemma 2.2 and the second one by assumption. Multiplying with (-1) yields (5.3).

If the one-stage cost functions are bounded and the monetary risk measures $\rho_0, \dots, \rho_{N-1}$ normalized, the stage-wise bounding functions $\underline{b}, \bar{b}$ can be chosen constant. Where we have used Lemma 2.23 or subadditivity in the proof of Lemma 5.4, one can then simply argue with translation invariance. Recall that normalization is no structural restriction for monetary risk measures due to the translation invariance.

With the bounding function $\mathbf{b}$ we define the function space

$$\mathbb{B}_{\mathbf{b}} = \{v : E \to \mathbb{R} \mid v \text{ measurable with } \lambda \in \mathbb{R}_+ \text{ s.t. } |v(x)| \le \lambda\,\mathbf{b}(x) \text{ for all } x \in E\}$$

as in Section 4.1. Endowing $\mathbb{B}_{\mathbf{b}}$ with the weighted supremum norm

$$\|v\|_{\mathbf{b}} = \sup_{x\in E} \frac{|v(x)|}{\mathbf{b}(x)}$$

makes $(\mathbb{B}_{\mathbf{b}}, \|\cdot\|_{\mathbf{b}})$ a Banach space, cf. Proposition 7.2.1 in Hernández-Lerma and Lasserre (1999). In case we have stage-wise bounding functions as in Lemma 5.4, it holds

$$\begin{aligned}\mathbb{B}_{\mathbf{b}} &= \{v : E \to \mathbb{R} \mid v \text{ measurable with } \lambda \in \mathbb{R}_+ \text{ s.t. } |v(x)| \le \lambda\,\mathbf{b}(x) \text{ for all } x \in E\}\\&= \{v : E \to \mathbb{R} \mid v \text{ measurable with } \lambda \in \mathbb{R}_+ \text{ s.t. } |v(x)| \le \lambda\,b(x) \text{ for all } x \in E\}\\&= \mathbb{B}_b\end{aligned}$$

and the weighted supremum norms $\|\cdot\|_{\mathbf{b}}$, $\|\cdot\|_b$ are equivalent.

Note that Assumption 5.2 (iii) is exactly the same as Assumption 4.3 (iii). It does not depend on where the bounding function $\mathbf{b}$ originates from and is in this sense independent of the optimality criterion. Thus, with the same arguments one can show that the statement of Lemma 4.7 holds here, too.

Lemma 5.6. *Let $v \in \mathbb{B}_{\mathbf{b}}$ and $n \in \{0, \dots, N-1\}$. Under Assumption 5.2 (iii) each sequence of random variables*

$$C_k = c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})) + v(T_n(x_k, a_k, Z_{n+1}))$$

induced by a convergent sequence $\{(x_k, a_k)\}_{k \in \mathbb{N}}$ in D_n has an L^p-bound $\bar{C}$, i.e. $|C_k| \leq \bar{C} \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ for all $k \in \mathbb{N}$.

Let us now consider specifically Markov policies $\pi \in \Pi^M$ of the controller. The subspace

$$\mathbb{B} = \{v \in \mathbb{B}_{\mathbf{b}} : \ v \text{ lower semicontinuous}\}$$

of $(\mathbb{B}_{\mathbf{b}}, \|\cdot\|_{\mathbf{b}})$ turns out to be the set of potential value functions under such policies. $(\mathbb{B}, \|\cdot\|_{\mathbf{b}})$ is a complete metric space since the subset of lower semicontinuous functions is closed in $(\mathbb{B}_{\mathbf{b}}, \|\cdot\|_{\mathbf{b}})$ by Lemma A.10. When we consider intervals $[\underline{v}, \bar{v}] \subseteq \mathbb{B}$ with $\underline{v}, \bar{v} : E \to \mathbb{R}$ s.t. $\underline{v}(x) \leq \bar{v}(x)$ for all $x \in E$, they are to be understood pointwise

$$[\underline{v}, \bar{v}] = \{v \in \mathbb{B} : \underline{v}(x) \leq v(x) \leq \bar{v}(x) \text{ for all } x \in E\}.$$

Note that $\underline{v}, \bar{v}$ need not be in $\mathbb{B}$. Such intervals are closed even w.r.t. pointwise convergence and therefore form a complete metric space as a closed subset of $(\mathbb{B}, \|\cdot\|_{\mathbf{b}})$. In the sequel, the interval

$$I = \left[\underline{\mathbf{b}}, \bar{\mathbf{b}}\right]$$

will be of interest. We define the following operators on $\mathbb{B}_{\mathbf{b}}$ and especially on $\mathbb{B}$.

Definition 5.7. For $v \in \mathbb{B}_b$ and a Markov decision rule d let

$$\begin{aligned} L_n v(x,a) &= \rho_n\Big(c_n(x,a,T_n(x,a,Z_{n+1})) + v(T_n(x,a,Z_{n+1}))\Big), && (x,a) \in D_n, \\ \mathcal{T}_{nd} v(x) &= L_n v(x, d(x)), && x \in E, \\ \mathcal{T}_n v(x) &= \inf_{a \in D_n(x)} L_n v(x,a), && x \in E. \end{aligned}$$

Note that the operators are monotone in v. Under a Markov policy $\pi = (d_0, \dots, d_{N-1}) \in \Pi^M$, the value iteration can be expressed with the operators. In order to distinguish from the history-dependent case, we denote policy values here with J. Setting $J_{N\pi}(x) = c_N(x),\ x \in E$, we obtain for $n = 0, \dots, N-1$ and $x \in E$

$$\begin{aligned} J_{n\pi}(x) &= \rho_n\Big(c_n(x, d_n(x), T_n(x, d_n(x), Z_{n+1})) + J_{n+1\pi}(T_n(x, d_n(x), Z_{n+1}))\Big) \\ &= \mathcal{T}_{nd_n} J_{n+1\pi}(x). \end{aligned}$$

Let us further define for $n = 0, \ldots, N-1$ the Markov value function

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

The next result shows that V_n satisfies a Bellman equation and proves that an optimal policy exists and is Markov.

Theorem 5.8. *Let Assumption 5.2 be satisfied. Then, for $n = 0, \ldots, N$, the value function V_n only depends on x_n, i.e. $V_n(h_n) = J_n(x_n)$ for all $h_n \in \mathcal{H}_n$, lies in $I = \left[\underline{\mathbf{b}}, \bar{\mathbf{b}}\right] \subseteq \mathbb{B}$ and satisfies the Bellman equation*

$$\begin{aligned} J_N(x) &= c_N(x), \\ J_n(x) &= \mathcal{T}_n J_{n+1}(x), \qquad x \in E. \end{aligned}$$

Furthermore, for $n = 0, \ldots, N-1$ there exist Markov decision rules d_n^* with $\mathcal{T}_{nd_n^*} J_{n+1} = \mathcal{T}_n J_{n+1}$ and every sequence of such minimizers constitutes an optimal policy $\pi = (d_0^*, \ldots, d_{N-1}^*)$.

Proof. The proof is by backward induction. At time N we have $V_N = J_N = c_N$ which is in $\mathbb{B}$ by Assumption 5.2 (ii). Assuming the assertion holds at time $n+1$, we have at time n:

$$\begin{aligned} V_n(h_n) &= \inf_{\pi \in \Pi} V_{n\pi}(h_n) \\ &= \inf_{\pi \in \Pi} \rho_n\Big(c_n(x_n, d_n(h_n), X_{n+1}^\pi) + V_{n+1\pi}(h_n, d_n(h_n), X_{n+1}^\pi)\Big) \\ &\geq \inf_{\pi \in \Pi} \rho_n\Big(c_n(x_n, d_n(h_n), X_{n+1}^\pi) + V_{n+1}(h_n, d_n(h_n), X_{n+1}^\pi)\Big) \end{aligned}$$

which equals by the induction hypothesis

$$\begin{aligned} &= \inf_{\pi \in \Pi} \rho_n\Big(c_n(x_n, d_n(h_n), X_{n+1}^\pi) + J_{n+1}(X_{n+1}^\pi)\Big) \\ &= \inf_{\pi \in \Pi} \rho_n\Big(c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})) + J_{n+1}(T_n(x_n, d_n(h_n), Z_{n+1}))\Big). \end{aligned}$$

Since the minimization does not depend on the entire policy but only on $a_n = d_n(h_n)$, this equals

$$= \inf_{a_n \in D(x_n)} \rho_n\Big(c_n(x_n, a_n, T_n(x_n, a_n, Z_{n+1})) + J_{n+1}(T_n(x_n, a_n, Z_{n+1}))\Big).$$

Here, objective and constraint depend on the history of the process only through x_n. Thus, given existence of a minimizing Markov decision rule d_n^*, one ontains the identity

$$= \mathcal{T}_{nd_n^*} J_{n+1}(x_n). \tag{5.4}$$

Again by the induction hypothesis there exists an optimal Markov policy $\pi^* \in \Pi^M$ such that

$$
\begin{aligned}
&= \mathcal{T}_{nd_n^*} J_{n+1\pi^*}(x_n)\\
&= J_{n\pi^*}(x_n)\\
&\geq J_n(x_n)\\
&\geq V_n(h_n).
\end{aligned}
$$

It remains to show the existence of a minimizing Markov decision rule d_n^* at (5.4) and that $J_n \in \mathbb{B}$. We want to apply Proposition A.25. The set-valued mapping $E \ni x \mapsto D_n(x)$ is compact-valued and upper semicontinuous. Next, we show that $D_n \ni (x,a) \mapsto L_n v(x,a)$ is lower semicontinuous for every $v \in \mathbb{B}$. Let $\{(x_k, a_k)\}_{k\in\mathbb{N}}$ be a convergent sequence in D_n with limit $(x^*, a^*) \in D_n$. By Lemma A.4 a) the function $D_n \ni (x,a) \mapsto c_n(x, a, T_n(x, a, Z_{n+1}(\omega))) + v(T_n(x, a, Z_{n+1}(\omega)))$ is lower semicontinuous for every $\omega \in \Omega$. Consequently,

$$
\begin{aligned}
&\lim_{k\to\infty} \inf_{\ell\geq k} c_n(x_\ell, a_\ell, T_n(x_\ell, a_\ell, Z_{n+1})) + v(T_n(x_\ell, a_\ell, Z_{n+1}))\\
&= \liminf_{k\to\infty} c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})) + v(T_n(x_k, a_k, Z_{n+1}))\\
&\geq c_n(x^*, a^*, T_n(x^*, a^*, Z_{n+1})) + v(T_n(x^*, a^*, Z_{n+1})). \qquad (5.5)
\end{aligned}
$$

The sequence of random variables $\{C_k\}_{k\in\mathbb{N}}$ with

$$
C_k(\omega) = \inf_{\ell\geq k} c_n(x_\ell, a_\ell, T_n(x_\ell, a_\ell, Z_{n+1}(\omega))) + v(T_n(x_\ell, a_\ell, Z_{n+1}(\omega)))
$$

is increasing for every $\omega \in \Omega$. Recall here that the ω-wise infimum of a countable number of random variables $Y_1, Y_2, \dots$ is again a random variable since

$$
\left\{\inf_{k\in\mathbb{N}} Y_k \leq y\right\} = \bigcup_{n\in\mathbb{N}} \{Y_k \leq y\}
$$

is measurable. By Lemma 5.6, there exists a nonnegative random variable $\bar{C} \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ such that $|C_k| \leq \bar{C}$ for all $k \in \mathbb{N}$. Hence, $\{C_k\}_{k\in\mathbb{N}}$ converges almost surely to some $C^* \in L^p(\Omega, \mathcal{A}, \mathbb{P})$. The Fatou property of the risk measure ρ_n implies

$$
\begin{aligned}
\liminf_{k\to\infty} L_n v(x_k, a_k) &= \liminf_{k\to\infty} \rho_n\Big(c_n(x_k, a_k, T_n(x_k, a_k, Z_{n+1})) + v(T_n(x_k, a_k, Z_{n+1}))\Big)\\
&\geq \liminf_{k\to\infty} \rho_n\Big(\inf_{\ell\geq k} c_n(x_\ell, a_\ell, T_n(x_\ell, a_\ell, Z_{n+1})) + v(T_n(x_\ell, a_\ell, Z_{n+1}))\Big)\\
&= \liminf_{k\to\infty} \rho_n(C_k)\\
&\geq \rho_n(C^*)\\
&\geq \rho_n\Big(c_n(x^*, a^*, T_n(x^*, a^*, Z_{n+1})) + v(T_n(x^*, a^*, Z_{n+1}))\Big)
\end{aligned}
$$

$$= L_n v(x^*, a^*).$$

The last inequality follows from (5.5) and the monotonicity of ρ_n. So we have shown the lower semicontinuity of $D_n \ni (x,a) \mapsto L_n v(x,a)$. Proposition A.25 yields the existence of a minimizing Markov decision rule d_n^* at (5.4) and that $J_n = TJ_{n+1}$ is lower semicontinuous. Furthermore, J_n is bounded by $\underline{\mathbf{b}}$ and $\bar{\mathbf{b}}$ according to Assumption 5.2 (ii). Thus, $J_n \in I$ and the proof is complete. □

5.2. Infinite Planning Horizon

In this section, we consider the risk-sensitive recursive cost minimization problem under an infinite planning horizon. To reiterate, this approach 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 abstract cost model with no terminal cost, i.e. D, T, ρ do not depend on n, the disturbances are identically distributed, the one-stage cost functions are of the form $c_n = \beta^n c$ with some discount factor $\beta \in (0,1]$ and $c_N \equiv 0$. Let Z be a representative of the disturbance distribution. The model with infinite planning horizon is derived as a limit of the one with finite horizon. So besides a stationary version of Assumption 5.2 we need some condition to ensure convergence of the value functions when the planning horizon tends to infinity.

For the risk measure ρ we require coherence as an additional property. Note that if ρ is finite on $L^p(\Omega, \mathcal{A}, \mathbb{P})$, the Fatou property is already implied by coherence, cf. Proposition 2.16. Within the wide class of distortion risk measures, which covers many of the risk measures with practical relevance, requiring coherence essentially means a restriction to spectral risk measures (Lemma 2.5 and Remark 2.7). For spectral risk measures, finiteness is guaranteed if the spectrum ϕ lies in L^q. We will see that in case the one-stage cost function is bounded, coherence can be dropped as a requirement on the risk measure. Then, i.a. all distortion risk measures with the Fatou property are admissible. For clarity, all assumptions of this section are summarized below.

Assumption 5.9. (i) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being continuous (case 1).

(ii) There exist $\alpha, \underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$ and measurable functions $\underline{b} : E \to (-\infty, -\underline{\epsilon}]$, $\bar{b} : E \to [\bar{\epsilon}, \infty)$ such that for all $(x,a) \in D$

$$\rho\big(c(x,a,T(x,a,Z))\big) \geq \underline{b}(x), \qquad \rho\big(-\underline{b}(T(x,a,Z))\big) \leq -\alpha \underline{b}(x),$$
$$\rho\big(c(x,a,T(x,a,Z))\big) \leq \bar{b}(x), \qquad \rho\big(\bar{b}(T(x,a,Z))\big) \leq \alpha \bar{b}(x).$$

(iii) We define $b : E \to [1, \infty)$, $b(x) = \bar{b}(x) - \underline{b}(x)$. For all $(\bar{x}, \bar{a}) \in D$ there exists an $\epsilon > 0$ and measurable functions $\Theta_1^{\bar{x},\bar{a}}, \Theta_2^{\bar{x},\bar{a}} : \mathcal{Z} \to \mathbb{R}_+$ such that $\Theta_1^{\bar{x},\bar{a}}(Z), \Theta_2^{\bar{x},\bar{a}}(Z) \in$

$L^p(\Omega, \mathcal{A}, \mathbb{P})$ and

$$|c(x,a,T(x,a,z))| \le \Theta_1^{\bar{x},\bar{a}}(z), \qquad b(T(x,a,z)) \le \Theta_2^{\bar{x},\bar{a}}(z)$$

for all $z \in \mathcal{Z}$ and $(x,a) \in B_\epsilon(\bar{x},\bar{a}) \cap D$. Here, $B_\epsilon(\bar{x},\bar{a})$ is the closed ball around $(\bar{x},\bar{a})$ w.r.t. an arbitrary product metric on $E \times A$.

(iv) The law-invariant risk measure $\rho : L^p(\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{R}}$ is proper, coherent and has the Fatou property.

(v) The discount factor β satisfies $\alpha\beta < 1$.

The second column of Assumption 5.9 (ii) is needed to ensure convergence of the value functions when the planning horizon tends to infinity. Given the need for such a condition and the coherence of ρ, it is natural to introduce stage-wise bounding functions as in Lemma 5.4. Due to discounting, the corresponding global bounding function are given by

$$\underline{\mathbf{b}} = \frac{1}{1-\alpha\beta}\underline{b}, \qquad \bar{\mathbf{b}} = \frac{1}{1-\alpha\beta}\bar{b} \qquad \text{and} \qquad \mathbf{b} = \frac{1}{1-\alpha\beta}b. \tag{5.6}$$

This can be seen as in the proof of Lemma 5.4.

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 $\pi = (d_1, d_2, \dots) \in \Pi^M$ due to Theorem 5.8. 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. It is only possible under Markov policies in a stationary model. There, the two ways of indexing are equivalent. In a non-stationary model or under a history-depended policy in a stationary model the distance-to-horizon indexing is not possible and a change of notation is therefore inevitable. The value of a policy $\pi = (d_0, d_1 \dots) \in \Pi^M$ up to a planning horizon $N \in \mathbb{N}$ now is

$$J_{N\pi}(x) = \mathcal{T}_{d_0} \circ \dots \circ \mathcal{T}_{d_{N-1}} 0(x), \qquad x \in E. \tag{5.7}$$

Note that Remark 4.14 applies here, too. In a non-stationary formulation the discounting is included in the one-stage cost functions and therefore calibrated w.r.t. the fixed reference time zero. Hence, it holds

$$J_n^{\text{non-stat}}(x) = \beta^n J_{N-n}^{\text{stat}}(x), \qquad x \in E,\ n = 0, \dots, N.$$

The reformulation (5.7) makes it necessary to write the value iteration in terms of the *shifted policy* $\vec{\pi} = (d_1, d_2, \dots)$ corresponding to $\pi = (d_0, d_1, \dots) \in \Pi^M$:

$$\begin{aligned} J_{N\pi}(x) &= \rho\Big(c(x, d_0(x), T(x, d_0(x), Z)) + \beta J_{N-1\vec{\pi}}(T(x, d_0(x), Z))\Big) \\ &= \mathcal{T}_{d_0} J_{N-1\vec{\pi}}(x), \end{aligned}$$

$x \in E$. The value function under planning horizon $N \in \mathbb{N}$ is given by

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

By Theorem 5.8, the value function satisfies the Bellman equation

$$J_N(x) = \mathcal{T} J_{N-1}(x) = \mathcal{T}^N 0(x), \qquad x \in E. \tag{5.8}$$

When the planning horizon is infinite, we define the value of a policy $\pi \in \Pi^M$ as

$$J_{\infty\pi}(x) = \lim_{N\to\infty} J_{N\pi}(x), \qquad x \in E. \tag{5.9}$$

Hence, the optimality criterion considered in this section is

$$J_\infty(x) = \inf_{\pi \in \Pi^M} J_{\infty\pi}(x), \qquad x \in E. \tag{5.10}$$

The next lemma shows that the infinite horizon policy value (5.9) and value function (5.10) are well-defined.

Lemma 5.10. *Under Assumption 5.9, the sequence $\{J_{N\pi}\}_{N\in\mathbb{N}}$ converges pointwise for every Markov policy $\pi \in \Pi^M$ and the limit function $J_{\infty\pi}$ is bounded by $\underline{\mathbf{b}}$ and $\bar{\mathbf{b}}$.*

Proof. First, we show by induction that for all $N \in \mathbb{N}$

$$J_{N\pi}(x) \geq J_{N-1\pi}(x) + (\alpha\beta)^{N-1}\underline{b}(x), \qquad x \in E. \tag{5.11}$$

For $N = 1$ we have by Assumption 5.9 (ii)

$$J_{1\pi}(x) \geq \underline{b}(x) = J_{0\pi}(x) + (\alpha\beta)^0\underline{b}(x).$$

For $N \geq 2$ it follows

$$\begin{aligned}
J_{N\pi}(x) &= \mathcal{T}_{d_0} J_{N-1\vec{\pi}}(x) \\
&= \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-1\vec{\pi}}(T(x, d_0(x), Z))\Big) \\
&\geq \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-2\vec{\pi}}(T(x, d_0(x), Z)) \\
&\qquad + \beta(\alpha\beta)^{N-2}\underline{b}(T(x, d_0(x), Z))\Big) \\
&\geq \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-2\vec{\pi}}(T(x, d_0(x), Z))\Big) \\
&\quad - \beta(\alpha\beta)^{N-2}\rho\Big(- \underline{b}(T(x, d_0(x), Z))\Big) \\
&\geq \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-2\vec{\pi}}(T(x, d_0(x), Z))\Big) + (\alpha\beta)^{N-1}\underline{b}(x) \\
&= J_{N-1\pi}(x) + (\alpha\beta)^{N-1}\underline{b}(x).
\end{aligned}$$

The first inequality is by the induction hypothesis, the second one is by Lemma 2.23

together with the positive homogeneity of ρ and the third one is due to Assumption 5.9 (ii). Thus, (5.11) holds. Applying this inequality repeatedly for $N, N-1, \dots, m$ yields

$$J_{N\pi}(x) \geq J_{m\pi}(x) + \sum_{k=m}^{N-1} (\alpha\beta)^k \underline{b}(x) \geq J_{m\pi}(x) + \delta_m(x),$$

where

$$\delta_m : E \to (-\infty, 0], \quad \delta_m(x) = \underline{b}(x) \sum_{k=m}^{\infty} (\alpha\beta)^k, \qquad m \in \mathbb{N}$$

are non-positive functions with $\lim_{m\to\infty} \delta_m(x) = 0$ for all $x \in E$. Hence, the sequence of functions $\{J_{N\pi}\}_{N\in\mathbb{N}}$ is weakly increasing and by Lemma A.9 a) convergent to a limit function $J_{\infty\pi}$. Clearly, the global bounds (5.6) also apply to the limit $J_{\infty\pi}$. □

Lemma 5.11. *Given Assumption 5.9, the Bellman operator $\mathcal{T}$ is a contraction on $I = [\underline{\mathbf{b}}, \bar{\mathbf{b}}]$ with modulus $\alpha\beta \in (0,1)$.*

Proof. Let $v \in I$. It has been established in the proof of Theorem 5.8 that $\mathcal{T}v$ is lower semicontinuous. Furthermore,

$$\begin{aligned}
\mathcal{T}v(x) &= \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \beta v(T(x,a,Z))\Big) \\
&\geq \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \frac{\beta}{1-\alpha\beta}\underline{b}(T(x,a,Z))\Big) \\
&\geq \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z))\Big) - \frac{\beta}{1-\alpha\beta}\rho\Big(-\underline{b}(T(x,a,Z))\Big) \\
&\geq \underline{b}(x) + \frac{\alpha\beta}{1-\alpha\beta}\underline{b}(x) \\
&= \underline{\mathbf{b}}(x).
\end{aligned}$$

The first inequality is by the monotonicity of ρ, the second one is by Lemma 2.23 together with the positive homogeneity of ρ and the third one is due to Assumption 5.9 (ii). Regarding the upper bounding function one can argue similarly, using the subadditivity of ρ instead of Lemma 2.23:

$$\begin{aligned}
\mathcal{T}v(x) &= \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \beta v(T(x,a,Z))\Big) \\
&\leq \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \frac{\beta}{1-\alpha\beta}\bar{b}(T(x,a,Z))\Big) \\
&\leq \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z))\Big) + \frac{\beta}{1-\alpha\beta}\rho\Big(\bar{b}(T(x,a,Z))\Big) \\
&\leq \bar{b}(x) + \frac{\alpha\beta}{1-\alpha\beta}\bar{b}(x) \\
&= \bar{\mathbf{b}}(x).
\end{aligned}$$

Hence, the operator $\mathcal{T}$ is an endofunction on I and it remains to verify the Lipschitz constant $\alpha\beta$. For $v_1, v_2 \in I$ it holds

$$
\begin{aligned}
|\mathcal{T}v_1(x) - \mathcal{T}v_2(x)| \le \sup_{a\in D(x)} \Big|&\rho\Big(c(x,a,T(x,a,Z)) + \beta v_1(T(x,a,Z))\Big) \\
&-\rho\Big(c(x,a,T(x,a,Z)) + \beta v_2(T(x,a,Z))\Big)\Big| \\
\le \beta \sup_{a\in D(x)} &\rho\Big(|v_1(T(x,a,Z)) - v_2(T(x,a,Z))| \Big) \\
\le \beta \sup_{a\in D(x)} &\rho\Big(\|v_1 - v_2\|_b b(T(x,a,Z))\Big) \\
= \beta\|v_1 - v_2\|_b &\sup_{a\in D(x)} \rho\Big(b(T(x,a,Z))\Big) \\
= \beta\|v_1 - v_2\|_b &\sup_{a\in D(x)} \rho\Big(\bar{b}(T(x,a,Z)) - \underline{b}(T(x,a,Z))\Big) \\
\le \beta\|v_1 - v_2\|_b &\sup_{a\in D(x)} \Bigg(\rho\Big(\bar{b}(T(x,a,Z))\Big) + \rho\Big(-\underline{b}(T(x,a,Z))\Big)\Bigg) \\
\le \alpha\beta\|v_1 - v_2\|_b &\Big(\bar{b}(x) - \underline{b}(x)\Big) \\
= \alpha\beta\|v_1 - v_2\|_b &b(x).
\end{aligned}
$$

Dividing by $b(x)$ and taking the supremum over $x \in E$ on the left hand side completes the proof. Note that the first inequality is by Lemma A.31, the second one is by Lemma 2.2, the third one is by definition of the weighted supremum norm, the fourth one is due to the subadditivity of ρ and the last one is by Assumption 5.9 (ii). □

Under a finite planning horizon $N \in \mathbb{N}$ we have characterized the value function with the Bellman equation (5.8). We will show that this is compatible with the optimality criterion of the infinite horizon model (5.10). To this end, we define the *limit value function*

$$J(x) = \lim_{N\to\infty} J_N(x), \qquad x \in E.$$

If existent, the limit value function lies in I due to Theorem 5.8. The existence follows from Theorem 5.12 below, which is the main result of this section.

Theorem 5.12. *Let Assumption 5.9 be satisfied. Then it holds:*

a) The limit value function J is the unique fixed point of the Bellman operator $\mathcal{T}$ in $I = [\underline{\mathbf{b}}, \bar{\mathbf{b}}]$.

b) There exists a Markov decision rule d^* such that

$$\mathcal{T}_{d^*} J(x) = \mathcal{T} J(x), \qquad x \in E.$$

c) Each stationary policy $\pi^* = (d^*, d^*, \dots)$ induced by a Markov decision rule d^* as in part b) is optimal for optimization problem (5.10) *and it holds $J_\infty = J$.*

Proof. a) The fact that J is the unique fixed point of the operator $\mathcal{T}$ in I follows directly from Banach's Fixed Point Theorem using Lemma 5.11.

b) The existence of a minimizing Markov decision rule follows from the respective result in the finite horizon case, cf. Theorem 5.8.

c) Let d^* be a Markov decision rule as in part b) and $\pi^* = (d^*, d^*, \dots)$. Then it holds

$$J(x) \le J_\infty(x) \le J_{\infty\pi^*}(x), \qquad x \in E.$$

The second inequality holds by definition. Regarding the first one note that for any $\pi \in \Pi^M$ we have $J_N(x) \le J_{N\pi}(x)$ for all $N \in \mathbb{N}_0$. Letting $N \to \infty$ yields $J(x) \le J_{\infty\pi}(x)$. Since $\pi \in \Pi^M$ was arbitrary we get $J(x) \le \inf_{\pi\in\Pi^M} J_{\infty\pi}(x) = J_\infty(x)$. It remains to show that

$$J_{\infty\pi^*}(x) \le J(x), \qquad x \in E. \tag{5.12}$$

To that end, we will prove by induction that for all $N \in \mathbb{N}_0$ and $x \in E$

$$J(x) \ge J_{N\pi^*}(x) + \frac{(\alpha\beta)^N}{1-\alpha\beta}\underline{b}(x). \tag{5.13}$$

Letting $N \to \infty$ in (5.13) yields (5.12) and concludes the proof.

For $N = 0$ equation (5.13) reduces to $J(x) \ge \frac{1}{1-\alpha\beta}\underline{b}(x) = \underline{\mathbf{b}}(x)$, which holds by part a). Now let $N \ge 1$. Then parts a) and b) together with the induction hypothesis yield

$$\begin{aligned}
J(x) &= \mathcal{T}_{d^*} J(x) \\
&= \rho\Big(c(x, d^*(x), T(x, d^*(x), Z)) + \beta J\big(T(x, d^*(x), Z)\big)\Big) \\
&\ge \rho\Big(c(x, d^*(x), T(x, d^*(x), Z)) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big) \\
&\qquad + \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\underline{b}\big(T(x, d^*(x), Z)\big)\Big) \\
&\ge \rho\Big(c(x, d^*(x), T(x, d^*(x), Z)) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big)\Big) \\
&\quad - \beta\frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\rho\Big(-\underline{b}\big(T(x, d^*(x), Z)\big)\Big) \\
&\ge \rho\Big(c(x, d^*(x), T(x, d^*(x), Z)) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big)\Big) + \frac{(\alpha\beta)^N}{1-\alpha\beta}\underline{b}(x) \\
&= J_{N\pi^*}(x) + \frac{(\alpha\beta)^N}{1-\alpha\beta}\underline{b}(x).
\end{aligned}$$

The second inequality is by Lemma 2.23 together with the positive homogeneity of ρ and the last one is by Assumption 5.9 (ii). □

Let us now consider the special case that the one-stage cost is bounded, i.e.

(B) there exist $\underline{b} \in \mathbb{R}_-$ and $\bar{b} \in \mathbb{R}_+$ such that $b = \bar{b} - \underline{b} > 0$ and $\underline{b} \leq c(x, a, T(x, a, Z)) \leq \bar{b}$ $\mathbb{P}$-f.s. for all $(x, a) \in D$.

Then, Assumption 5.9 (ii), (iii) are satisfied with $\alpha = 1$ for every normalized monetary risk measure. Part (v) of the assumption reduces to $\beta < 1$. In fact, all results of this section then hold for normalized monetary risk measures with the Fatou property.

Corollary 5.13. *Given (B), Lemmata 5.10, 5.11 and Theorem 5.12 hold for any normalized monetary risk measure with the Fatou property.*

Proof. The steps in the proofs that where justified by Lemma 2.23, subadditivity, positive homogeneity or Assumption 5.9 (ii) now all hold due to translation invariance and normalization. Apart from that, nothing has to be changed. Results from Section 5.1 can be applied since ρ has the necessary properties. □

5.3. Connection to Distributionally Robust Cost Minimization

We consider the stationary version of the abstract cost model with no terminal cost under both finite and infinite horizon in this section. If the planning horizon is finite, stationarity is only assumed for convenience and everything can be transferred to a non-stationary setting purely by notational changes. Let the risk measure ρ be proper and coherent with the Fatou property. By inserting the dual representation $\rho(X) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X],\ X \in L^p$, in the Bellman equation

$$\begin{aligned} J_N(x) &= 0, \\ J_n(x) &= \inf_{a \in D(x)} \rho\Big(c(x, a, T(x, a, Z)) + \beta J_{n+1}(T(x, a, Z))\Big), \qquad x \in E, \end{aligned}$$

we get

$$\begin{aligned} J_N(x) &= 0, \\ J_n(x) &= \inf_{a \in D(x)} \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\Big[c(x, a, T(x, a, Z)) + \beta J_{n+1}(T(x, a, Z))\Big], \qquad x \in E, \end{aligned}$$

i.e. the Bellman equation of the distributionally robust model of Chapter 4. Under some minor technical assumptions the two models can indeed be seen as special cases of each other. That is, the two optimality criteria induce the same optimal policy and the Markovian value functions coincide. This allows us to give a global interpretation of the recursively (locally) defined risk-sensitive optimality criterion.

Due to stationarity it is natural to make the following comparison based on the Assumptions 4.13 and 5.9 of the respective infinite horizon setting. Regarding the differences when the assumptions of the finite horizon case are taken as a basis, see Remark 5.15.

Theorem 5.14. *a) Consider the distributionally robust cost minimization of Chapter 4 with Assumption 4.13 being fulfilled. Let the ambiguity set $\mathcal{Q}$ be weak* closed, then we have a special case of the risk-sensitive recursive cost minimization of Chapter 5. That is, Assumption 5.9 is fulfilled and the value functions and the controller's optimal policies coincide.*

b) Consider the risk-sensitive recursive cost minimization of Chapter 5. Let Assumption 5.9 be fulfilled with the following tightening in part (ii):

$$\rho(c^-(x,a,T(x,a,Z))) \leq -\underline{b}(x), \quad \rho(c^+(x,a,T(x,a,Z))) \leq \bar{b}(x), \quad (x,a) \in D.$$

Furthermore, let the underlying probability space have a product structure

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{n=1}^{\infty} (\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$$

with $Z_n(\bar{\omega}) = Z_n(\omega_n)$ only depending on component ω_n of $\bar{\omega} = (\omega_1, \omega_2, \dots) \in \Omega$ and let the probability measure $\mathbb{P}_1$ on $(\Omega_1, \mathcal{A}_1)$ be separable. Then we have a special case of the distributionally robust cost minimization of Chapter 4. That is, Assumption 4.13 is fulfilled and the value functions and the controller's optimal policies coincide.

c) Given a fixed policy $\pi \in \Pi^M$ of the controller, the recursive risk measure constitutes in both cases a coherent risk measure

$$\tilde{\rho}(X) = \sup_{\mathbb{Q} \in \mathfrak{Q}_\pi} \mathbb{E}^{\mathbb{Q}}[X], \qquad X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$$

on the product space $(\Omega, \mathcal{A}, \mathbb{P})$ with ambiguity set $\mathfrak{Q}_\pi = \{\mathbb{Q}_x^{\pi\gamma} : \gamma \in \Gamma\}$. It is applied to the discounted total cost

$$\sum_{k=0}^{\infty} \beta^k c(X_k^\pi, d_k(X_k^\pi), X_{k+1}^\pi).$$

Proof. a) The ambiguity set is norm bounded and weak* closed, i.e. weak* compact by the Theorem of Banach-Alaoglu (Aliprantis and Border; 2006, 6.21). By Proposition 2.21, $\rho : L^p(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \to \mathbb{R}$ defined by

$$\rho(X) = \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X]$$

is a law-invariant, proper coherent risk measure with the Fatou property. Hence, the Bellman equations are equivalent and it remains to verify Assumption 5.9.

(i) This equals Assumption 4.13 (i).

(ii) It holds by Assumption 4.13 (ii) for all $\mathbb{Q} \in \mathcal{Q}$ and $(x,a) \in D$

$$\mathbb{E}^{\mathbb{Q}}\left[-c^-(x,a,T(x,a,Z))\right] \geq \underline{b}(x), \qquad \mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T(x,a,Z))\right] \geq \alpha \underline{b}(x),$$
$$\mathbb{E}^{\mathbb{Q}}\left[c^+(x,a,T(x,a,Z))\right] \leq \bar{b}(x), \qquad \mathbb{E}^{\mathbb{Q}}\left[\bar{b}(T(x,a,Z))\right] \leq \alpha \bar{b}(x).$$

The first inequality implies

$$\begin{aligned}\rho\big(c(x,a,T(x,a,Z))\big) &\geq \rho\big(-c^-(x,a,T(x,a,Z))\big)\\ &= \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[-c^-(x,a,T(x,a,Z))\right]\\ &\geq \underline{b}(x),\end{aligned}$$

the second one directly yields $\rho\big(-\underline{b}(T(x,a,Z))\big) \leq -\alpha\underline{b}(x)$, the third one implies

$$\begin{aligned}\rho\big(c(x,a,T(x,a,Z))\big) &\leq \rho\big(c^+(x,a,T(x,a,Z))\big)\\ &= \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c^+(x,a,T(x,a,Z))\right]\\ &\leq \bar{b}(x)\end{aligned}$$

and the fourth inequality again directly yields $\rho\big(\bar{b}(T(x,a,Z))\big) \leq \alpha\bar{b}(x)$. Hence, part (ii) is satisfied.

(iii) This equals Assumption 4.13 (iii).

(iv) The properties of the risk measure have been verified above.

(v) This equals Assumption 4.13 (vi).

b) By Proposition 2.21, ρ has a dual representation

$$\rho(X) = \max_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}[X],$$

where $\mathcal{Q} \subseteq \mathcal{M}_1^q(\Omega,\mathcal{A},\mathbb{P})$ is weak* compact and therefore norm bounded by the Theorem of Banach-Alaoglu (Aliprantis and Border; 2006, 6.21). Thus, the Bellman equations are equivalent and it remains to check the Assumption 4.13:

(i) This equals Assumption 5.9 (i).

(ii) It holds by Assumption 5.9 (ii) with the required tightening for all $(x,a) \in D$

$$\rho\big(c^-(x,a,T(x,a,Z))\big) \leq -\underline{b}(x), \qquad \rho\big(-\underline{b}(T(x,a,Z))\big) \leq -\alpha\underline{b}(x),$$
$$\rho\big(c^+(x,a,T(x,a,Z))\big) \leq \bar{b}(x), \qquad \rho\big(\bar{b}(T(x,a,Z))\big) \leq \alpha\bar{b}(x).$$

The first inequality implies for all $\mathbb{Q} \in \mathcal{Q}$

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[-c^-(x,a,T(x,a,Z))\right] &\geq \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[-c^-(x,a,T(x,a,Z))\right]\\ &= -\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c^-(x,a,T(x,a,Z))\right]\\ &= -\rho\big(c^-(x,a,T(x,a,Z))\big)\\ &\geq \underline{b}(x).\end{aligned}$$

The second inequality implies for all $\mathbb{Q} \in \mathcal{Q}$

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T(x,a,Z))\right] &\geq \inf_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[\underline{b}(T(x,a,Z))\right] \\ &= -\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[-\underline{b}(T(x,a,Z))\right] \\ &= -\rho\big(-\underline{b}(T(x,a,Z))\big) \\ &\geq \alpha \underline{b}(x).\end{aligned}$$

The third inequality implies for all $\mathbb{Q} \in \mathcal{Q}$

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[c^+(x,a,T(x,a,Z))\right] &\leq \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[c^+(x,a,T(x,a,Z))\right] \\ &= \rho\big(c^+(x,a,T(x,a,Z))\big) \\ &\leq \underline{b}(x).\end{aligned}$$

Finally, the last inequality yields for all $\mathbb{Q} \in \mathcal{Q}$

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\bar{b}(T(x,a,Z))\right] &\leq \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[\bar{b}(T(x,a,Z))\right] \\ &= \rho\big(\bar{b}(T(x,a,Z))\big) \\ &\leq \alpha \bar{b}(x).\end{aligned}$$

Thus, part (ii) is satisfied.

(iii) This equals Assumption 5.9 (iii).

(iv) This holds as a prerequisite.

(v) It has been verified above that $\mathcal{Q}$ is norm bounded.

(vi) This equals Assumption 5.9 (v).

c) The axioms of a coherent risk measure are readily checked for $\tilde{\rho}$. Note that compactness of the ambiguity set is only needed for finiteness and continuity properties. □

The prerequisite of Theorem 5.14 b) is indeed a tightening of Assumption 5.9 (ii) since it implies Assumption 4.13 by part b) of the theorem which in turn implies Assumption 5.9 by part a) of the theorem.

Remark 5.15. a) For the comparison of the distributionally robust and the risk-sensitive recursive cost minimization to make sense, one needs a weak* closed ambiguity set or a proper coherent risk measure with the Fatou property, respectively. Then the bounding function of the risk-sensitive recursive model can be constructed stage-wise, cf. Lemma 5.4. Now, the equivalence of the two optimality criteria can also be shown in the non-stationary case with finite planning horizon analogously to Theorem 5.14.

b) Under a finite planning horizon, the risk-sensitive recursive optimality criterion allows us to work with risk measures which do not possess a dual representation. In that sense, the recursive model is more general. On the other hand, the distributionally robust optimality criterion allows for non-compact ambiguity sets.

c) Ambiguity sets induced by a coherent risk measure guarantee the existence of an optimal policy of nature, cf. Proposition 2.21 and Theorem 4.12.

d) The global ambiguity set $\mathfrak{Q}_\pi$ is rectangular in the sense of Iyengar (2005) or Shapiro (2016). Its elements are absolutely continuous w.r.t. $\mathbb{P}$ due to component-wise absolute continuity. Note that there are no "π-factors" in (4.2) under a deterministic Markov policy.

5.4. Real Line as State Space

As for the distributionally robust model of Chapter 4, the continuity assumption on the transition functions can be relaxed to semicontinuity if the state space is the real line and the transition and one-stage cost function satisfy some form of monotonicity. For some applications as e.g. in Section 5.5, this relaxation is relevant. Moreover, the monotonicity properties allow for weaker assumptions on the risk measure if the one-stage cost function is bounded from below. To ease the notational burden, we consider the stationary model with no terminal cost under both finite and infinite horizon in this section.

5.4.1. Finite Planning Horizon

If the planning horizon is finite, all results can be transferred to a non-stationary setting by mere notational changes. We make the following assumptions.

Assumption 5.16. (i) The state space is the real line $E = \mathbb{R}$.

(ii) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being lower semicontinuous (case 2).

(iii) The model data has the following monotonicity properties:

(iii a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is decreasing.

(iii b) The transition function T is increasing in x.

(iii c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is increasing for all (a, z).

(iv) There exist $\underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$ and measurable functions $\underline{\mathbf{b}} : \mathbb{R} \to (-\infty, -\underline{\epsilon}]$, $\bar{\mathbf{b}} : \mathbb{R} \to [\bar{\epsilon}, \infty)$ such that it holds for all policies $\pi \in \Pi$ and all $n = 0, \dots, N$

$$\underline{\mathbf{b}}(x_n) \leq V_{n\pi}(h_n) \leq \bar{\mathbf{b}}(x_n), \qquad h_n \in \mathcal{H}_n.$$

(v) We define $\mathbf{b} : \mathbb{R} \to [1, \infty)$, $\mathbf{b}(x) = \bar{\mathbf{b}}(x) - \underline{\mathbf{b}}(x)$. For all $(\bar{x}, \bar{a}) \in D$ there exists an $\epsilon > 0$ and measurable functions $\Theta_1^{\bar{x},\bar{a}}, \Theta_2^{\bar{x},\bar{a}} : \mathcal{Z} \to \mathbb{R}_+$ such that $\Theta_1^{\bar{x},\bar{a}}(Z), \Theta_2^{\bar{x},\bar{a}}(Z) \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and

$$|c(x, a, T(x, a, z))| \leq \Theta_1^{\bar{x},\bar{a}}(z), \qquad \mathbf{b}(T(x, a, z)) \leq \Theta_2^{\bar{x},\bar{a}}(z)$$

for all $z \in \mathcal{Z}$ and $(x,a) \in B_\epsilon(\bar{x},\bar{a}) \cap D$. Here, $B_\epsilon(\bar{x},\bar{a})$ is the closed ball around $(\bar{x},\bar{a})$ w.r.t. an arbitrary product metric on $\mathbb{R} \times A$.

(vi) The monetary risk measures $\rho : L^p(\Omega,\mathcal{A},\mathbb{P}) \to \bar{\mathbb{R}}$ is law invariant and has the Fatou property.

The one-stage cost function c being increasing both in x and x' is sufficient for Assumption 5.16 (iii c) to hold since the transition function is increasing in x. Besides, if c is increasing in x', it is sufficient for Continuity and Compactness Assumption 3.1 (iii) that c is lower semicontinuous due to Lemma A.4 b). If the risk measure is additionally positive homogeneous and comonotonic additive, the existence of a global upper and lower bounding function can be guaranteed by suitable stage-wise bounding functions. This is similar to Lemma 5.4, however due to the real state space and monotonicity properties of the model, weaker conditions on the risk measure are sufficient. E.g. all distortion risk measures with the Fatou property are allowed.

Lemma 5.17. *Instead of Assumption 5.16 (iii c) let the one-stage cost function c be increasing in x, x' and in addition to (vi) let ρ be positive homogeneous and comonotonic additive. If there exist $\underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$, increasing functions $\underline{b} : \mathbb{R} \to (-\infty, -\underline{\epsilon}]$, $\bar{b} : \mathbb{R} \to [\bar{\epsilon}, \infty)$ and a constant $\alpha > 0$ such that $\alpha\beta \in (0,1)$ and*

$$\rho\big(c(x,a,T(x,a,Z))\big) \geq \underline{b}(x), \qquad \rho\big(\underline{b}(T(x,a,Z))\big) \geq \alpha \underline{b}(x),$$
$$\rho\big(c(x,a,T(x,a,Z))\big) \leq \bar{b}(x), \qquad \rho\big(\bar{b}(T(x,a,Z))\big) \leq \alpha \bar{b}(x),$$

for all $(x,a) \in D$, then

$$\underline{\mathbf{b}} = \frac{1}{1-\alpha\beta}\underline{b} \qquad \text{and} \qquad \bar{\mathbf{b}} = \frac{1}{1-\alpha\beta}\bar{b}$$

is a global lower and upper bounding function, respectively, and Assumption 5.16 (iv) holds.

Proof. We proceed by backward induction. At time N there is nothing to show. Assuming the assertion holds at time $n+1$, it follows for time n:

$$\begin{aligned}
V_{n\pi}(h_n) &= \rho\Big(c(x_n, d_n(h_n), T(x_n, d_n(h_n), Z)) + \beta V_{n+1\pi}(h_n, d_n(h_n), T(x_n, d_n(h_n), Z))\Big) \\
&\geq \rho\Big(c(x_n, d_n(h_n), T(x_n, d_n(h_n), Z)) + \frac{\beta}{1-\alpha\beta}\underline{b}(T(x_n, d_n(h_n), Z))\Big) \\
&= \rho\Big(c(x_n, d_n(h_n), T(x_n, d_n(h_n), Z))\Big) + \frac{\beta}{1-\alpha\beta}\rho\Big(\underline{b}(T(x_n, d_n(h_n), Z))\Big) \\
&\geq \underline{b}(x_n) + \frac{\alpha\beta}{1-\alpha\beta}\underline{b}(x_n) \\
&= \underline{\mathbf{b}}(x_n),
\end{aligned}$$

$\pi \in \Pi,\ h_n \in \mathcal{H}_n$. The first inequality is by the induction hypothesis and the monotonicity of ρ. The equality thereafter is by the comonotonic additivity and positive homogeneity of ρ. Regarding the upper bounding function one argues analogously. □

In Lemma 5.17, the stage-wise bounding functions are assumed to be increasing, which was not necessary in Lemma 5.4. Note that increasing functions are Borel measurable. Moreover, note that we only have to require $\rho(\underline{b}(T(x,a,Z))) \geq \alpha\underline{b}(x)$, $(x,a) \in D$ which is weaker than the corresponding assumption for the model with general state space, cf. Lemma 5.4 and Remark 5.5.

Since Assumption 5.16 (v) equals Assumption 4.13 (iii) and is independent of the optimality criterion, the separation condition of Corollary 4.20 applies here, too. The proof is exactly the same.

Corollary 5.18. *Let there be upper semicontinuous functions $\vartheta_1, \vartheta_2 : D \to \mathbb{R}_+$ and measurable functions $\Theta_1, \Theta_2 : \mathcal{Z} \to \mathbb{R}_+$ which fulfill $\Theta_1(Z), \Theta_2(Z) \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and*

$$|c(x,a,T(x,a,z))| \leq \vartheta_1(x,a) + \Theta_1(z), \qquad b(T(x,a,z)) \leq \vartheta_2(x,a) + \Theta_2(z)$$

for every $(x,a,z) \in D \times \mathcal{Z}$. Then Assumption 5.16 (v) is satisfied.

Apart from stationarity, Assumptions 5.2 and 5.16 differ only to the extend that the continuity of the transition function has been replaced by Assumption 5.16 (i) to (iii). How does this affect the validity of the results in Section 5.1? Lemmata 5.4 and 5.6 were proven without using the continuity of T. Thus, only Theorem 5.8 needs to be looked at.

Proposition 5.19. *The assertion of Theorem 5.8 remains true under Assumption 5.16. Moreover, the value functions J_n are increasing and the set of potential value functions can therefore be replaced by*

$$\mathbb{B} = \{v \in \mathbb{B}_{\mathbf{b}} : \ v \textit{ lower semicontinuous and increasing}\}.$$

Proof. The subset of increasing functions in $\{v \in \mathbb{B}_{\mathbf{b}} : \ v \text{ lower semicontinuous}\}$ is closed w.r.t. pointwise convergence, so especially w.r.t. $\|\cdot\|_{\mathbf{b}}$. Hence, $(\mathbb{B}, \|\cdot\|_{\mathbf{b}})$ is a complete metric space as a closed subset of complete metric space.

The proof of Theorem 5.8 uses the continuity of T only to show that $D \ni (x,a) \mapsto Lv(x,a)$ is lower semicontinuous for every $v \in \mathbb{B}$. Due to the monotonicity assumptions,

$$D \ni (x,a) \mapsto c\big(x,a,T(x,a,Z(\omega))\big) + \beta v\big(T(x,a,Z(\omega))\big)$$

is lower semicontinuous for every $\omega \in \Omega$ by part b) of Lemma A.4 (instead of part a) which is used in the proof of Theorem 5.8). Now, the lower semicontinuity of $D \ni (x,a) \mapsto Lv(x,a)$ and the existence of a minimizing decision rule follow as in the proof of Theorem 5.8. The fact that $\mathcal{T}v$ is increasing for every $v \in \mathbb{B}$ follows from Lemma A.19. □

When we refer to the interval $I = [\underline{\mathbf{b}}, \bar{\mathbf{b}}]$ in the following, it is to be understood as a subset of the modified function space $\mathbb{B}$ as in Proposition 5.19, i.e. it consists of increasing functions.

5.4.2. Infinite Planning Horizon

Now, let us consider an infinite planning horizon. Again the question is, how replacing the continuity of the transition function by Assumption 5.16 (i) to (iii) affects the results of Section 5.2. In detail, our assumptions are

Assumption 5.20. (i) The state space is the real line $E = \mathbb{R}$.

(ii) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being lower semicontinuous (case 2).

(iii) The model data has the following monotonicity properties:

(iii a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is decreasing.

(iii b) The transition function T is increasing in x.

(iii c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is increasing for all (a, z).

(iv) There exist $\alpha, \underline{\epsilon}, \bar{\epsilon} \geq 0$ with $\underline{\epsilon} + \bar{\epsilon} = 1$ and measurable functions $\underline{b} : \mathbb{R} \to (-\infty, -\underline{\epsilon}]$, $\bar{b} : \mathbb{R} \to [\bar{\epsilon}, \infty)$ such that for all $(x, a) \in D$

$$\rho\big(c(x, a, T(x, a, Z))\big) \geq \underline{b}(x), \qquad \rho\big(-\underline{b}(T(x, a, Z))\big) \leq -\alpha \underline{b}(x),$$
$$\rho\big(c(x, a, T(x, a, Z))\big) \leq \bar{b}(x), \qquad \rho\big(\bar{b}(T(x, a, Z))\big) \leq \alpha \bar{b}(x).$$

(v) We define $b : \mathbb{R} \to [1, \infty)$, $b(x) = \bar{b}(x) - \underline{b}(x)$. For all $(\bar{x}, \bar{a}) \in D$ there exists an $\epsilon > 0$ and measurable functions $\Theta_1^{\bar{x},\bar{a}}, \Theta_2^{\bar{x},\bar{a}} : \mathcal{Z} \to \mathbb{R}_+$ such that $\Theta_1^{\bar{x},\bar{a}}(Z), \Theta_2^{\bar{x},\bar{a}}(Z) \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and

$$|c(x, a, T(x, a, z))| \leq \Theta_1^{\bar{x},\bar{a}}(z), \qquad b(T(x, a, z)) \leq \Theta_2^{\bar{x},\bar{a}}(z)$$

for all $z \in \mathcal{Z}$ and $(x, a) \in B_\epsilon(\bar{x}, \bar{a}) \cap D$. Here, $B_\epsilon(\bar{x}, \bar{a})$ is the closed ball around $(\bar{x}, \bar{a})$ w.r.t. an arbitrary product metric on $\mathbb{R} \times A$.

(vi) The law-invariant risk measure $\rho : L^p(\Omega, \mathcal{A}, \mathbb{P}) \to \bar{\mathbb{R}}$ is proper, coherent and has the Fatou property.

(vii) The discount factor β satisfies $\alpha\beta < 1$.

The proofs of Lemmata 5.10 and 5.11 and Corollary 5.13 do not rely on the continuity of the transition function. Theorem 5.12 uses to continuity of T only indirectly through Theorem 5.8. In view of Proposition 5.19 we can conclude the following without proof.

Proposition 5.21. *Lemmata 5.10, 5.11, Theorem 5.12 and Corollary 5.13 hold under Assumption 5.20, too.*

In case the one-stage cost is bounded, Corollary 5.13 shows that a coherent risk measure is not necessary to solve the infinite horizon risk-sensitive recursive cost minimization problem. This result is very general regarding the risk measure but very restrictive concerning the one-stage cost. The monotone model with real state space allows for a middle course.

(B^-) There exist $\underline{b} \leq 0$, $\bar{\epsilon} \geq 0$ and $\alpha \geq 1$ with $\bar{\epsilon} - \underline{b} = 1$ and an increasing function

$\bar{b} : \mathbb{R} \to [\bar{\epsilon}, \infty)$ such that $c(x, a, T(x, a, Z)) \geq \underline{b}$ $\mathbb{P}$-f.s. and

$$\rho\big(c(x, a, T(x, a, Z))\big) \leq \bar{b}(x), \qquad\qquad \rho\big(\bar{b}(T(x, a, Z))\big) \leq \alpha \bar{b}(x).$$

for all $(x, a) \in D$.

W.l.o.g. we assume $\alpha \geq 1$ since then $\rho(-\underline{b}) = -\underline{b} \leq \alpha \underline{b}$ due to translation invariance and normalization. Otherwise one would need separate alphas for the lower and upper stage-wise bounding function.

If the one-stage cost function c is increasing in x' and the risk measure is comonotonic additive and positive homogeneous, the objective function is globally bounded under (B$^-$) due to Lemma 5.17. In that case Assumption 5.16 (iv) for the finite horizon can be replaced by (B$^-$) and the assertion of Theorem 5.8 remains true. Under an infinite planning horizon, Assumption 5.20 (iv) is clearly implied by (B$^-$). In that case, the results of Section 5.2 can be proven without requiring a coherent risk measure.

Proposition 5.22. *Let the one-stage cost function c be increasing in x' and let the Assumption 5.20 be satisfied with the modification that part (iv) is replaced by (B$^-$) and part (vi) by the requirement that ρ is a law invariant, comonotonic additive and positive homogeneous monetary risk measure with the Fatou property. Then it holds:*

a) The sequence $\{J_{N\pi}\}_{N \in \mathbb{N}}$ converges pointwise for every Markov policy $\pi \in \Pi^M$ and the limit function $J_{\infty\pi}$ is bounded by $\underline{\mathbf{b}}$ and $\bar{\mathbf{b}}$.

b) The Bellman operator $\mathcal{T}$ is a contraction on $I = [\underline{\mathbf{b}}, \bar{\mathbf{b}}] \subseteq \mathbb{B}$ with modulus $\alpha\beta \in (0, 1)$ and the limit value function J is the unique fixed point of $\mathcal{T}$ in I.

c) There exists a Markov decision rule d^* such that

$$\mathcal{T}_{d^*} J(x) = \mathcal{T} J(x), \qquad x \in \mathbb{R}.$$

Each stationary policy $\pi^* = (d^*, d^*, \dots)$ induced by such a Markov decision rule is optimal for optimization problem (5.10) *and it holds $J_\infty = J$.*

Proof. a) We show by induction that for all $N \in \mathbb{N}$

$$J_{N\pi}(x) \geq J_{N-1\pi}(x) + (\alpha\beta)^{N-1} \underline{b}, \qquad x \in \mathbb{R}. \tag{5.14}$$

For $N = 1$ it holds due to (B$^-$) that $J_{1\pi}(x) \geq \underline{b} = J_{0\pi}(x) + (\alpha\beta)^0 \underline{b}$. For $N \geq 2$ it follows with the monotonicity and translation invariance of ρ that

$$\begin{aligned} J_{N\pi}(x) &= \mathcal{T}_{d_0} J_{N-1\vec{\pi}}(x) \\ &= \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-1\vec{\pi}}(T(x, d_0(x), Z))\Big) \\ &\geq \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-2\vec{\pi}}(T(x, d_0(x), Z)) + \beta(\alpha\beta)^{N-2} \underline{b}\Big) \\ &= \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-2\vec{\pi}}(T(x, d_0(x), Z))\Big) + \beta(\alpha\beta)^{N-2} \underline{b} \end{aligned}$$

$$\geq \rho\Big(c(x, d_0(x), T(x, d_0(x), Z) + \beta J_{N-2\vec{\pi}}(T(x, d_0(x), Z))\Big) + (\alpha\beta)^{N-1}\underline{b}$$
$$= J_{N-1\pi}(x) + (\alpha\beta)^{N-1}\underline{b}.$$

Thus, (5.14) holds. Applying this inequality repeatedly for $N, N-1, \dots, m$ yields

$$J_{N\pi}(x) \geq J_{m\pi}(x) + \sum_{k=m}^{N-1}(\alpha\beta)^k\underline{b} \geq J_{m\pi}(x) + \sum_{k=m}^{\infty}(\alpha\beta)^k\underline{b}.$$

Since $\sum_{k=m}^{\infty}(\alpha\beta)^k\underline{b}$ is non-positive and converges to zero for $m \to \infty$, the sequence $\{J_{N\pi}\}_{N\in\mathbb{N}}$ is weakly increasing and by Lemma A.9 a) convergent to a limit function $J_{\infty\pi}$. Clearly, the global bounds $\underline{\mathbf{b}}, \bar{\mathbf{b}}(\cdot)$ also apply to the limit $J_{\infty\pi}$.

b) Let $v \in I$. Due to Proposition 5.19 $\mathcal{T}v$ is increasing and lower semicontinuous. Furthermore, the monotonicity and translation invariance of ρ imply

$$\begin{aligned}
\mathcal{T}v(x) &= \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \beta v(T(x,a,Z))\Big) \\
&\geq \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \frac{\beta}{1-\alpha\beta}\underline{b}\Big) \\
&= \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z))\Big) + \frac{\beta}{1-\alpha\beta}\underline{b} \\
&\geq \underline{b} + \frac{\alpha\beta}{1-\alpha\beta}\underline{b} = \underline{\mathbf{b}}.
\end{aligned}$$

Regarding the upper bounding function it holds

$$\begin{aligned}
\mathcal{T}v(x) &= \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \beta v(T(x,a,Z))\Big) \\
&\leq \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z)) + \frac{\beta}{1-\alpha\beta}\bar{b}(T(x,a,Z))\Big) \\
&= \inf_{a\in D(x)} \rho\Big(c(x,a,T(x,a,Z))\Big) + \frac{\beta}{1-\alpha\beta}\rho\Big(\bar{b}(T(x,a,Z))\Big) \\
&\leq \bar{b}(x) + \frac{\alpha\beta}{1-\alpha\beta}\bar{b}(x) = \bar{\mathbf{b}}(x).
\end{aligned}$$

Here, the second equality is by the comonotonic additivity and positive homogeneity of the risk measure. Thus, the operator $\mathcal{T}$ is an endofunction on I and it remains to verify the Lipschitz constant $\alpha\beta$. For $v_1, v_2 \in I$ it holds

$$\begin{aligned}
\mathcal{T}v_1(x) - \mathcal{T}v_2(x) \leq \sup_{a\in D(x)} \; &\rho\Big(c(x,a,T(x,a,Z)) + \beta v_1(T(x,a,Z))\Big) \\
&- \rho\Big(c(x,a,T(x,a,Z)) + \beta v_2(T(x,a,Z))\Big) \\
= \beta \sup_{a\in D(x)} \; &\rho\Big(v_1(T(x,a,Z))\Big) - \rho\Big(v_2(T(x,a,Z))\Big) \\
= \beta \sup_{a\in D(x)} \; &\rho\Big(v_1(T(x,a,Z)) - v_2(T(x,a,Z)) + v_2(T(x,a,Z))\Big)
\end{aligned}$$

$$
\begin{aligned}
&\qquad - \rho\Big(v_2(T(x,a,Z))\Big) \\
&\leq \beta \sup_{a \in D(x)} \rho\Big(\|v_1 - v_2\|_b b(T(x,a,Z)) + v_2(T(x,a,Z))\Big) \\
&\qquad - \rho\Big(v_2(T(x,a,Z))\Big) \\
&= \beta \sup_{a \in D(x)} \rho\Big(\|v_1 - v_2\|_b b(T(x,a,Z))\Big) + \rho\Big(v_2(T(x,a,Z))\Big) \\
&\qquad - \rho\Big(v_2(T(x,a,Z))\Big) \\
&= \|v_1 - v_2\|_b \beta \sup_{a \in D(x)} \rho\Big(b(T(x,a,Z))\Big) \\
&= \|v_1 - v_2\|_b \beta \sup_{a \in D(x)} \left(\rho\Big(\bar{b}(T(x,a,Z))\Big) - \underline{b}\right) \\
&\leq \alpha\beta \|v_1 - v_2\|_b (\bar{b}(x) - \underline{b}) \\
&= \alpha\beta \|v_1 - v_2\|_b b(x).
\end{aligned}
$$

The first inequality is by Lemma A.31 and the equality thereafter is by comonotonic additivity and positive homogeneity. Since $\underline{b}$ is constant, $b(\cdot) = \bar{b}(\cdot) - \underline{b}$ is an increasing function and so is v_2. Therefore, the third equality is again by comonotonic additivity. The last inequality is by (B^-) using $\alpha \geq 1$. Interchanging the roles of v_1 and v_2 yields

$$|\mathcal{T} v_1(x) - \mathcal{T} v_2(x)| \leq \alpha\beta \|v_1 - v_2\|_b b(x).$$

Finally, dividing by $b(x)$ and taking the supremum over $x \in \mathbb{R}$ on the left hand side gives

$$\|\mathcal{T} v_1 - \mathcal{T} v_2\|_b \leq \alpha\beta \|v_1 - v_2\|_b.$$

Now, Banach's Fixed Point Theorem states that J is the unique fixed point of $\mathcal{T}$ in I.

c) The existence of a minimizing Markov decision rule follows from the finite horizon case, cf. Proposition 5.19.

With the same argument as in the proof of Theorem 5.12, the relation $J \leq J_\infty \leq J_{\infty\pi}$ holds for any policy and it remains to show that $J_{\infty\pi^*} \leq J$ for the specific policy π^*. To that end, we will prove by induction that for all $N \in \mathbb{N}_0$ and $x \in \mathbb{R}$

$$J(x) \geq J_{N\pi^*}(x) + \frac{(\alpha\beta)^N}{1 - \alpha\beta} \underline{b}.$$

Then, letting $N \to \infty$ concludes the proof. The case $N = 0$, i.e. $J(x) \geq \frac{1}{1-\alpha\beta}\underline{b}$, holds by part b). For $N \geq 1$ we have

$$
\begin{aligned}
J(x) &= \mathcal{T}_{d^*} J(x) \\
&= \rho\Big(c(x, d^*(x), T(x, d^*(x), Z)) + \beta J(T(x, d^*(x), Z))\Big)
\end{aligned}
$$

$$
\begin{aligned}
&\geq \rho\Big(c\big(x, d^*(x), T(x, d^*(x), Z)\big) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big) + \beta \frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\underline{b}\Big) \\
&= \rho\Big(c\big(x, d^*(x), T(x, d^*(x), Z)\big) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big)\Big) + \beta \frac{(\alpha\beta)^{N-1}}{1-\alpha\beta}\underline{b} \\
&\geq \rho\Big(c\big(x, d^*(x), T(x, d^*(x), Z)\big) + \beta J_{N-1\pi^*}\big(T(x, d^*(x), Z)\big)\Big) + \frac{(\alpha\beta)^{N}}{1-\alpha\beta}\underline{b} \\
&= J_{N\pi^*}(x) + \frac{(\alpha\beta)^{N}}{1-\alpha\beta}\underline{b}.
\end{aligned}
$$

The first inequality is by the induction hypothesis and the monotonicity of ρ, the equality thereafter is by translation invariance and the second inequality holds since $\alpha \geq 1$. □

The prerequisite of Proposition 5.22 is satisfied by any distortion risk measure with the Fatou property due to Lemma 2.5. As in Section 4.3, the monotonicity requirements in Assumptions 5.16 and 5.20 are only one option. The following alternative is relevant i.a. for the dynamic reinsurance models introduced in Section 3.2.

Corollary 5.23. *Assumption 5.16 (ii) and (iii) and Assumption 5.20 (ii) to (iii) can be replaced by*

(ii') The model data has the Continuity and Compactness Properties 3.1 with the transition function T being upper semicontinuous (case 3).

(iii') The model data has the following monotonicity properties:

- *(iii' a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is increasing.*
- *(iii' b) The transition function T is increasing in x.*
- *(iii' c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is decreasing for all (a, z).*

Then, the assertion of Theorems 5.8 and 5.12 still hold. Moreover, the value functions J_n are decreasing and the set of potential value functions is

$$\mathbb{B} = \{v \in \mathbb{B}_{\mathbf{b}} : \ v \text{ lower semicontinuous and decreasing}\}.$$

Lemma 5.17 and Proposition 5.22 remain true, too, with the adaption that the stage-wise bounding functions need to be decreasing in x.

The proof is analogous to the one of Corollary 4.22. Requiring that the one-stage cost function c is decreasing both in x and x' is sufficient for (iii' c) since the transition function is increasing in x. While this condition might seem more natural, assuming the monotonicity only for the composition is relevant for some applications. E.g. in the dynamic reinsurance model for minimization of the cost of solvency capital (Section 3.2.1), the incremental version of one-stage cost function $c(x, f, x') = x - x'$ is not decreasing in x but the composition $c(x, f, T(x, f, y, z)) = f(y) + \pi_R(f) - z$ is. Besides, if c is decreasing in x', it is sufficient for Continuity and Compactness Assumption 3.1 (iii) that c is lower semicontinuous due to Remark A.5.

5.5. Cost of Capital Minimization of an Insurance Company

As an application of the risk-sensitive recursive cost minimization we consider the minimization of the cost of capital of an insurance company in the dynamic reinsurance model of Section 3.2.1. This is a dynamic extension in discrete time of the static optimal reinsurance problem

$$\min_{f \in \mathcal{F}} \quad r_{\text{CoC}} \cdot \rho\big(f(Y) + \pi_R(f)\big), \tag{5.15}$$

which has been studied extensively in the literature, starting with Cai and Tan (2007) and generalizations i.a. by Chi and Tan (2013) and Cui et al. (2013). Here, the cost of solvency capital is calculated as the cost of capital rate $r_{\text{CoC}} \in (0,1]$ times the solvency capital requirement which is determined by applying the risk measure ρ to the insurer's effective risk after reinsurance consisting of the retained loss and the cost of reinsurance.

In the terminal period $[N-1, N)$ of the dynamic reinsurance model of Section 3.2.1, the insurer faces the same problem: minimizing the cost of solvency capital

$$J_{N-1}(x) = \min_{f \in D(x)} r_{\text{CoC}} \cdot \rho\Big(f(Y_N) + \pi_R(f) - Z_N - x\Big)$$

for the effective risk consisting of the retained loss and the cost of reinsurance minus the premium income during the period and the capital at the beginning of the period, i.e. the state of the surplus process, over all admissible reinsurance treaties $f \in D(x)$. In terms of the abstract cost model this means that the one-stage cost function is given by the loss (negative surplus) of the next stage $c(x, f, x') = -x'$.

In any earlier period $[n, n+1)$, the effective risk consists of the risk for that period plus the discounted future cost of capital which is a random variable as a measurable function of the next state of the surplus process. To simplify the notation, we assume that the cost of capital rate r_{CoC} is included in the discount factor β and obtain the minimization problem

$$J_n(x) = \min_{f \in D(x)} \rho\Big(f(Y_{n+1}) + \pi_R(f) - Z_{n+1} - x + \beta J_{n+1}\big(x + Z_{n+1} - f(Y_{n+1}) - \pi_R(f)\big)\Big).$$

In the first period, one has to multiply once more with the cost of capital rate in order to obtain the overall recursive cost of capital, but for the minimization this is of course not relevant.

Remark 5.24. Assuming that the cost of capital rate is included in the discount factor means that β is of the form

$$\beta = r_{\text{CoC}} \cdot \frac{1}{1+r},$$

where $r \in (0,1]$ is the risk-free interest rate per period. Hence, the discount factor still is a quantity in $(0,1]$ and our simplification of the notation entails no restriction.

It has been shown in Lemma 3.3 that the dynamic reinsurance model of Section 3.2.1 is a special case of the abstract cost model of Section 3.1. In order to formally introduce the dynamic optimal reinsurance problem as a special case of the abstract risk-sensitive recursive cost minimization, we have to specify the value of a policy $\pi = (d_0, \dots, d_{N-1}) \in \Pi$ with $d_n : \mathcal{H}_n \to \mathcal{F}$ s.t. $d_n(h_n) \in D(x_n)$ for all $h_n \in \mathcal{H}_n$:

$$\begin{aligned}
V_N(h_N) &= 0,\\
V_{n\pi}(h_n) &= \rho\left(-X^\pi_{n+1} + \beta V_{n+1\pi}(h_n, d_n(h_n), X^\pi_{n+1})\right)\\
&= \rho\Big(d_n(h_n)(Y_{n+1}) + \pi_R(d_n(h_n)) - Z_{n+1} - x_n\\
&\qquad + \beta V_{n+1\pi}\big(h_n, d_n(h_n), x_n + Z_{n+1} - d_n(h_n)(Y_{n+1}) - \pi_R(d_n(h_n))\big)\Big).
\end{aligned}$$

The corresponding value functions are

$$V_n(h_n) = \inf_{\pi\in\Pi} V_{n\pi}(h_n), \qquad h_n \in \mathcal{H}_n,$$

and the optimization objective is to determine the optimal recursive cost of solvency capital

$$V_0(x) = \inf_{\pi\in\Pi} V_{0\pi}(x), \qquad x \in \mathbb{R}. \tag{5.16}$$

Due to the real state space we want to apply Corollary 5.23 for solving the optimization problem. Let us verify the assumptions. The numbering is as in the corollary.

(i) The state space is the real line $E = \mathbb{R}$.

(ii') The Continuity and Compactness Properties 3.1 with upper semicontinuous transition function have been verified in Section 3.2.1.

(iii') Monotonicity properties:

(iii' a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x) = \{f \in \mathcal{F} : \pi_R(f) \le x^+\}$ is increasing.
(iii' b) The transition function $T : \mathbb{R} \times \mathcal{F} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, $T(x, f, y, z) = x - f(y) - \pi_R(f) + z$ is increasing in x.
(iii' c) The one-stage cost function $c : \mathbb{R} \times \mathcal{F} \times \mathbb{R} \to \mathbb{R}$, $c(x, f, x') = -x'$ is decreasing in x' and the composition $\mathbb{R} \ni x \mapsto c(x, f, T(x, f, y, z)) = f(y) + \pi_R(f) - z - x$ is decreasing for all (f, y, z).

(iv) It will be shown that

$$\begin{aligned}
\underline{b}(x) &= \frac{-1}{1-\beta}x^+ + \frac{1}{(1-\beta)^2}\underline{\eta}, && \underline{\eta} = -\operatorname{ess\,sup}(Z) < 0,\\
\bar{b}(x) &= \frac{1}{1-\beta}x^- + \frac{1}{(1-\beta)^2}\bar{\eta}, && \bar{\eta} = \rho(Y) + \pi_R(Y) > 0,\\
b(x) = \bar{b}(x) - \underline{b}(x) &= \frac{1}{1-\beta}|x| + \frac{1}{(1-\beta)^2}\eta, && \eta = \rho(Y) + \pi_R(Y) + \operatorname{ess\,sup}(Z)
\end{aligned}$$

$x \in \mathbb{R}$, are decreasing stage-wise bounding functions in the sense of Lemma 5.17 and Assumption 5.20 (iv), where $\alpha = \frac{1-(1-\beta)^2}{\beta}$. Note that the risk measure ρ has the necessary properties for a stage-wise bounding approach, see (vii), and that $\beta \in (0,1)$, see (viii). For normalization of the weighted supremum norm we assume w.l.o.g. $\eta \geq 1$. It holds for $(x,f) \in D$:

$$\begin{aligned}
\rho\big(c(x,f,T(x,f,Y,Z))\big) &= \rho\big(f(Y) + \pi_R(f) - Z - x\big) \\
&\geq \rho(-Z) - x \\
&\geq \underline{b}(x), \\
\rho\big(\underline{b}(T(x,f,Y,Z))\big) &= \frac{\underline{\eta}}{(1-\beta)^2} + \frac{1}{1-\beta}\rho\big(\min\{f(Y) + \pi_R(f) - Z - x, 0\}\big) \\
&\geq \frac{\underline{\eta}}{(1-\beta)^2} + \frac{1}{1-\beta}\rho\big(-Z - x^+\big) \\
&\geq \frac{-x^+}{1-\beta} + \underline{\eta}\left(\frac{1}{1-\beta} + \frac{1}{(1-\beta)^2}\right) \\
&= -x^+\frac{1}{\beta}\left(\frac{1}{1-\beta} - 1\right) + \underline{\eta}\frac{1}{\beta}\left(\frac{1}{(1-\beta)^2} - 1\right) \\
&= \frac{1}{\beta}\left(\underline{b}(x) + x^+ - \underline{\eta}\right) \\
&\geq \frac{1}{\beta}\left(\underline{b}(x) + (1-\beta)x^+ - \underline{\eta}\right) \\
&= \frac{1-(1-\beta)^2}{\beta}\underline{b}(x).
\end{aligned}$$

The second equality holds since

$$\frac{1}{1-\beta} = 1 + \frac{\beta}{1-\beta} \qquad \text{and} \qquad 1 + \frac{\beta}{1-\beta} + \frac{\beta}{(1-\beta)^2} = \frac{1}{(1-\beta)^2}.$$

Regarding the upper bounding function one argues analogously:

$$\begin{aligned}
\rho\big(c(x,f,T(x,f,Y,Z))\big) &= \rho\big(f(Y) + \pi_R(f) - Z - x\big) \\
&\leq \rho(Y) + \pi_R(Y) - x \\
&\leq \bar{b}(x), \\
\rho\big(\bar{b}(T(x,f,Y,Z))\big) &= \frac{\bar{\eta}}{(1-\beta)^2} + \frac{1}{1-\beta}\rho\big(\max\{f(Y) + \pi_R(f) - Z - x, 0\}\big) \\
&\leq \frac{\bar{\eta}}{(1-\beta)^2} + \frac{1}{1-\beta}\big(\rho(Y) + \pi_R(Y) + x^-\big) \\
&= \frac{x^-}{1-\beta} + \bar{\eta}\left(\frac{1}{1-\beta} + \frac{1}{(1-\beta)^2}\right) \\
&= x^-\frac{1}{\beta}\left(\frac{1}{1-\beta} - 1\right) + \bar{\eta}\frac{1}{\beta}\left(\frac{1}{(1-\beta)^2} - 1\right) \\
&= \frac{1}{\beta}\left(\bar{b}(x) - x^- - \bar{\eta}\right)
\end{aligned}$$

$$\leq \frac{1}{\beta}\left(\bar{b}(x) - (1-\beta)x^- - \bar{\eta}\right)$$
$$= \frac{1-(1-\beta)^2}{\beta}\bar{b}(x).$$

(v) Here, we use the separation condition of Lemma 5.18:

$$\begin{aligned}|c(x,f,T(x,f,Y,Z))| &= |f(Y) + \pi_R(f) - Z - x| \\ &\leq f(Y) + \pi_R(f) + Z + |x| \\ &\leq Y + Z + \pi(Y) + |x|,\end{aligned}$$

i.e $\vartheta_1(x) = \pi(Y) + |x|$, which is continuous, and $\Theta_1(y,z) = y + z$, which satisfies $\Theta_1(Y,Z) \in L^p(\Omega, \mathcal{A}, \mathbb{P})$. Furthermore,

$$\begin{aligned}b(T(x,f,Y,Z) &= \frac{\eta}{(1-\beta)^2} + \frac{1}{1-\beta}\,|x + Z - f(Y) - \pi_R(f)| \\ &\leq \frac{\eta}{(1-\beta)^2} + \frac{1}{1-\beta}\,(|x| + Z + Y + \pi_R(Y)) \\ &= \frac{Y+Z}{1-\beta} + \frac{\eta}{(1-\beta)^2} + \frac{|x| + \pi_R(Y)}{1-\beta},\end{aligned}$$

implying that $\vartheta_2(x) = \frac{\eta}{(1-\beta)^2} + \frac{|x|+\pi_R(Y)}{1-\beta}$, which is continuous, and $\Theta_2(y,z) = \frac{y+z}{1-\beta}$, which satisfies $\Theta_2(Y,Z) \in L^p(\Omega, \mathcal{A}, \mathbb{P})$.

(vi) We assume that ρ is a law-invariant monetary risk measure with the Fatou property. Under a finite planning horizon we have to require positive homogeneity and comonotonic additivity as additional properties, while under an infinite planning horizon we require properness and coherence. Possible examples include all distortion risk measures with the Fatou property under a finite planning horizon and all spectral risk measures with a spectrum in L^q under an infinite planning horizon, cf. Lemma 2.5 and Corollary 2.17.

(vii) We assume $\beta \in (0,1)$. Thus, it holds for the modulus of the Bellman operator $\alpha\beta = 1 - (1-\beta)^2 \in (0,1)$.

Hence, Corollary 5.23 implies that it is sufficient for the insurer to minimize over all Markov policies, the value functions lie in the interval $I = [\underline{\mathbf{b}}, \bar{\mathbf{b}}] = \left[\frac{\underline{b}}{(1-\beta)^2}, \frac{\bar{b}}{(1-\beta)^2}\right]$ and satisfy the Bellman equation

$$J_N(x) = 0,$$
$$J_n(x) = \inf_{f \in D(x)} \rho\Big(f(Y) + \pi_R(f) - Z - x + \beta J_{n+1}(x + Z - f(Y) - \pi_R(f))\Big), \qquad x \in \mathbb{R},$$

for $n = 0, \dots, N-1$. There exists a Markov Decision rule $d_n^* : \mathbb{R} \to \mathcal{F}$ minimizing J_{n+1} and every sequence $\pi = (d_0^*, \dots, d_{N-1}^*) \in \Pi^M$ of such minimizers is at solution to (5.16).

Furthermore, under the necessary assumptions for an infinite planning horizon the Bellman operator

$$\mathcal{T} : I \to I,\ \mathcal{T}v(x) = \inf_{f \in D(x)} \rho\Big(f(Y) + \pi_R(f) - Z - x + \beta v(x + Z - f(Y) - \pi_R(f))\Big)$$

is a contraction with modulus $1 - (1-\beta)^2$ and the limit value function J is its unique fixed point. Every stationary policy $\pi = (d^*, d^*, \dots) \in \Pi^S$ induced by a minimizer d^* of J is optimal for the insurer under an infinite planning horizon.

Remark 5.25. Here, we have the special situation that $\underline{b}, \bar{b}$ are even global bounding functions. This can be seen by backward induction. Let $\pi \in \Pi$ be arbitrary. Obviously, it holds $\underline{b}(x_N) \le V_{N\pi}(h_N) = 0 \le \bar{b}(x_N)$ for all $h_N \in \mathcal{H}_N$. Now assuming the assertion holds at time $n+1$, it follows at time n for all $h_n \in \mathcal{H}_n$

$$\begin{aligned}
V_{n\pi}(h_n) &= \rho\left(-X_{n+1}^{\pi} + \beta V_{n+1\pi}(h_n, d_n(h_n), X_{n+1}^{\pi})\right) \\
&\ge \rho\left(-X_{n+1}^{\pi} + \beta \underline{b}(X_{n+1}^{\pi})\right) \\
&= \rho\left(-X_{n+1}^{\pi}\right) + \beta\rho\left(\underline{b}(X_{n+1}^{\pi})\right) \\
&= \rho\Big(d_n(h_n)(Y) + \pi_R(d_n(h_n)) - Z - x_n\Big) \\
&\quad + \beta\rho\Big(\underline{b}(x_n + Z - d_n(h_n)(Y) - \pi_R(d_n(h_n))\Big) \\
&\ge \underline{\eta} - x_n + \beta\rho\Big(\underline{b}(x_n + Z)\Big) \\
&= \underline{\eta} - x_n + \beta\rho\left(\frac{\underline{\eta}}{(1-\beta)^2} + \frac{\min\{-Z - x_n, 0\}}{1-\beta}\right) \\
&\ge \underline{\eta}\left(1 + \frac{\beta}{1-\beta} + \frac{\beta}{(1-\beta)^2}\right) - x_n^+\left(1 + \frac{\beta}{1-\beta}\right) \\
&= \underline{b}(x_n).
\end{aligned}$$

For the second equality we used comonotonic additivity and positive homogeneity. If ρ is coherent one argues instead with Lemma 2.23. Analogously, one can show $V_{n\pi}(h_n) \le \bar{b}(x_n)$. Consequently, the value functions J_n and J lie in the smaller subinterval $[\underline{b}, \bar{b}] \subseteq I$.

In Lemmata 5.4 and 5.17 it has been shown that every stage-wise (lower/ upper) bounding function induces a global one through the relation

$$\underline{\mathbf{b}} = \frac{1}{(1-\beta)^2}\underline{b} \qquad \text{and} \qquad \bar{\mathbf{b}} = \frac{1}{(1-\beta)^2}\bar{b}.$$

Here, we can see that the converse is not true in general: $\underline{b}, \bar{b}$ are global bounding functions, but $(1-\beta)^2\underline{b}, (1-\beta)^2\bar{b}$ are not stage-wise bounding functions. E.g.

$$\rho\big(c(x, f, T(x, f, Y, Z))\big) = \rho\big(f(Y) + \pi_R(f) - Z\big) - x \not\ge -(1-\beta)x^+ + \underline{\eta}$$

for sufficiently large $x \ge 0$.

In Section 3.2.1, we introduced the possibility of no budget constraint in the dynamic reinsurance model. It significantly simplifies the optimization problem.

Remark 5.26. In case of no budget constraint $D(x) = \mathcal{F}$ for all $x \in \mathbb{R}$, the dynamic optimization problem (5.16) reduces to a static problem and there is a constant optimal action. This can be seen by backward induction. At time $N-1$, the Bellman equation reads due to the translation invariance of ρ

$$J_{N-1}(x) = \min_{f \in \mathcal{F}} \rho\Big(f(Y) - Z\Big) + \pi_R(f) - x,$$

i.e. the minimization does not depend on the state of the surplus process x. Therefore, the value function is of the form

$$J_{N-1}(x) = c - x$$

with a constant $c = \min_{f \in \mathcal{F}} \rho(f(Y) - Z) + \pi_R(f)$ and the optimal decision rule d^*_{N-1} is constant

$$d^*_{N-1}(x) = \operatorname{argmin}_{f \in \mathcal{F}} \rho(f(Y) - Z) + \pi_R(f) =: f^*, \qquad x \in \mathbb{R}.$$

Proceeding to the previous time step, we get due to translation invariance and positive homogeneity of ρ

$$\begin{aligned} J_{N-2}(x) &= \min_{f \in \mathcal{F}} \rho\Big(f(Y) + \pi_R(f) - Z - x + \beta J_{N-1}(x + Z - f(Y) - \pi_R(f))\Big) \\ &= \min_{f \in \mathcal{F}} \rho\Big(f(Y) + \pi_R(f) - Z - x + \beta(c + f(Y) + \pi_R(f) - Z - x)\Big) \\ &= \min_{f \in \mathcal{F}} (1+\beta)\Big(\rho(f(Y) - Z) + \pi_R(f)\Big) - (1+\beta)x + \beta c. \end{aligned}$$

Again, the minimization does not depend on x, the value function is given by

$$J_{N-2}(x) = (1 + 2\beta)c - (1+\beta)x$$

and the optimal decision rule is $d^*_{N-2} \equiv f^*$. Proceeding with the induction, one finds that the value functions are affine and structurally related to the bounding functions

$$\begin{aligned} J_n(x) &= c \sum_{k=0}^{N-n-1} (k+1)\beta^k - x \sum_{k=0}^{N-n-1} \beta^k, && x \in \mathbb{R},\ n = 0, \dots, N-1, \\ J(x) &= \frac{c}{(1-\beta)^2} - \frac{x}{1-\beta}, && x \in \mathbb{R}. \end{aligned}$$

Moreover, there is a retained loss function $f^* \in \mathcal{F}$ which is optimal at each point in time independently from the state of the surplus process. It can be determined by solving the classical static optimal reinsurance problem

$$\min_{f \in \mathcal{F}} \ \rho(f(Y) - Z) + \pi_R(f).$$

In order to prove this, it remains to verify the induction step. Due to translation invariance and positive homogeneity of ρ it follows

$$\begin{aligned}
J_n(x) &= \min_{f\in\mathcal{F}} \rho\Big(f(Y) + \pi_R(f) - Z - x + \beta J_{n+1}(x + Z - f(Y) - \pi_R(f))\Big) \\
&= \min_{f\in\mathcal{F}} \rho\Big(f(Y) + \pi_R(f) - Z - x + c\beta \sum_{k=0}^{N-n-2} (k+1)\beta^k \\
&\qquad + \beta(f(Y) + \pi_R(f) - Z - x) \sum_{k=0}^{N-n-2} \beta^k\Big) \\
&= \min_{f\in\mathcal{F}} \Big(\rho(f(Y) - Z) + \pi_R(f)\Big) \sum_{k=0}^{N-n-1} \beta^k + c\beta \sum_{k=0}^{N-n-2} (k+1)\beta^k - x \sum_{k=0}^{N-n-1} \beta^k \\
&= c\left(\sum_{k=0}^{N-n-1} \beta^k + \beta \sum_{k=0}^{N-n-2} (k+1)\beta^k\right) - x \sum_{k=0}^{N-n-1} \beta^k \\
&= c \sum_{k=0}^{N-n-1} (k+1)\beta^k - x \sum_{k=0}^{N-n-1} \beta^k.
\end{aligned}$$

Having a constant optimal action in case of no budget constraint especially means that the optimal policy is myopic. The following example studies Value-at-Risk as a concrete choice for the risk measure ρ. This choice has particular practical relevance with regard to Solvency II. Due to specific properties of Value-at-Risk, we obtain a myopic optimal policy even in combination with a budget constraint, but not a constant optimal action.

Example 5.27. Let $\rho = \mathrm{VaR}_\alpha$. We consider an arbitrary premium principle from the large class of Wang premium principles

$$\pi_R(X) = (1+\theta) \int_0^\infty g(S_X(x))\,\mathrm{d}\,x, \quad \theta \geq 0,$$

where we only assume that the distortion function g is left-continuous. This includes any PH premium, especially the expected premium principle. Furthermore, it is assumed that the insurer's premium income is deterministic, i.e. $Z \equiv z \in \mathbb{R}_+$. Economically, this means that the insurer either has customers with a good credit rating or a large homogeneous portfolio such that fluctuations of individual premium payments (approximately) cancel out.

We have to solve the Bellman equation

$$\begin{aligned}
J_n(x) &= \inf_{f\in D(x)} \mathrm{VaR}_\alpha\Big(f(Y) + \pi_R(f) - z - x + \beta J_{n+1}\big(-(f(Y) + \pi_R(f) - z - x)\big)\Big) \\
&= \inf_{f\in D(x)} \mathrm{VaR}_\alpha\Big(h(f(Y) + \pi_R(f) - z - x)\Big) \\
&= \inf_{f\in D(x)} h\Big(\mathrm{VaR}_\alpha\,(f(Y) + \pi_R(f) - z - x)\Big) \\
&= \inf_{f\in D(x)} h\Big(f(\mathrm{VaR}_\alpha(Y)) + \pi_R(f) - z - x\Big).
\end{aligned}$$

Here, we defined $h(x) = x + \beta J_{n+1}(-x)$ and applied Lemma B.9. Note that h is increasing and left-continuous since $x \mapsto J_{n+1}(-x)$ is increasing and lower semicontinuous, cf. Lemma A.6. The increasing transformation h can be dropped and it remains to solve

$$\inf_{f \in D(x)} f\big(\operatorname{VaR}_\alpha(Y)\big) + \pi_R(f). \tag{5.17}$$

This is a static optimal reinsurance problem with budget constraint. I.e., the dynamic reinsurance problem (5.16) possesses a myopic and stationary optimal policy.

We will first reduce (5.17) to a finite dimensional problem, extending an approach used in Chi and Tan (2013) and Bäuerle and Glauner (2018) to problems with constraints. Then, we will derive an explicit solution of the reduced problem. Define

$$h_a(x) = \max\left\{\min\{a, x\},\ x - \operatorname{VaR}_\alpha(Y) + a\right\}, \qquad x \in \mathbb{R}_+,\ 0 \le a \le \operatorname{VaR}_\alpha(Y).$$

This is the retained loss function corresponding to a layer reinsurance treaty with deductible a and upper bound $\operatorname{VaR}_\alpha(Y) - a$. Clearly, $h_a \in \mathcal{F}$ for all $a \in [0, \operatorname{VaR}_\alpha(Y)]$. Fix $f \in \mathcal{F}$. We write h_f short hand for h_a when $a = f(\operatorname{VaR}_\alpha(Y))$. Observe that $f(\operatorname{VaR}_\alpha(Y)) \in [0, \operatorname{VaR}_\alpha(Y)]$. Simply by inserting we get

$$\begin{aligned} h_f(\operatorname{VaR}_\alpha(Y)) = \max\Big\{ &\min\{f(\operatorname{VaR}_\alpha(Y)), \operatorname{VaR}_\alpha(Y)\}, \\ &\operatorname{VaR}_\alpha(Y) - \operatorname{VaR}_\alpha(Y) + f(\operatorname{VaR}_\alpha(Y))\Big\} \\ = f(&\operatorname{VaR}_\alpha(Y)). \end{aligned}$$

Moreover, it holds $\pi_R(h_f) \le \pi_R(f)$. This can be seen as follows. If $0 \le x < f(\operatorname{VaR}_\alpha(Y))$, then $h_f(x) = x \ge f(x)$ as f is bounded by the identity. If $f(\operatorname{VaR}_\alpha(Y)) \le x < \operatorname{VaR}_\alpha(Y)$, then $h_f(x) = f(\operatorname{VaR}_\alpha(Y)) \ge f(x)$ since f is increasing. Finally if $x \ge \operatorname{VaR}_\alpha(Y)$, then $h_f(x) = x - \operatorname{VaR}_\alpha(Y) + f(\operatorname{VaR}_\alpha(Y)) \ge f(x)$ as f is 1-Lipschitz, cf. Lemma 3.3 a). Consequently, $Y - h_f(Y) \le Y - f(Y)$ and by monotonicity $\pi_R(h_f) \le \pi_R(f)$. I.e. h_f is weakly better than f with respect to the objective function of (5.17) and satisfies the constraint if f does. Therefore, it suffices to consider the reduced problem

$$\inf_{0 \le a \le \operatorname{VaR}_\alpha(Y)} a + \pi_R(h_a) \quad \text{such that} \quad \pi_R(h_a) \le x^+. \tag{5.18}$$

In order to determine the premium, let us consider the survival function of $Y - h_a(Y) = \min\{(Y-a)^+, \operatorname{VaR}_\alpha(Y) - a\}$:

$$\mathbb{P}(Y - h_a(Y) > y) = \begin{cases} \mathbb{P}((Y-a)_+ > y) = S_Y(y+a), & 0 \le y < \operatorname{VaR}_\alpha(Y) - a, \\ 0, & y \ge \operatorname{VaR}_\alpha(Y) - a. \end{cases}$$

It follows

$$\pi_R(h_a) = (1+\theta) \int_0^\infty g\big(\mathbb{P}(Y - h_a(Y) > y)\big) \,\mathrm{d}\, y$$

$$
\begin{aligned}
&= (1+\theta) \int_0^{\mathrm{VaR}_\alpha(Y)-a} g\big(S_Y(y+a)\big)\,\mathrm{d}\,y \\
&= (1+\theta) \int_a^{\mathrm{VaR}_\alpha(Y)} g\big(S_Y(y)\big)\,\mathrm{d}\,y
\end{aligned}
$$

The derivative of the objective function

$$
\psi(a) = a + (1+\theta) \int_a^{\mathrm{VaR}_\alpha(Y)} g\big(S_Y(y)\big)\,\mathrm{d}\,y, \qquad 0 \le a \le \mathrm{VaR}_\alpha(Y)
$$

is given by $\psi'(a) = 1 - (1+\theta)g(S_Y(a))$. Since the distortion function g is left-continuous, $g \circ S_Y$ is itself a survival function. Thus, ψ' is increasing and right continuous. I.e. its generalized inverse

$$
\psi'^{-1}(z) = \inf\{a \in [0, \mathrm{VaR}_\alpha(Y)] : \ \psi'(a) \ge z\}
$$

is well-defined for every z in the range of ψ'. Let us distinguish two cases:
Case 1: $g(1-\alpha) < \frac{1}{1+\theta}$
By Lemma B.8 we have $S_Y(\mathrm{VaR}_\alpha(Y)) \le 1-\alpha$. Since g is increasing it follows

$$
\begin{aligned}
\psi'(\mathrm{VaR}_\alpha(Y)) &= 1 - (1+\theta)g\big(S_Y(\mathrm{VaR}_\alpha(Y))\big) \\
&\ge 1 - (1+\theta)g(1-\alpha) \\
&> 1 - (1+\theta)\frac{1}{1+\theta} = 0.
\end{aligned}
$$

Hence, ψ is strictly increasing on $[\psi'^{-1}(0), \mathrm{VaR}_\alpha(Y)]$.
Case 2: $g(1-\alpha) \ge \frac{1}{1+\theta}$
Let $a < \mathrm{VaR}_\alpha(Y)$. Then $S_Y(a) > 1-\alpha$ by Lemma B.8 and as g is increasing

$$
\begin{aligned}
\psi'(a) &= 1 - (1+\theta)g\big(S_Y(a)\big) \\
&\le 1 - (1+\theta)g(1-\alpha) \\
&\le 1 - (1+\theta)\frac{1}{1+\theta} = 0.
\end{aligned}
$$

I.e., ψ is decreasing on $[0, \mathrm{VaR}_\alpha(Y)]$.
Note that in practice α is chosen very close to 1 and θ smaller than 1, so only the first case is actually relevant. Let us define

$$
a^* = \begin{cases} \psi'^{-1}(0), & \text{if } g(1-\alpha) < \frac{1}{1+\theta}, \\ \mathrm{VaR}_\alpha(Y), & \text{otherwise.} \end{cases}
$$

Note that $a = \mathrm{VaR}_\alpha(Y)$ is always feasible for optimization problem (5.18) and that $a \mapsto \pi_R(h_{a,n})$ is a continuous mapping. Therefore, taking into account the budget constraint we obtain as an optimal solution of (5.18):

$$\delta(x) = \min\{a \in [a^*, \mathrm{VaR}_\alpha(Y)] : \ \pi_R(h_a) \leq x^+\}.$$

Consequently, an optimal policy for the dynamic reinsurance problem (5.16) is given by $\pi = (d^*, \dots, d^*)$, where $d^*(x) = h_{\delta(x)}$. We have seen that it is an optimal policy to buy a layer reinsurance treaty at each time step where the single parameter is chosen as close to the optimal parameter of the corresponding problem without constraint as the current surplus allows. For a low surplus, this means that the insurer should invest all capital in reinsurance to mitigate future insurance claims rather than saving capital to pay for them himself. Consequently, it is sufficient to act optimally in every period as if the optimization problem were static. Long term planning is not necessary.

Our second example studies the behavior of coherent and especially spectral risk measures in the cost of capital minimization problem (5.16) using the connection to distributionally robust MDP discussed in Section 5.3.

Example 5.28. Let ρ be a proper coherent risk measure with the Fatou property. We want to apply Theorem 5.14 b) in order to treat the recursive cost of capital minimization as a distributionally robust MDP. Since the claims $Y_1, Y_2, \dots$ and premium income $Z_1, Z_2, \dots$ are i.i.d. we can w.l.o.g. assume that the probability space has a product structure

$$(\Omega, \mathcal{A}, \mathbb{P}) = \bigotimes_{n=1}^{\infty} (\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$$

with $(Y_n, Z_n)(\bar{\omega}) = (Y_n, Z_n)(\omega_n)$ only depending on component ω_n of $\bar{\omega} = (\omega_1, \omega_2, \dots) \in \Omega$. The probability measure $\mathbb{P}_1$ on $(\Omega_1, \mathcal{A}_1)$ can be assumed as separable since $\mathcal{B}(\mathbb{R}^2_+)$ is countably generated (apply Lemma B.5 and a canonical construction). Also the additional assumptions on the one-stage bounding functions are satisfied:

$$\begin{aligned}
\rho\big(c^-(x, f, T(x, f, Y, Z))\big) &= \rho\big(\max\{0, x + Z - f(Y) - \pi_R(f)\}\big) \\
&\leq x^+ + \operatorname{ess\,sup}(Z) \\
&\leq -\underline{b}(x), \\
\rho\big(c^+(x, f, T(x, f, Y, z))\big) &= \rho\big(\max\{0, f(Y) + \pi_R(f) - z - x\}\big) \\
&\leq \rho(Y) + \pi(Y) + x^- \\
&\leq \bar{b}(x),
\end{aligned}$$

for all $(x, a) \in D$. Hence, we have a special case of the distributionally robust cost minimization of Chapter 4 and get an expression in closed from for the recursively defined optimality criterion (5.16):

$$\inf_{\pi \in \Pi^M} \sup_{\gamma \in \Gamma^M} \mathbb{E}_{0x}^{\pi\gamma} \left[\sum_{k=0}^{N-1} \beta^k \Big(d_k(X_k)(Y_{k+1}) + \pi_R(d_k(X_k)) - Z_{k+1} - X_k \Big) \right]$$

$$
\begin{aligned}
&= -\sup_{\pi\in\Pi^M}\inf_{\gamma\in\Gamma^M}\mathbb{E}_{0x}^{\pi\gamma}\left[\sum_{k=0}^{N-1}\beta^k\Big(X_k + Z_{k+1} - d_k(X_k)(Y_{k+1}) - \pi_R(d_k(X_k))\Big)\right]\\
&= \sum_{k=0}^{N-1}\beta^k x\\
&\quad -\sup_{\pi\in\Pi^M}\inf_{\gamma\in\Gamma^M}\mathbb{E}^{\pi\gamma}\left[\sum_{k=0}^{N-1}\Big(\sum_{j=k}^{N-1}\beta^j\Big)\Big(Z_{k+1} - d_k(X_k)(Y_{k+1}) - \pi_R(d_k(X_k))\Big)\right]. \qquad (5.19)
\end{aligned}
$$

The second equality can be obtained inductively by inserting the representation

$$
X_k = X_{k-1} + Z_k - d_{k-1}(X_{k-1})(Y_k) - \pi_R(d_{k-1}(X_{k-1})), \qquad k = 1,\dots,N-1
$$

given by the transition function. Here, we have a robust maximization of total profit with higher weights on earlier periods. This addresses a fundamental criticism of cost of capital minimization as an optimality criterion for reinsurance design by Albrecher et al. (2017, Sec. 8.4). The authors state that if the minimization of the cost of capital was the driving criterion of the insurer, it would be optimal in the long run to stay out of business altogether and thereby achieve zero cost of capital. This viewpoint brings the suitability of the recursive optimality criterion (5.16) into question since it would be applied over several periods. However, under a coherent risk measure the calculations above show that the recursive criterion is indeed in accordance with the primary target of any insurer: profit maximization.

Now consider the special case that ρ is a spectral risk measure with spectrum $\phi \in L^q$ and the premium income $Z \equiv z$ is deterministic. By Lemma 4.32 and the monotonicity properties of the model, $\phi(U_Y)$ defines a constant optimal action for nature and it remains to solve a risk-neutral MDP with disturbance distribution $\mathrm{d}\,\mathbb{Q}^* = \phi(U_Y)\,\mathrm{d}\,\mathbb{P}^Y$. I.e. the recursive cost of capital is given by the expected discounted loss under a new probability measure $\widehat{\mathbb{Q}} = \bigotimes_{k=1}^{\infty}\mathbb{Q}^*$. Furthermore, optimization problem (5.16) is equivalent to

$$
\inf_{\pi\in\Pi^M}\rho\left(\sum_{k=0}^{N-1}\beta^k\Big(d_k(X_k)(Y_{k+1}) + \pi_R(d_k(X_k)) - z - X_k\Big)\right)
$$

with a coherent risk measure $\hat{\rho}(X) = \sup_{\mathbb{Q}\in\mathfrak{Q}}\mathbb{E}^{\mathbb{Q}}[X],\ X \in L^p(\Omega,\mathcal{A},\mathbb{P})$, as in (4.33) where

$$
\mathfrak{Q} = \left\{\bigotimes_{k=1}^{\infty}\mathbb{Q}_k :\ \mathrm{d}\,\mathbb{Q}_k = Y_k\,\mathrm{d}\,\mathbb{P}_1,\ Y_k \in L^q(\Omega_1,\mathcal{A}_1,\mathbb{P}_1),\ Y_k \le_{cx} \phi(U),\ U \sim \mathcal{U}(0,1)\right\}.
$$

Here, it might seem unnatural to sum over states of the surplus process. As we have seen in (5.19), this gives a higher weight to income and claims of earlier periods. An alternative is to use the other one-stage cost function $c(x,f,x') = x' - x$ introduced in Section 3.2.1. Then the cost of capital or capital requirement is calculated for the present value of the total loss

$$\inf_{\pi\in\Pi^M} \rho\left(\sum_{k=0}^{N-1} \beta^k \Big(d_k(X_k)(Y_{k+1}) + \pi_R(d_k(X_k)) - z\Big)\right) - x_0.$$

Globally, this approach might be more natural. But stage-wise, we have the Bellman equation

$$J_n(x) = \inf_{f\in D(x)} \rho\Big(f(Y) + \pi_R(f) - Z + \beta J_{n+1}\big(x + Z - f(Y) - \pi_R(f)\big)\Big),$$

i.e. the insurer's current capital is no longer directly taken into account for determining the recursive capital requirement. Only the initial capital reduces the capital requirement at time zero.

CHAPTER 6

Risk-Sensitive Total Cost Minimization

In Lemma 4.32 and subsequent remarks together with Theorem 5.14 we have seen that, given a product structure of the underlying probability space and sufficient monotonicity properties of the Markov decision model, the distributionally robust cost minimization of Chapter 4 with an ambiguity set induced by a spectral risk measure or equivalently the risk-sensitive recursive cost minimization of Chapter 5 with a spectral risk measure can be perceived as the minimization of some non-standard coherent risk measure $\tilde{\rho}$ applied to the total discounted cost

$$\inf_{\pi \in \Pi} \tilde{\rho}\left(\sum_{k=0}^{N-1} \beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\right),$$

where $\tilde{\rho}(X) = \sup_{\mathbb{Q} \in \mathfrak{Q}} \mathbb{E}^{\mathbb{Q}}[X],\ X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and

$$\mathfrak{Q} = \left\{ \bigotimes_{k=1}^{\infty} \mathbb{Q}_k :\ \mathrm{d}\,\mathbb{Q}_k = Y_k \,\mathrm{d}\,\mathbb{P}_1,\ Y_k \in L^q(\Omega_1, \mathcal{A}_1, \mathbb{P}_1),\ Y_k \leq_{cx} \phi(U),\ U \sim \mathcal{U}(0,1) \right\}.$$

Example 5.28 illustrated this observation with the dynamic cost of capital minimization of an insurance company.

In this chapter, we proceed in the reverse order and consider the minimization of a spectral risk measure ρ_ϕ applied to the total discounted cost

$$\inf_{\pi \in \Pi} \rho_\phi\left(\sum_{k=0}^{N-1} \beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\right). \tag{6.1}$$

Dynamic programming techniques cannot be applied straightforwardly to problem (6.1) since it does not admit a value iteration. This is due to the fact that spectral risk measures in general lack a tower property like the one of conditional expectation. In Chapter 5, this difficulty was avoided since when applying the risk measure recursively a value iteration holds by construction. Due to the dual representation of spectral risk measures in Proposition 2.24 one can reformulate (6.1) to

$$\inf_{\pi\in\Pi}\sup_{\mathbb{Q}\in\mathcal{Q}}\mathbb{E}^{\mathbb{Q}}\left[\sum_{k=0}^{N-1}\beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\right]$$

with $\mathcal{Q}=\{\mathbb{Q}\in\mathcal{M}_1^q(\Omega,\mathcal{A},\mathbb{P}) : \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \leq_{cx} \phi(U),\ U\sim\mathcal{U}(0,1)\}$, i.e. a distributionally robust cost minimization, however with a non-rectangular ambiguity set in the sense of Iyengar (2005). Hence, the results of Chapter 4 do not apply and we will therefore not consider the problem from a robust viewpoint.

Optimization problem (6.1) has been studied by Bäuerle and Ott (2011) in the special case that ρ_ϕ is the Expected Shortfall. Using the infimum representation (2.4) of Expected Shortfall and interchanging infima

$$\begin{aligned}
&\inf_{\pi\in\Pi}\mathrm{ES}_\alpha\left(\sum_{k=0}^{N-1}\beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\right)\\
&=\inf_{\pi\in\Pi}\inf_{q\in\mathbb{R}}\left\{q+\frac{1}{1-\alpha}\mathbb{E}\left[\left(\sum_{k=0}^{N-1}\beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)-q\right)^+\right]\right\}\\
&=\inf_{q\in\mathbb{R}}\inf_{\pi\in\Pi}\left\{q+\frac{1}{1-\alpha}\mathbb{E}\left[\left(\sum_{k=0}^{N-1}\beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)-q\right)^+\right]\right\}\\
&=\inf_{q\in\mathbb{R}}\left\{q+\frac{1}{1-\alpha}\inf_{\pi\in\Pi}\mathbb{E}\left[\left(\sum_{k=0}^{N-1}\beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)-q\right)^+\right]\right\} \qquad (6.2)
\end{aligned}$$

they showed that the inner optimization problem of (6.2) can be solved as an ordinary Markov Decision Process on an extended state space. Earlier, Bäuerle and Mundt (2009) solved a mean-Expected Shortfall problem for an investor in a binomial financial market. There, the risk measure appears in the constraint but becomes part of the objective function through a Lagrangian approach.

Here, we are going to generalize the results of Bäuerle and Ott (2011) to spectral risk measures. Moreover, we are able to relax their assumptions on the one-stage cost functions and allow for unbounded above costs. In principle, the approach remains the same: We will use the infimum representation for spectral risk measures of Proposition 2.11 instead of equation (2.4). In the inner optimization problem, the functions $\mathbb{R}\ni x\mapsto(x-q)^+$ are then replaced by general increasing convex functions $g:\mathbb{R}\to\mathbb{R}$. Therefore, also the outer optimization problem becomes harder since it is no longer parametric but one has to minimize over an infinite dimensional function space. For the new outer problem we discuss both existence and an algorithmic approximation.

6.1. Inner Problem

In this section, we separate (6.1) into an inner and outer problem and solve the inner one using a state space extension. First, we consider a finite planning horizon in Section 6.1.1 and then an infinite planning horizon in Section 6.1.2. Additional model properties in case of a real state space are studied in Section 6.1.3.

6.1.1. Finite Planning Horizon

Under a finite planning horizon $N \in \mathbb{N}$, we consider the non-stationary version of the abstract cost model introduced in Section 3.1 with deterministic policies $\pi \in \Pi$ of the controller. The Markov Decision Process therefore has the functional representation (3.3). Here, it is more convenient to index the process and its random history with the policy since we will not explicitly refer to the law of motion. Even though the model is non-stationary we will explicitly introduce discounting by a factor $\beta \in (0, 1]$ 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. The total discounted cost generated by a policy $\pi \in \Pi$ if the initial state is $X_0^\pi = x$, is denoted by

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

In the following, it is assumed that the one-stage cost $c_n(x, a, T_n(x, a, Z_{n+1}))$ is bounded below by a constant $\underline{c} \in \mathbb{R}$ for all $(x, a) \in D_n$, $n = 0, \ldots, N-1$ and the same applies to the terminal cost $c_N(T_{N-1}(x, a, Z_N))$ for all $(x, a) \in D_{N-1}$. Due to the translation invariance of ρ_ϕ, we have for every policy $\pi \in \Pi$

$$\rho_\phi(C_N^{\pi x}) = \rho_\phi\left(\sum_{k=0}^{N-1} \beta^k \big(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi) - \underline{c}\big) + \beta^N \big(c_N(X_N^\pi - \underline{c})\big)\right) + \sum_{k=0}^{N} \beta^k \underline{c}.$$

Since $c_n(x, a, T_n(x, a, Z_{n+1})) - \underline{c} \geq 0,\ k = 0, \ldots, N-1$ and $c_N(T_{N-1}(x, a, Z_N)) - \underline{c} \geq 0$, we can assume w.l.o.g. that the one-stage and terminal cost is non-negative.

Using Proposition 2.11, we can reformulate optimization problem (6.1) to

$$\begin{aligned}
\inf_{\pi \in \Pi} \rho_\phi\left(C_N^{\pi x}\right) &= \inf_{\pi \in \Pi} \inf_{g \in G} \left\{ \mathbb{E}[g(C_N^{\pi x})] + \int_0^1 g^*(\phi(u)) \,\mathrm{d}\, u \right\} \\
&= \inf_{g \in G} \inf_{\pi \in \Pi} \left\{ \mathbb{E}[g(C_N^{\pi x})] + \int_0^1 g^*(\phi(u)) \,\mathrm{d}\, u \right\} \\
&= \inf_{g \in G} \left\{ \inf_{\pi \in \Pi} \mathbb{E}[g(C_N^{\pi x})] + \int_0^1 g^*(\phi(u)) \,\mathrm{d}\, u \right\}, \qquad (6.3)
\end{aligned}$$

where G denotes the set of increasing convex functions $g : \mathbb{R} \to \mathbb{R}$. For fixed $g \in G$ we will refer to

$$\inf_{\pi \in \Pi} \mathbb{E}[g(C_N^{\pi x})] \tag{6.4}$$

as *inner optimization problem.* Since an increasing convex function $g : \mathbb{R} \to \mathbb{R}$ can be viewed as a disutility function, optimality criterion (6.4) 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 $g^{-1}(\mathbb{E}[g(C_N^{\pi x})])$. 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

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

see Example 2.13. 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 $u^{-1}(\rho_\phi(u(X)))$. 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 (6.4), only g is replaced by $g(u(\cdot))$. At least for bounded costs, 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.* In this special case, our results apply without further conditions.

The following assumptions are made in this section.

Assumption 6.1. (i) The model data has the Continuity and Compactness Properties 3.1 with the transition function T_n being continuous in (x, a) for $n = 0, \dots, N-1$ (case 1).

(ii) The one-stage cost $c_n(x, a, T_n(x, a, Z_{n+1}))$ and the terminal cost $c_N(T_{N-1}(x, a, Z_N))$ are non-negative for all $(x, a) \in D_n$, $n = 0, \dots, N-1$.

(iii) The family of random variables $\{C_N^{\pi x} : \pi \in \Pi, x \in E\}$ is uniformly integrable.

(iv) The spectrum ϕ is bounded, i.e. $\phi(1) < \infty$.

Since spectral risk measures preserve the increasing convex order (Lemma 2.5), the following equivalent characterization of Assumption 6.1 (iii) will be useful.

Lemma 6.2. *Let Assumption 6.1 (ii) be statisfied. Then, Assumption 6.1 (iii) is equivalent to the existence of a non-negative random variable $\bar{C} \in L^1_+$ on some probability space such that*

$$C_N^{\pi x} \leq_{icx} \bar{C}$$

for all policies $\pi \in \Pi$ and initial states $x \in E$.

Proof. By Assumption 6.1 (ii) the random variables $C_N^{\pi x}$, $\pi \in \Pi$, $x \in E$, are non-negative. Now the assertion follows from Theorem 1 in Leskelä and Vihola (2013). □

Assumption 6.1 (iii) is rather general. The next lemma gives a sufficient condition in terms of properties of the model data.

Lemma 6.3. *If there exists a measurable function $\bar{c} : \mathcal{Z} \to \mathbb{R}_+$ such that*

$$\begin{aligned} c_n(x, a, T_n(x, a, z)) &\leq \bar{c}(z), & (x, a, z) &\in D_n \times \mathcal{Z},\ n = 0, \dots, N-1, \\ c_N(T_{N-1}(x, a, z)) &\leq \bar{c}(z), & (x, a, z) &\in D_{N-1} \times \mathcal{Z}, \end{aligned}$$

and $\bar{c}(Z_{n+1}) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, $n = 0, \dots, N-1$, then Assumption 6.1 (iii) is satisfied.

Proof. We have for all policies $\pi \in \Pi$, initial states $x \in E$ and time points $n = 0, \dots, N-1$

$$c_n(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) = c_n(X_n^\pi, d_n(H_n^\pi), T_n(X_n^\pi, d_n(H_n^\pi), Z_{n+1})) \leq \bar{c}(Z_{n+1}).$$

The inequality for the terminal cost in analogous. It follows

$$C_N^{\pi x} = \sum_{k=0}^{N-1} \beta^k c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi) + \beta^N c_N(X_N^\pi) \leq \sum_{k=0}^{N-1} \beta^k \bar{c}(Z_{k+1}) + \beta^N \bar{c}(Z_N) =: \bar{C} \in L^1.$$

A common integrable majorant is sufficient for uniform integrability. □

Spectral risk measures are finite for risks in L^p if the spectrum is in L^q, see Lemma 2.17. Since we require the spectrum ϕ to be bounded, there is no need to restrict the cost to L^p for some $p > 1$. Hence, we only require integrability or $\bar{C} \in L^1$.

The bounded spectrum enables us to reduce the function space G. The reduction guarantees finite policy values in (6.4). Moreover, it will be needed under an infinite planning horizon and to solve the outer optimization problem (6.3).

Lemma 6.4. *Under Assumption 6.1 it is sufficient to consider functions $g \in G$ which are $\phi(1)$-Lipschitz and satisfy $0 \leq g(s) \leq \bar{g}(s)$, $s \in \mathbb{R}$, where*

$$\bar{g}(s) = \phi(1) s^+ + \rho_\phi(\bar{C}).$$

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

Proof. Fix $\pi \in \Pi$, $x \in E$ and set $C = C_N^{\pi x}$ to simplify the notation. We know from the proof of Proposition 2.11 that the optimal $g \in G$ corresponding to C is

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

with μ from Proposition 2.9. Clearly, it is sufficient to consider functions $g \in G$ which are optimal for at least one $C = C_N^{\pi x}$. Since $C \geq 0$ it follows

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

Furthermore, we have

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

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

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

is by definition of μ. The second inequality holds due to Lemma 6.2, since spectral risk measures preserve the increasing convex order (Lemma 2.5). As a convex function, $g_{\phi,C}$ is almost everywhere differentiable with derivative $g'_{\phi,C}(s) = \phi(F_C(s)) \leq \phi(1)$, cf. the proof of Proposition 2.11. This establishes the Lipschitz continuity with constant $L = \phi(1)$. □

In the sequel, we are going to solve the inner optimization problem (6.4) for an arbitrary but fixed function $g \in \mathcal{G}$. Lemma 6.4 guarantees that the optimization problem is well-defined under Assumption 6.1. Indeed, we have for every initial state $x \in E$, policy $\pi \in \Pi$ and $g \in \mathcal{G}$

$$\begin{aligned} &\mathbb{E}[g^-(C_N^{\pi x})] \geq 0, \\ &\mathbb{E}[g^+(C_N^{\pi x})] \leq \mathbb{E}[g^+(\bar{C})] \leq \mathbb{E}[\bar{g}(\bar{C})] = \phi(1)\mathbb{E}[\bar{C}] + \rho_\phi(\bar{C}) < \infty \end{aligned}$$

since $\bar{C} \in L^1$. Here, we have used that g^+ is increasing convex and $\bar{g}$ non-negative.

As the functions $g \in \mathcal{G}$ are in general non-linear, the inner optimization problem (6.4) does not directly admit a value iteration. This can be overcome by extending the state space to

$$\widehat{E} = E \times \mathbb{R}_+ \times (0, 1]$$

with corresponding Borel σ-algebra following Bäuerle and Rieder (2014). A generic element of $\widehat{E}$ 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, $n=0,\dots,N-1$ remain unchanged. Formally, one defines

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

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

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

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

$$h_n = \begin{cases} (x_0,s_0,t_0), & n=0, \\ (x_0,s_0,t_0,a_0,x_1,s_1,t_1,a_1,\dots,x_n,s_n,t_n), & n \geq 1, \end{cases}$$

where $a_k \in \widehat{D}_k(x_k,s_k,t_k)$, $k=0,\dots,N-1$, and the set of such histories is denoted by $\widehat{\mathcal{H}}_n$. With $\widehat{\Pi}$ and $\widehat{\Pi}^M$ we denote the sets of history-dependent and Markov policies for the decision model with extended state space. We will write $\mathbb{E}_{nh_n}$ for a conditional expectation given $H_n^\pi = h_n,\ h_n \in \widehat{\mathcal{H}}_n$. The value of a policy $\pi \in \widehat{\Pi}$ at time $n=0,\dots,N$ is defined as

$$\begin{aligned} V_{N\pi}(h_N) &= g(s_N + t_N c_N(x_N)), \\ V_{n\pi}(h_n) &= \mathbb{E}_{nh_n}\left[g\left(s_n + t_n \left(\sum_{k=n}^{N-1} \beta^{k-n} c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi) + \beta^{N-n} c_N(X_N^\pi) \right)\right)\right], \end{aligned} \tag{6.5}$$

where $h_n \in \widehat{H}_n$. The corresponding value functions are

$$V_n(h_n) = \inf_{\pi \in \widehat{\Pi}} V_{n\pi}(h_n), \qquad h_n \in \widehat{\mathcal{H}}_n. \tag{6.6}$$

In the end, the quantity of interest is $V_0(x,0,1)$ which agrees with the infimal value of the original inner optimization problem (6.4). But how do we get an optimal policy for problem (6.4)? When starting in $(x_0,0,1) \in \widehat{E}$, a history $(x_0,a_0,x_1,a_1,\dots,x_N) \in \mathcal{H}_N$ of the original decision model uniquely determines the history $(x_0,s_0,t_0,a_0,x_1,s_1,t_1,a_1,\dots,x_N,s_N,t_N) \in \widehat{\mathcal{H}}_N$ of the decision model with extended state space through

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

Hence, for the initial state $(x_0,0,1) \in \widehat{E}$, a Markov policy $\pi = (d_0,\dots,d_{N-1}) \in \widehat{\Pi}^M$

with $d_n : \widehat{E} \to \mathcal{A}$, which will turn out to be optimal for (6.6), can be perceived as a history-dependent policy $\pi' = (d_0', \dots, d_{N-1}') \in \Pi$ of the original decision model, since we can find measurable functions $d_n' : \mathcal{H}_n \to A$ satisfying $d_n'(h_n) \in D_n(x_n)$ and

$$d_n'(x_0, a_0, x_1, \dots, x_n) = d_n\left(x_n, \sum_{k=0}^{n-1} \beta^k c_k(x_k, a_k, x_{k+1}), \beta^n\right).$$

Analogously, a history-dependent policy $\pi \in \widehat{\Pi}$ can be regarded as a history-dependent policy of the original decision model.

We can now proceed to deriving an iteration for the policy values (6.5).

Proposition 6.5. *Under Assumption 6.1 the value of a policy $\pi \in \widehat{\Pi}$ can be calculated recursively for $n = 0, \dots, N-1$ and $h_n \in \widehat{\mathcal{H}}_n$ as*

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

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 for time n

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

$$= \mathbb{E}\Big[V_{n+1\pi}\Big(h_n,\, d_n(h_n),\, T_n(x_n, d_n(h_n), Z_{n+1}),$$
$$s_n + t_n c_n(x_n, d_n(h_n), T_n(x_n, d_n(h_n), Z_{n+1})),\, \beta t\Big)\Big]. \qquad \square$$

Remark 6.6. 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 Expected Shortfall, one only has to consider the functions $g_q(x) = (x-q)^+,\ q \in \mathbb{R}$. 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).

Let us now consider specifically Markov policies $\pi \in \widehat{\Pi}^M$. The function space

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

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 $\mathbb{M}$. Note that all $v \in \mathbb{M}$ are non-negative and thus at least quasi-integrable.

Definition 6.7. For $v \in \mathbb{M}$ and a Markov decision rule $d : \widehat{\mathbb{E}} \to A$ we define

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

Note that the operators are monotone in v. Under a Markov policy $\pi = (d_0, \dots, d_{N-1}) \in \widehat{\Pi}$ the value iteration can be expressed with the operators. In order to distinguish from the history-depended case, we denote the policy values with J. Setting $J_{N\pi}(x,s,t) = g(s + tc_N(x)),\ (x,s,t) \in \widehat{E}$, we obtain for $n = 0, \dots, N-1$ and $(x,s,t) \in \widehat{E}$

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

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

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

The next result shows that V_n satisfies a Bellman equation and proves that an optimal policy exists and is Markov.

Theorem 6.8. *Let Assumption 6.1 be satisfied. Then, for $n = 0, \dots, N$ the value function V_n only depends on (x_n, s_n, t_n), i.e. $V_n(h_n) = J_n(x_n, s_n, t_n)$ for all $h_n \in \widehat{\mathcal{H}}_n$, lies in $\mathbb{M}$ and satisfies the Bellman equation*

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

Furthermore, for $n = 0, \dots, N-1$ there exist Markov decision rules $d_n^* : \widehat{E} \to A$ with $\mathcal{T}_{nd_n^*} J_{n+1} = \mathcal{T}_n J_{n+1}$ and every sequence of such minimizers constitutes an optimal policy $\pi = (d_0^*, \dots, d_{N-1}^*) \in \widehat{\Pi}^M$.

Proof. The proof is by backward induction. At time N we have $V_N(h_N) = J_N(x_N, s_N, t_N) = g(s_N + t_N c_N(x_N)),\ h_N \in \widehat{\mathcal{H}}_n$, which is

- lower semicontinuous by Lemma A.4 b) since g is increasing and continuous (as a convex function on $\mathbb{R}$) and c_N is lower semicontinuous,
- continuous and increasing in (s_N, t_N) since g is continuous and 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. in $\mathbb{M}$. Assuming the assertion holds at time $n+1$ we have at time n for $h_n \in \widehat{\mathcal{H}}_n$

$$\begin{aligned} V_n(h_n) &= \inf_{\pi \in \widehat{\Pi}} V_{n\pi}(h_n) \\ &= \inf_{\pi \in \widehat{\Pi}} \mathbb{E}\Big[V_{n+1\pi}\Big(h_n,\, d_n(h_n),\, \widehat{T}_n(x_n, s_n, t_n, d_n(h_n), Z_{n+1})\Big)\Big] \\ &\geq \inf_{\pi \in \widehat{\Pi}} \mathbb{E}\Big[V_{n+1}\Big(h_n,\, d_n(h_n),\, \widehat{T}_n(x_n, s_n, t_n, d_n(h_n), Z_{n+1})\Big)\Big] \end{aligned}$$

which equals by the induction hypothesis

$$= \inf_{\pi \in \widehat{\Pi}} \mathbb{E}\Big[J_{n+1}\Big(\widehat{T}_n(x_n, s_n, t_n, d_n(h_n), Z_{n+1})\Big)\Big].$$

Since the minimization here does not depend on the entire policy but only on $a_n = d_n(h_n)$, this equals

$$= \inf_{a_n \in D_n(x_n)} \mathbb{E}\Big[J_{n+1}\Big(\widehat{T}_n(x_n, s_n, t_n, a_n, Z_{n+1})\Big)\Big].$$

Here, objective and constraint depend on the history of the process only through x_n. Thus, given existence of a minimizing Markov decision rule $d_n^* : \widehat{E} \to A$, one ontains the identity

$$= \mathcal{T}_{nd_n^*} J_{n+1}(x_n, s_n, t_n). \tag{6.7}$$

Again by the induction hypothesis, there exists an optimal Markov policy $\pi^* \in \widehat{\Pi}^M$ such that

$$\begin{aligned}
&= \mathcal{T}_{nd_n^*} J_{n+1\pi^*}(x_n, s_n, t_n) \\
&= J_{n\pi^*}(x_n, s_n, t_n) \\
&\geq J_n(x_n, s_n, t_n) \\
&\geq V_n(h_n).
\end{aligned}$$

It remains to show the existence of a minimizing Markov decision rule d_n^* at (6.7) and that $J_n \in \mathbb{M}$. We want to apply Proposition A.25. The set-valued mapping $\widehat{E} \ni (x, s, t) \mapsto D_n(x)$ is compact-valued and upper semicontinuous. Next, we show that $\widehat{D}_n \ni (x, s, t, a) \mapsto L_n v(x, s, t, a)$ 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 $\widehat{D}_n$ with limit $(x^*, s^*, t^*, a^*) \in \widehat{D}_n$. The mapping

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

is lower semicontinuous for every $\Omega \in \Omega$ by Lemma A.4 a,b) applied simultaneously. Since $v \geq g \geq 0$, we can apply Fatou's Lemma B.1 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. With Proposition A.25 follows the existence of a minimizing decision rule d_n^* at (6.7) and the lower semicontinuity of $\mathcal{T}_n v$.

Now fix $(x, a) \in D_n$. By the monotonicity of expectation $(s, t) \mapsto L_n v(x, s, t, a)$ is increasing. Consequently, $(s, t) \mapsto \mathcal{T}_n v(x, s, t)$ is increasing by Lemma A.19. To see that this mapping is continuous it suffices to show upper semicontinuity or equivalently right continuity (Lemma A.6). Let $(s_k, t_k) \downarrow (s^*, t^*)$ as $k \to \infty$. Then the sequence

$$\Big\{v\Big(T_n(x, a, Z_{n+1}),\ s_k + t_k c_n(x, a, T_n(x, a, Z_{n+1})),\ \beta t_k\Big)\Big\}_{k\in\mathbb{N}}$$

is decreasing and it follows from monotone convergence that

$$\begin{aligned}
&\lim_{k\to\infty} L_n v(x, s_k, t_k, a) \\
&= \lim_{k\to\infty} \mathbb{E}\Big[v\Big(T_n(x, a, Z_{n+1}),\ s_k + t_k c_n(x, a, T_n(x, a, Z_{n+1})),\ \beta t_k\Big)\Big]
\end{aligned}$$

$$= \mathbb{E}\Big[\lim_{k\to\infty} v\Big(T_n(x,a,Z_{n+1}),\; s_k + t_k c_n(x,a,T_n(x,a,Z_{n+1})),\; \beta t_k\Big)\Big]$$
$$= L_n v(x, s^*, t^*, a),$$

where the last inequality is due to the continuity of v in the second and third argument. I.e. $(s,t) \mapsto L_n v(x,s,t,a)$ is continuous and especially upper semicontinuous. Hence, $(s,t) \mapsto \mathcal{T}_n v(x,s,t)$ is upper semicontinuous as an infimum of upper semicontinuous functions (Corollary A.3).

The inequality $\mathcal{T}_n v(x,s,t) \geq g(s),\ (x,s,t) \in \widehat{E}$, is obvious. Taken together, we have $\mathcal{T}_n v \in \mathbb{M}$ and the proof is complete. □

Remark 6.9. From Theorem 6.8 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).

6.1.2. Infinite Planning Horizon

In this section, we consider the inner optimization problem of the risk-sensitive total cost minimization under an infinite planning horizon. To reiterate, this approach 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 abstract cost model with no terminal cost, i.e. D, T, c do not depend on n, $c_N \equiv 0$, the disturbances are identically distributed and the discount factor β lies in $(0,1)$. Let Z be a representative of the disturbance distribution. The model with infinite planning horizon is derived as a limit of the one with finite horizon, i.e. the total discounted cost under a policy $\pi \in \Pi$ for initial state $x \in E$ is given by

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

As in Section 6.1.1, we assume that the one-stage cost $c(x,a,T(x,a,Z))$ is bounded below by a constant $\underline{c} \in \mathbb{R}$ for all $(x,a) \in D$. Due to the translation invariance of ρ_ϕ we have w.l.o.g. $\underline{c} = 0$ and $C_\infty^{\pi x} \geq 0,\ \pi \in \Pi$. Theorem B.3 guarantees that $C_\infty^{\pi x}$ is well-defined as an almost sure limit.

With Proposition 2.11, the initial optimization problem under an infinite planning

$$\inf_{\pi\in\Pi} \rho_\phi(C_\infty^{\pi x}) \tag{6.8}$$

can again be reformulated to

$$\begin{aligned}
\inf_{\pi\in\Pi} \rho_\phi\left(C_\infty^{\pi x}\right) &= \inf_{\pi\in\Pi}\inf_{g\in G}\left\{\mathbb{E}[g(C_\infty^{\pi x})] + \int_0^1 g^*(\phi(u))\,\mathrm{d}\,u\right\} \\
&= \inf_{g\in G}\inf_{\pi\in\Pi}\left\{\mathbb{E}[g(C_\infty^{\pi x})] + \int_0^1 g^*(\phi(u))\,\mathrm{d}\,u\right\} \\
&= \inf_{g\in G}\left\{\inf_{\pi\in\Pi}\mathbb{E}[g(C_\infty^{\pi x})] + \int_0^1 g^*(\phi(u))\,\mathrm{d}\,u\right\}, \qquad (6.9)
\end{aligned}$$

where G denotes the set of increasing convex functions $g : \mathbb{R} \to \mathbb{R}$. For fixed $g \in G$ we will refer to

$$\inf_{\pi\in\Pi} \mathbb{E}[g(C_\infty^{\pi x})] \qquad (6.10)$$

as *infinite horizon inner optimization problem.* The remarks in Section 6.1.1 regarding connections to the minimization of (rank-dependent) expected disutilities and corresponding certainty equivalents apply in the infinite horizon case as well. The following assumptions are made in this section.

Assumption 6.10. (i) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being continuous in (x, a) (case 1).

(ii) The one-stage cost $c(x, a, T(x, a, Z))$ is non-negative for all $(x, a) \in D$.

(iii) The family of random variables $\{C_\infty^{\pi x} : \ \pi \in \Pi,\ x \in E\}$ is uniformly integrable.

(iv) The spectrum ϕ is bounded, i.e. $\phi(1) < \infty$.

(v) The discount factor β lies in $(0, 1)$.

As in the finite horizon case, Assumption 6.10 (iii) can be equivalently characterized in terms of the increasing convex order. The proof is the same as for Lemma 6.2.

Lemma 6.11. *Let Assumption 6.10 (ii) be statisfied. Then, Assumption 6.10 (iii) is equivalent to the existence of a non-negative random variable $\bar{C} \in L^1_+$ on some probability space such that*

$$C_\infty^{\pi x} \leq_{icx} \bar{C}$$

for all policies $\pi \in \Pi$ and initial states $x \in E$.

Also the sufficient condition for Assumption 6.1 (iii) in Lemma 6.3 applies to Assumption 6.10 analogously.

Lemma 6.12. *If there exists a measurable function $\bar{c} : \mathcal{Z} \to \mathbb{R}_+$ such that*

$$c(x, a, T(x, a, z)) \leq \bar{c}(z), \qquad (x, a, z) \in D \times \mathcal{Z}$$

and $\bar{c}(Z) \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, then Assumption 6.10 (iii) is satisfied.

Proof. We have for all policies $\pi \in \Pi$ and initial states $x \in E$

$$c(X_n^\pi, d_n(H_n^\pi), X_{n+1}^\pi) = c(X_n^\pi, d_n(H_n^\pi), T(X_n^\pi, d_n(H_n^\pi), Z_{n+1})) \leq \bar{c}(Z_{n+1}).$$

It follows

$$C_\infty^{\pi x} = \sum_{k=0}^{\infty} \beta^k c(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi) \leq \sum_{k=0}^{\infty} \beta^k \bar{c}(Z_{k+1}) =: \bar{C}$$

Since $\bar{c}(Z)$ is non-negative and in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ it follows from Theorem B.3 that $\bar{C}$ is well-defined as an almost sure limit and

$$\mathbb{E}[\bar{C}] = \sum_{k=0}^{\infty} \beta^k \mathbb{E}[\bar{c}(Z_{k+1})] = \frac{\mathbb{E}[\bar{c}(Z)]}{1-\beta} < \infty,$$

i.e. $\bar{C} \in L^1_+(\Omega, \mathcal{A}, \mathbb{P})$. A common integrable majorant is clearly sufficient for uniform integrability. □

Due to the bounded spectrum, the function space G can again be reduced. The proof is the same as for Lemma 6.4. Recall that $\rho_\phi(\bar{C})$ is finite since $\bar{C} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Lemma 6.13. *Under Assumption 6.10 it is sufficient to consider functions $g \in G$ which are $\phi(1)$-Lipschitz and satisfy $0 \leq g(s) \leq \bar{g}(s),\ s \in \mathbb{R}$, where*

$$\bar{g}(s) = \phi(1)s^+ + \rho_\phi(\bar{C}).$$

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

We are now going to solve the infinite horizon inner optimization problem (6.10) for an arbitrary but fixed function $g \in \mathcal{G}$. Lemma 6.13 guarantees that the optimization problem is well-defined under Assumption 6.10. As under a finite planning horizon, we have for every initial state $x \in E$, policy $\pi \in \Pi$ and $g \in \mathcal{G}$

$$\mathbb{E}[g^-(C_\infty^\pi)] \geq 0 \qquad \text{and} \qquad \mathbb{E}[g^+(C_\infty^\pi)] \leq \phi(1)\mathbb{E}[\bar{C}] + \rho_\phi(\bar{C}) < \infty.$$

In order to obtain a value iteration, the state space is extended to $\widehat{E} = E \times \mathbb{R}_+ \times (0,1]$ as in Section 6.1.1. The action space A and the admissible state-action combinations D remain unchanged, i.e. $\widehat{D} = \{(x,s,t,a) \in \widehat{E} \times A : \ a \in D(x)\}$ and $\widehat{D}(x,s,t) = D(x),\ (x,s,t) \in \widehat{E}$. The transition function on the new state space is given by $\widehat{T} : \widehat{D} \times \mathcal{Z} \to \widehat{E}$,

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

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 $\pi = (d_1, d_2, \dots) \in \widehat{\Pi}^M$ due to Theorem 6.8. For the relevant initial state $(x_0, 0, 1) \in \widehat{E}$, a Markov policy $\pi \in \widehat{\Pi}^M$ can be perceived as a history-dependent policy of the original decision model, cf. Section 6.1.1. 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 $\pi = (d_0, d_1 \dots) \in \widehat{\Pi}^M$ up to a planning horizon $N \in \mathbb{N}$ now is

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

where $(X_n^\pi, \mathbf{s}_n^\pi, \mathbf{t}_n^\pi)_{n\in\mathbb{N}}$ is the extended decision process under policy $\pi \in \widehat{\Pi}^M$ with initial state $(x,s,t) \in \widehat{E}$. The change of indexing makes it necessary to write the value iteration in terms the *shifted policy* $\vec{\pi} = (d_1, d_2, \dots)$ corresponding to $\pi = (d_0, d_1, \dots) \in \widehat{\Pi}^M$:

$$\begin{aligned} J_{N\pi}(x,s,t) &= E\Big[J_{N-1\vec{\pi}}\Big(T(x, d_0(x,s,t), Z),\ s + tc(x, d_0(x,s,t), T(x, d_0(x,s,t), Z)),\ \beta t\Big)\Big] \\ &= \mathcal{T}_{d_0} J_{N-1\vec{\pi}}(x), \end{aligned} \tag{6.11}$$

$(x,s,t) \in \widehat{E}$. The value function for finite planning horizon $N \in \mathbb{N}$ is given by

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

and satisfies due to Theorem 6.8 the Bellman equation

$$J_N(x,s,t) = \mathcal{T} J_{N-1}(x,s,t) = \mathcal{T}^N 0(x,s,t), \qquad (x,s,t) \in \widehat{E}.$$

The value of a policy $\pi \in \widehat{\Pi}^M$ under an infinite planning horizon is defined as

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

$(x,s,t) \in \widehat{E}$. Note that $J_{\infty\pi}$ is well-defined due to Theorem B.3. This section's optimality criterion is

$$J_\infty(x,s,t) = \inf_{\pi\in\Pi^M} J_{\infty\pi}(x,s,t), \qquad (x,s,t) \in \widehat{E}. \tag{6.12}$$

Lemma 6.14. *Under Assumption 6.10 the sequences* $\{J_{N\pi}\}_{N\in\mathbb{N}}$, $\pi \in \widehat{\Pi}^M$, *and* $\{J_N\}_{N\in\mathbb{N}}$ *are increasing and pointwise convergent. It holds* $\lim_{N\to\infty} J_{N\pi} = J_{\infty\pi}$.

Proof. Since the one-stage cost is non-negative, $\{J_{N\pi}\}_{N\in\mathbb{N}}$ and hence $\{J_N\}_{N\in\mathbb{N}}$ are increasing. They converge by Lemma A.9 a). Also the sequence of random variables

$$\sum_{k=0}^{N-1}\beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi), \qquad n \in \mathbb{N},$$

is increasing and converges almost surely to $\sum_{k=0}^{\infty}\beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)$ by Theorem B.3. Thus, monotone convergence yields

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

where the second equality is by the Continuous Mapping Theorem. □

The yet unknown limit $J(x) = \lim_{N\to\infty} J_N(x),\ x \in E$, is referred to as *limit value function.* We introduce an upper bounding function $\bar{b} : \mathbb{R}_+ \times (0,1] \to \mathbb{R}$ given by

$$\bar{b}(s,t) = \mathbb{E}\left[g\left(s + t\frac{\bar{C}}{1-\beta}\right)\right].$$

Note that $\bar{b}$ is well-defined since for $g \in \mathcal{G}$ we have by Lemma 6.13

$$\mathbb{E}\left[g\left(s + t\frac{\bar{C}}{1-\beta}\right)\right] \le \phi(1)\left(s + t\frac{E[\bar{C}]}{1-\beta}\right) + \rho_\phi(\bar{C}) < \infty$$

for all $(s,t) \in \mathbb{R}_+ \times (0,1]$ as $\bar{C} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$. We will see that if $\bar{C}$ has a structure as in Lemma 6.12, it is not necessary to divide $\bar{C}$ by $1-\beta$ in the definition of $\bar{b}$. Finally note that $\bar{b}$ is indeed an upper bounding function and a lower one is not needed: It holds

$$g(s) \le J_{N\pi}(x,s,t) \le J_{\infty\pi}(x,s,t) \le \bar{b}(s,t) \tag{6.13}$$

for all $N \in \mathbb{N}_0,\ \pi \in \Pi^M$ and $(x,s,t) \in \widehat{E}$. The first inequality is by the non-negativity of the one-stage cost, the second one by Lemma 6.14 and the third one by Lemma 6.11 and the definition of the increasing convex order.

Theorem 6.15. *Let Assumption 6.10 be satisfied. Then it holds:*

a) The infinite horizon value function J_∞ is the unique fixed point of the Bellman operator $\mathcal{T}$ in $\mathbb{B} = \{v \in \mathbb{M} : v(x,s,t) \le \bar{b}(s,t)$ for all $(x,s,t) \in \widehat{E}\}$ and $J_\infty = J$. Moreover, we have $\mathcal{T}^N g \uparrow J_\infty$ and $\mathcal{T}^N \bar{b} \downarrow J_\infty$ as $N \to \infty$.

b) There exists a Markov decision rule d^* such that

$$\mathcal{T}_{d^*} J_\infty(x,s,t) = \mathcal{T} J_\infty(x,s,t), \qquad (x,s,t) \in \widehat{E}.$$

c) Each stationary policy $\pi^* = (d^*, d^*, \dots)$ induced by a Markov decision rule d^* as in part b) is optimal for optimization problem (6.12).

Proof. a) First, we show that $J_\infty = J$. Since the function $g : \mathbb{R} \to \mathbb{R}_+$ is convex, it is almost everywhere differentiable and in these points the derivative agrees with the everywhere existing right derivative g'_+. By Lemma 6.13 we have $g'_+ \leq \phi(1)$. Consequently,

$$\begin{aligned} g(s_1 + s_2) &= g(s_1) + \int_{s_1}^{s_1+s_2} g'_+(s)\,\mathrm{d}\,s \\ &= g(s_1) + \left(\frac{1}{s_2}\int_{s_1}^{s_1+s_2} g'_+(s)\,\mathrm{d}\,s\right) s_2 \\ &\leq g(s_1) + \phi(1)s_2 \end{aligned} \tag{6.14}$$

for all $s_1 \in \mathbb{R}$ and $s_2 > 0$. For $s_2 = 0$ the inequality holds trivially. It follows for every $N \in \mathbb{N}_0$, $(x, s, t) \in \widehat{E}$ and $\pi \in \widehat{\Pi}^M$

$$\begin{aligned} J_N(x,s,t) &\leq J_{N\pi}(x,s,t) \\ &\leq J_{\infty\pi}(x,s,t) \\ &= \mathbb{E}_{0x}\left[g\left(s + t\sum_{k=0}^{\infty}\beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right)\right] \\ &= \mathbb{E}_{0x}\left[g\left(s + t\sum_{k=0}^{N-1}\beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right.\right. \\ &\qquad\qquad \left.\left. + t\sum_{k=N}^{\infty}\beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right)\right] \\ &\leq \mathbb{E}_{0x}\left[g\left(s + t\sum_{k=0}^{N-1}\beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right)\right] \\ &\quad + \phi(1)t\beta^N\mathbb{E}_{0x}\left[\sum_{k=N}^{\infty}\beta^{k-N}c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right] \\ &= J_{N\pi}(x,s,t) + \phi(1)t\beta^N\mathbb{E}_{0x}\left[\sum_{k=N}^{\infty}\beta^{k-N}c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right] \\ &\leq J_{N\pi}(x,s,t) + \phi(1)\beta^N\mathbb{E}[\bar{C}]. \end{aligned} \tag{6.15}$$

The last inequality is true since $t \in (0, 1]$ and

$$\begin{aligned} &\mathbb{E}_{0x}\left[\sum_{k=N}^{\infty}\beta^{k-N}c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right] \\ &= \int \mathbb{E}_{Nh_N}\left[\sum_{k=N}^{\infty}\beta^{k-N}c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi)\right]\mathbb{Q}^\pi_{(x,s,t)}(\mathrm{d}\,h_N) \\ &= \int \mathbb{E}_{0(x_N,s_N,t_N)}\left[C_\infty^{\pi x_N}\right]\mathbb{Q}^\pi_{(x,s,t)}(\mathrm{d}\,h_N) \\ &\leq \mathbb{E}[\bar{C}], \end{aligned}$$

where the second equality is by the stationarity of the model and the inequality by

Lemma 6.11. Taking the infimum over $\pi \in \widehat{\Pi}^M$ in (6.15) and letting $N \to \infty$ yields

$$J(x,s,t) \leq J_\infty(x,s,t) \leq J(x,s,t), \qquad (x,s,t) \in \widehat{E}.$$

By Lemma 6.14, it follows $\mathcal{T}^N g \uparrow J_\infty$.

Next, we show that $\mathcal{T}^N \bar{b} \downarrow J_\infty$. Let (X_1, X_2) be a random vector and (X_1^c, X_2^c) be a comonotonic random vector with the same marginal distributions. Then $X_1 + X_2 \leq_{icx} X_1^c + X_2^c$, cf. Müller and Stoyan (2002, 8.3.4). Also note that $\bar{b}$ is increasing convex in s. Hence,

$$\begin{aligned}
\mathcal{T}\bar{b}(s,t) &= \inf_{a \in D(x)} E\Big[\bar{b}\Big(s + tc(x,a,T(x,a,Z)),\ \beta t\Big)\Big] \\
&\leq \inf_{a \in D(x)} E\Big[\bar{b}\Big(s + t\bar{C},\ \beta t\Big)\Big] \\
&= \iint g\left(s + tc_2 + \beta t \frac{c_1}{1-\beta}\right) \mathbb{P}^{\bar{C}}(\mathrm{d}\, c_1)\, \mathbb{P}^{\bar{C}}(\mathrm{d}\, c_2) \\
&\leq \iint g\left(s + tc_2 + \beta t \frac{c_1}{1-\beta}\right) \mathbb{P}^{(\bar{C}^c, \bar{C}^c)}(\mathrm{d}(c_1, c_2)) \\
&= \int g\left(s + t\frac{c}{1-\beta}\right) \mathbb{P}^{\bar{C}}(\mathrm{d}\, c) \\
&= \bar{b}(s,t)
\end{aligned}$$

where the first inequality is by Lemma 6.11. Now the monotonicity of the operator $\mathcal{T}$ implies that the sequence $\{\mathcal{T}^N \bar{b}\}_{N \in \mathbb{N}}$ is decreasing and convergent since 0 is a lower bound. Note that $\bar{b} \in \mathbb{M}$. Therefore, Theorem 6.8 yields that

$$\begin{aligned}
\mathcal{T}^N \bar{b}(x,s,t) &= \inf_{\pi \in \widehat{\Pi}^M} \mathbb{E}_{0x}\left[\bar{b}\left(s + t\sum_{k=0}^{N-1} \beta^k c(X_k^\pi, d_k(X_k^\pi, \mathbf{s}_k^\pi, \mathbf{t}_k^\pi), X_{k+1}^\pi),\ \beta^N t\right)\right] \\
&= \inf_{\pi \in \widehat{\Pi}^M} \mathbb{E}_{0x}[\bar{b}(s + tC_N^{\pi x},\ \beta^N t)].
\end{aligned}$$

Consequently, we have by Lemma A.31 a), Tonelli's Theorem B.2 and equation (6.14)

$$\begin{aligned}
0 &\leq \mathcal{T}^N \bar{b}(x,s,t) - J_N(x,s,t) \\
&= \mathcal{T}^N \bar{b}(x,s,t) - \mathcal{T}^N g(x,s,t) \\
&\leq \sup_{\pi \in \widehat{\Pi}^M} \mathbb{E}_{0x}\Big[\bar{b}(s + tC_N^{\pi x},\ \beta^N t) - g(s + tC_N^{\pi x})\Big] \\
&= \sup_{\pi \in \widehat{\Pi}^M} \iint g(s + tc_2 + t\beta^N \frac{c_1}{1-\beta}) - g(s + tc_2)\, \mathbb{P}^{\bar{C}}(\mathrm{d}\, c_1)\, \mathbb{P}^{C_N^{\pi x}}(\mathrm{d}\, c_2) \\
&\leq \phi(1) \frac{t\beta^N}{1-\beta} \mathbb{E}[\bar{C}] \to 0 \qquad \text{as } N \to \infty.
\end{aligned}$$

Hence, it holds $\lim_{N \to \infty} \mathcal{T}^N \bar{b} = J_\infty$.

Finally, we show that J_∞ is the unique fixed point of $\mathcal{T}$ in $\mathbb{B}$. By Theorem 6.8 and

(6.13) we have $J_N \in \mathbb{B}$ for all $N \in \mathbb{N}$. Hence, $g(s) \le J_\infty(x,s,t) \le \bar{b}(s,t),\ (x,s,t) \in \widehat{E}$. Moreover, J_∞ is lower semicontinuous by Lemma A.9 b). We already noted that $\bar{b} \in \mathbb{M}$, i.e. the function $(s,t) \mapsto \bar{b}(s,t)$ is upper semicontinuous and increasing. By monotone convergence $(s,t) \mapsto \mathbb{E}_{0x}[\bar{b}(s + tC_N^{\pi x},\ \beta^N t)]$ is upper semicontinuous, too, and

$$(s,t) \mapsto \mathcal{T}^N \bar{b} = \inf_{\pi \in \widehat{\Pi}^M} \mathbb{E}_{0x}[\bar{b}(s + tC_N^{\pi x},\ \beta^N t)]$$

remains so as an infimum of upper semicontinuous functions. Since $\mathcal{T}^N \bar{b} \downarrow J_\infty$, Lemma A.9 b) (mutatis mutandis) yields that $(s,t) \mapsto J_\infty(x,s,t)$ is upper semicontinuous for all $x \in E$. To sum up, we have $J_\infty \in \mathbb{B}$.
As the sequence $\{J_N\}_{N\in\mathbb{N}}$ is increasing, it holds $J_N \le J_\infty$ for all $N \in \mathbb{N}$. Now the monotonicity of $\mathcal{T}$ yields $J_{N+1} = \mathcal{T} J_N \le \mathcal{T} J_\infty$ and implying $J_\infty \le \mathcal{T} J_\infty$ as $N \to \infty$. Conversely, by taking the infimum over $\pi \in \Pi^M$ in (6.15) and then applying $\mathcal{T}$ we get

$$J_{N+1} + \phi(1)\beta^N \mathbb{E}[\bar{C}] = \mathcal{T}\Big(J_N + \phi(1)\beta^N \mathbb{E}[\bar{C}]\Big) \ge \mathcal{T} J_\infty$$

implying $J_\infty \ge \mathcal{T} J_\infty$ as $N \to \infty$. It remains to verify that the fixed point is unique. Assume $v \in \mathbb{B}$ is a further one. Then $g(s) \le v(x,s,t) \le \bar{b}(s,t),\ (x,s,t) \in \widehat{E}$, and by the monotonicity of $\mathcal{T}$

$$\mathcal{T}^N g \le T^N v = v = \mathcal{T}^N v \le \mathcal{T}^N \bar{b}.$$

Letting $N \to \infty$ yields $J_\infty \le v \le J_\infty$.

b) Since $J_\infty \in \mathbb{B} \subseteq \mathbb{M}$, the existence of a minimizing Markov decision rule follows from Theorem 6.8.

c) It holds $J_\infty(x,s,t) \ge g(s),\ (x,s,t) \in \widehat{E}$, since $J_\infty \in \mathbb{M}$. Consequently, we have

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

i.e. π^* is optimal. The first equality is by parts a) and b), the inequality thereafter by the monotonicity of the operator $\mathcal{T}_{d^*}$, the second equality by the value iteration (6.11) and the third one by Lemma 6.14. □

Note that inequality (6.14) differs from the standard inequality for convex functions $g(s_1 + s_2) \ge g(s_1) + g'_+(s_1)s_2,\ s_1, s_2 \in \mathbb{R}$, only by a parallel translation of the tangent.

6.1.3. Real Line as State Space

As for the distributionally robust and the risk-sensitive recursive cost minimization of Chapters 4 and 5, the continuity assumption on the transition functions can be relaxed to semicontinuity if the state space is the real line and the transition and one-stage cost function satisfy some form of monotonicity. For some applications as e.g. in Section 6.3

this relaxation is relevant. To ease the notational burden, we consider the stationary model with no terminal cost under both finite and infinite horizon in this section. The results can be transferred to a non-stationary setting by mere notational changes if the planning horizon is finite.

Assumption 6.16. (i) The original state space is the real line $E = \mathbb{R}$.

(ii) The model data has the Continuity and Compactness Properties 3.1 with the transition function T being lower semicontinuous in (x, a) (case 2).

(iii) The model data has the following monotonicity properties:

  (iii a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is decreasing.

  (iii b) The transition function T is increasing in x.

  (iii c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is increasing for all (a, z).

(iv) Assumptions 6.10 (ii) to (v) hold.

Requiring that the one-stage cost function c is increasing both in x and x' is sufficient for Assumption 6.16 (iii c) to hold since the transition function is increasing in x. Besides, if c is increasing in x', it is sufficient for Continuity and Compactness Assumption 3.1 (iii) that c is lower semicontinuous due to Lemma A.4 b).

The question is, how replacing Assumption 6.10 (i) by Assumption 6.16 (i) to (iii) affects the validity of all previous results. 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 6.8 and 6.15. All other statements are unaffected.

Proposition 6.17. *The assertions of Theorems 6.8 and 6.15 hold under Assumption 6.16, too. Moreover, the value functions J_n and J_∞ are increasing. The set of potential value functions can therefore be replaced by*

$$\begin{aligned}\mathbb{B} = \{v : \widehat{E} \to \mathbb{R} \mid\ & v \textit{ is lower semicontinuous and increasing,} \\ & v(x, \cdot, \cdot) \textit{ is continuous for all } x \in \mathbb{R}, \\ & g(s) \le v(x, s, t) \le \bar{b}(s, t) \textit{ for } (x, s, t) \in \widehat{E}\}.\end{aligned}$$

Proof. In Theorem 6.8, the continuity of T is used to show that $\widehat{D} \ni (x, s, t, a) \mapsto Lv(x, s, t, a)$ is lower semicontinuous for every $v \in \mathbb{B}$. Due to the monotonicity assumptions, the mapping

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

is lower semicontinuous by Lemma A.4 b). Now, the lower semicontinuity of $\widehat{D} \ni (x, s, t, a) \mapsto Lv(x, s, t, a)$ and the existence of a minimizing decision rule follow as in the proof of Theorem 6.8. The fact that $\mathcal{T}v$ is increasing for every $v \in \mathbb{B}$ follows from Lemma A.19.

In Theorem 6.15, the continuity of T is only used indirectly through Theorem 6.8. One only has to note that $J_\infty \in \mathbb{B}$ since the pointwise limit of increasing functions remains increasing. □

The monotonicity requirements in Assumption 6.16 (iii) are only one option. The following alternative is relevant i.a. for the dynamic reinsurance models introduced in Section 3.2.

Corollary 6.18. *Assumptions 6.16 (ii) and (iii) can be replaced by*

(ii') The model data has the Continuity and Compactness Properties 3.1 with the transition function T being upper semicontinuous (case 3).

(iii') The model data has the following monotonicity properties:

- *(iii' a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x)$ is increasing.*
- *(iii' b) The transition function T is increasing in x.*
- *(iii' c) The function $\mathbb{R} \ni x \mapsto c(x, a, T(x, a, z))$ is decreasing for all (a, z).*

Then, the assertions of Theorems 6.8 and 6.15 still hold. Moreover, the value functions J_n and J_∞ are decreasing in x and increasing in (s, t). The set of potential value functions is therefore

$$\begin{aligned}\mathbb{B} = \{v : \widehat{E} \to \mathbb{R} \mid\ & v \text{ is lower semicontinuous,}\\ & v(\cdot, s, t) \text{ is decreasing for all } (s, t) \in \mathbb{R}_+ \times (0, 1],\\ & v(x, \cdot, \cdot) \text{ is continuous and increasing for all } x \in \mathbb{R},\\ & g(s) \le v(x, s, t) \le \bar{b}(s, t) \text{ for } (x, s, t) \in \widehat{E}\}.\end{aligned}$$

Proof. One argues analogously to the proof of Proposition 6.17. In order to show that $\widehat{D} \ni (x, s, t, a) \mapsto Lv(x, s, t, a)$ is lower semicontinuous for every $v \in \mathbb{B}$ one uses Remark A.5 for proving that the mapping

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

is lower semicontinuous. □

Requiring that the one-stage cost function c is decreasing both in x and x' is sufficient for (iii' c) to hold since the transition function is increasing in x. Besides, if c is decreasing in x', it is sufficient for Continuity and Compactness Assumption 3.1 (iii) that c is lower semicontinuous due to Remark A.5.

The monotonicity properties of Assumption 6.16 (iii) can be used to construct a convex model. Due to the state space extension, Proposition 2.4.18 of Bäuerle and Rieder (2011) cannot be applied directly as in Lemma 4.23 but has to be slightly modified.

Lemma 6.19. *Let Assumption 6.16 be satisfied, A be a subset of a vector space, the admissible state-action-combinations D be a convex set, the transition function T be convex in (x, a) and the composition $D \ni (x, a) \mapsto c(x, a, T(x, a, z))$ 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 $t \in (0, 1]$.*

Proof. We prove by backward induction that J_n is convex in (x,s) for $n = 0, \dots, N$. Then J_∞ is convex as a pointwise limit of convex functions (after switching to forward indexing).

For $n = N$ we know that $J_N(x,s,t) = g(s)$ is convex in (x,s). Now assume that J_{n+1} is convex in (x,s). Recall that J_{n+1} increasing by Proposition 6.17. Hence, for every $\omega \in \Omega$ and $t \in (0,1]$ the function

$$(x,s,a) \mapsto J_{n+1}\Big(T(x,a,Z(\omega)),\ s + tc(x,a,T(x,a,Z(\omega))),\ \beta t\Big)$$

is convex as a composition of an increasing convex with a convex function. By the linearity of expectation $(x,s,a) \mapsto LJ_{n+1}(x,s,t,a)$ is convex, too, for every $t \in (0,1]$. Fix $t \in (0,1]$ and let $(x_1,s_1),(x_2,s_2) \in \mathbb{R} \times \mathbb{R}_+$ and $\lambda \in [0,1]$. By Theorem 6.8 there exist $a_i \in D(x_i)$ such that $LJ_{n+1}(x_i,s_i,t,a_i) = \mathcal{T} J_{n+1}(x_i,s_i,t)$, $i = 1,2$. The convexity of D implies $\lambda a_1 + (1-\lambda)a_2 \in D(\lambda x_1 + (1-\lambda)x_2)$. Hence, we have

$$\begin{aligned}
&\mathcal{T} J_{n+1}\Big(\lambda x_1 + (1-\lambda)x_2,\ \lambda s_1 + (1-\lambda)s_2,\ t\Big) \\
&= \inf_{a \in D(\lambda x_1 + (1-\lambda)x_2)} LJ_{n+1}\Big(\lambda x_1 + (1-\lambda)x_2,\ \lambda s_1 + (1-\lambda)s_2,\ t,\ a\Big) \\
&\le LJ_{n+1}\Big(\lambda x_1 + (1-\lambda)x_2,\ \lambda s_1 + (1-\lambda)s_2,\ t,\ \lambda a_1 + (1-\lambda)a_2\Big) \\
&\le \lambda LJ_{n+1}(x_1,s_1,t,a_1) + (1-\lambda) LJ_{n+1}(x_2,s_2,t,a_2) \\
&= \lambda \mathcal{T} J_{n+1}(x_1,s_1,t) + (1-\lambda)\mathcal{T} J_{n+1}(x_2,s_2,t).
\end{aligned}$$

□

If c is increasing in x', it is sufficient to require that c and T are convex in (x,a). In the decreasing setting of Corollary 6.18, one needs a concave transition function in order to obtain a convex model.

Corollary 6.20. *Let the assumptions of Corollary 6.18 be satisfied, A be a subset of a vector space, the admissible state-action-combinations D be a convex set, the transition function T be concave in (x,a) and the composition $D \ni (x,a) \mapsto c(x,a,T(x,a,z))$ 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 $t \in (0,1]$.*

6.2. Outer Problem

In this Section, we discuss the solution of the outer optimization problem (6.3) for a finite and (6.9) for an infinite planning horizon. Given a solution of the respective inner problem for every $g \in \mathcal{G}$, the two outer problems are essentially the same and therefore treated together. For a fixed policy $\pi \in \widehat{\Pi}^M$ the optimal solution of the outer problem is already given by Proposition 2.11 as

$$g_{\phi, C_N^{\pi x}}(s) = \int_0^1 F^{-1}_{C_N^{\pi x}}(\alpha) + \frac{1}{1-\alpha}\Big(s - F^{-1}_{C_N^{\pi x}}(\alpha)\Big)^+ \mu(\mathrm{d}\,\alpha), \qquad s \in \mathbb{R},\ N \in \mathbb{N} \cup \{\infty\}.$$

However, we solved the inner problem for arbitrary but fixed $g \in \mathcal{G}$. Hence, the optimal policy depends on g and Proposition 2.11 is not helpful. First, we consider the existence of a solution to the outer problem in Section 6.2.1 and then its algorithmic approximation in Section 6.2.2.

6.2.1. Existence

As a first step in ensuring the existence of a solution of the outer problem, 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 6.21. *$(\mathcal{G}, m)$ 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 A.32. 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. Non-negativity and the pointwise upper bound $\bar{g}$ imply that assumption (i) of Theorem A.32 is satisfied and the common Lipschitz constant that assumption (ii) holds. □

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

Lemma 6.22. *Let Assumption 6.1 be satisfied. Then the functional $\mathcal{G} \times \widehat{E} \ni (g, x, s, t) \mapsto J_n^g(x, s, t)$ 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

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

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

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

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

$$\begin{aligned}\liminf_{k\to\infty} g_k(s_k + t_k c_N(x_k)) &\geq \lim_{k\to\infty} g_k(s_k + t_k c_k)\\ &= g(s + t\hat{c})\\ &\geq g(s + t c_N(x)).\end{aligned}$$

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

$$\begin{aligned}\liminf_{k\to\infty} g_k(s_k + t_k c_N(x_k)) &\geq \liminf_{k\to\infty} g_k(s_k + t_k c_k)\\ &\geq \lim_{k\to\infty} g_k(s_k + t_k c_N(x))\\ &= g(s + t c_N(x)).\end{aligned}$$

Now assume the assertion holds for $n+1$. By Theorem 6.8 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 and Lemma A.4. Hence, if $(g_k,x_k,s_k,t_k) \to (g,x,s,t)$, Fatou's Lemma B.1 and the monotonicity of expectation yield

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

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 A.25. □

If there is an almost sure integrable upper bound for the integrand (e.g. under the conditions of Lemma 6.3), the value functions are continuous in g by dominated convergence and Lemma A.3 b) additionally to the joint lower semicontinuity.

Lemma 6.23. *Let Assumption 6.10 be satisfied. Then the functional* $\mathcal{G} \times \widehat{E} \ni (g, x, s, t) \mapsto J_\infty^g(x, s, t)$ *is lower semicontinuous for all* $(x, s, t) \in \widehat{E}$.

Proof. Under an infinite planning horizon we consider a stationary model and use forward indexing for the value functions J_N^g. They are lower semicontinuous in (g, x, s, t) by Lemma 6.22. Note that the induction basis holds especially for $c_N \equiv 0$. Since $J_N^g \uparrow J_\infty^g$ as $N \to \infty$ by Lemma 6.14 and Theorem 6.15, the assertion follows from Lemma A.9 b). ☐

Under the conditions of Lemma 6.12, the infinite horizon value function J_∞^g is continuous in g, too. This follows by applying Lemma A.9 b) mutatis mutandis to the decreasing sequence of upper semicontinuous functions $g \mapsto \mathcal{T}^N \bar{b}$ which converges to J_∞^g as $N \to \infty$ by Theorem 6.15.

For initial state $x \in E$ and finite planning horizon $N \in \mathbb{N}$ the outer problem (6.3) is given by $\inf_{g \in \mathcal{G}} J_0^g(x, 0, 1) + \int_0^1 g^*(\phi(u)) \,\mathrm{d}\, u$ and for infinite planning horizon the outer problem (6.9) is given by $\inf_{g \in \mathcal{G}} J_\infty^g(x, 0, 1) + \int_0^1 g^*(\phi(u)) \,\mathrm{d}\, u$. In the following, we will only use the semicontinuity of the value functions in g. Hence, we write

$$\inf_{g \in \mathcal{G}} J(g) + \int_0^1 g^*(\phi(u)) \,\mathrm{d}\, u \tag{6.16}$$

for a generic outer problem and suppress initial state and planning horizon.

Theorem 6.24. *Under Assumption 6.1 or 6.10, respectively, there exists a solution for the the outer optimization problem* (6.16)*.*

Proof. We want to apply Weierstraß' extrem value Theorem A.7. In view of Lemmata 6.21, 6.22 and 6.23 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 $u \in [0, 1]$

$$\begin{aligned}
\liminf_{k \to \infty} g_k^*(\phi(u)) &= \lim_{k \to \infty} \inf_{\ell \geq k} g_\ell^*(\phi(u)) \\
&= \lim_{k \to \infty} \inf_{\ell \geq k} \sup_{s \in \mathbb{R}} \{\phi(u)s - g_\ell(s)\} \\
&\geq \lim_{k \to \infty} \sup_{s \in \mathbb{R}} \inf_{\ell \geq k} \{\phi(u)s - g_\ell(s)\} \\
&= \sup_{s \in \mathbb{R}} \lim_{k \to \infty} \inf_{\ell \geq k} \{\phi(u)s - g_\ell(s)\} \\
&= \sup_{s \in \mathbb{R}} \{\phi(u)s - \limsup_{k \to \infty} g_k(s)\} \\
&= \sup_{s \in \mathbb{R}} \{\phi(u)s - g(s)\} \\
&= g^*(\phi(u)).
\end{aligned} \tag{6.17}$$

The inequality holds generally for the interchange of infimum and supremum, the equality thereafter by Lemma A.9 c) 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

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

Now, Fatou's Lemma B.1 and (6.17) yield 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. □

6.2.2. Numerical Approximation

As we know now that a solution to the outer optimization problem (6.16) exists, this section aims 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 (6.16) our method coincides with the *Fast Legendre-Fenchel Transform (FLT)* algorithm studied i.a. by Corrias (1996).

For unbounded cost $C_N^{\pi x}$ with $N \in \mathbb{N} \cup \{\infty\},\ \pi \in \Pi,\ x \in E$, the functions $g \in \mathcal{G}$ would have to be approximated on the whole non-negative real line. This is numerically not feasible.

Assumption 6.25. If $N \in \mathbb{N}$, we require additionally to Assumption 6.1 that the conditions of Lemma 6.3 are satisfied with constant $\bar{c}$. If $N = \infty$, we require that additionally to Assumption 6.10 the conditions of Lemma 6.12 are satisfied with constant $\bar{c}$.

Consequently, it holds $0 \leq C_N^{\pi x} \leq \hat{c}$ for all $N \in \mathbb{N} \cup \{\infty\},\ \pi \in \Pi$ and $x \in E$, where we define

$$\hat{c} = \begin{cases} \sum_{k=0}^N \beta^k \bar{c} & \text{for finite planning horizon } N \in \mathbb{N}, \\ \frac{\bar{c}}{1-\beta} & \text{for infinite planning horizon } N = \infty. \end{cases}$$

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 $\bar{\mathbb{R}}$-valued function defined by $g^*(y) = \sup_{s\in\mathbb{R}}\{sy - g(s)\},\ y \in \mathbb{R}$.

Lemma 6.26. *a) Under Assumption 6.25, a minimizer of the outer optimization problem* (6.16) *lies in*

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

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

Proof. a) Fix $\pi \in \Pi, x \in E$ and set $C = C_N^{\pi x}$ to simplify the notation. We know from the proof of Proposition 2.11 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 2.9. Clearly, it is sufficient to consider functions $g \in G$ which are optimal for at least one $C = C_N^{\pi x}$. Since $0 \le C \le \hat{c}$ we have $0 \le F_C^{-1}(\alpha) \le \hat{c}$. 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. the proof of Proposition 2.11, 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 $[0, \hat{c}]$. □

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 $[0, \hat{c}]$. For the value iteration in Lemma 6.5 and equation (6.11) it may be necessary to evaluate g in some $s > \hat{c}$, but here the function is determined as a linear continuation with slope $\phi(1)$. On the interval $I = [0, \hat{c}]$, the metric of compact convergence reduces to the supremum norm $\|\cdot\|_\infty$. 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 (6.16) to $\widehat{\mathcal{G}}_m$ it is convenient to define for $g \in \widehat{\mathcal{G}}$

$$\begin{aligned} K_m(g) &= J(p_m(g)) + \int p_m(g)^*(\phi(u))\,\mathrm{d}\,u, \\ K(g) &= J(g) + \int g^*(\phi(u))\,\mathrm{d}\,u. \end{aligned}$$

Proposition 6.27. *It holds*

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

Proof. The first inequality follows from Lemma A.31 b) and it remains to prove the second. We have for $N \in \mathbb{N} \cup \{\infty\}$, $x \in E$ and $g \in \widehat{\mathcal{G}}$

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

Also by Lemma A.31 b) it holds for $y \in [0, \phi(1)]$

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

Finally, the assertion follows with

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

□

The proposition shows that the infimum of K_m converges to the one of K. The error of restricting the outer problem (6.16) 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 $s_1, \dots, s_m$. Hence, we can identify $\widehat{\mathcal{G}}_m$ with the compact set

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

Note that due to translation invariance of ρ_ϕ it holds under Assumption 6.25 for $g \in \widehat{\mathcal{G}}$ that $g(0) \leq \bar{g}(0) = \rho(\bar{C}) = \rho(\hat{c}) = \hat{c}$. Thus, the outer problem (6.16) restricted to $\widehat{\mathcal{G}}_m$ becomes finite dimensional:

$$\inf_{y \in \Gamma_m} J(g_y) + \int_0^1 g_y^*(\phi(u)), \tag{6.18}$$

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

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

How to evaluate $J(\cdot)$ in $g_y,\ y \in \Gamma_m$, has been discussed in Sections 6.1.1 and 6.1.2. The next Lemma simplifies the evaluation of the second summand of the objective function (6.18) to calculating the integrals

$$\int_{u_k}^{u_{k+1}} \phi(u) \,\mathrm{d}\, u, \qquad k = 0, \dots, m,$$

where $u_0 = 0$, $u_k = \phi^{-1}\left(\frac{y_{k+1}-y_k}{s_{k+1}-s_k}\right)$, $k = 1, \dots, m-1$ and $u_m = \phi(1)$.

Lemma 6.28. *The convex conjugate of $g_y,\ y \in \Gamma_m$, in $\xi \in [0, \phi(1)]$ is given by*

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

Proof. By Lemma 6.26 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

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

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

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

The last equality holds, since $c_1 \leq \dots \leq c_{m-1}$ and $c_\ell \leq \xi \leq c_{\ell+1}$ is equivalent to

$\xi s_\ell - y_\ell \leq \xi s_{\ell+1} - y_{\ell+1} \geq \xi s_{\ell+2} - y_{\ell+2}$. 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 $c_1 \leq \dots \leq c_{m-1}$, this implies the last equality. The third case $c_{m-1} < \xi$ is analogous. □

The results of this section can be summarized in the following schematic algorithm.

Algorithm: Outer problem

Data: Markov Decision Model

Result: Optimal policy π^*, minimal risk-sensitive cost $\rho_\phi(C_N^{\pi^* x})$

1. Select an approximation error $\epsilon > 0$ and set $m = \left\lceil \frac{2\phi(1)\hat{c}}{\epsilon} \right\rceil + 1$.
2. Solve (6.18) with an algorithm for global optimization.
   if $N \in \mathbb{N}$ **then** For each evaluation of $J(\cdot)$ solve the inner problem (6.4) with Theorem 6.8.
   if $N = \infty$ **then** For each evaluation of $J(\cdot)$ solve the inner problem (6.10) with Theorem 6.15.

6.3. Cost of Capital Minimization of an Insurance Company

In Section 5.5, a dynamic extension in discrete time of the classical static reinsurance problem (5.15) has been developed by applying the risk-sensitive recursive optimality criterion to the dynamic reinsurance model of Section 3.2.1. In this section, we construct an alternative dynamic extension based on the risk-sensitive total cost criterion. The aim is to choose the reinsurance treaties such that the cost of capital for the total discounted loss

$$\inf_{\pi \in \Pi} \quad r_{\text{CoC}} \cdot \rho_\phi \left(\sum_{k=0}^{N-1} \beta^k \Big(d_k(H_k^\pi)(Y_{k+1}) + \pi_R(d_k(H_k^\pi)) - Z_{k+1} \Big) \right) \tag{6.19}$$

is minimized under a spectral risk measure ρ_ϕ with bounded spectrum. 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. Recall that in the dynamic reinsurance model the decision process describes the development of the insurer's surplus, i.e. the transition function is given by $T(x, f, y, z) = x - f(y) - \pi_R(f) + z$, where $f \in \mathcal{F}$ is the reinsurance contract, y the claims arriving at the end of the period, $\pi_R(f)$ the cost of reinsurance and z the premium income at the beginning of the next period. Here, the one-stage cost function is given by the incremental loss $c(x, f, x') = x - x'$. As discussed in Example 5.28, this is the natural choice for a total loss optimization criterion. For $\beta = 1$ we have

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

i.e. due to translation invariance of spectral risk measures the objective reduces to minimizing the capital requirement for the loss at the planing horizon $-X_N^\pi$. This is reminiscent of the static reinsurance problem (5.15), however here the loss distribution at the planing horizon can be controlled by interim action.

Throughout the chapter, we have required that the one-stage cost $c(x, f, T(x, f, Y, Z)) = f(Y) + \pi_R(f) - Z$ is non-negative. As $f(Y)$ and $\pi_R(f)$ are non-negative for all $f \in \mathcal{F}$ 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 $Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P})$ 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 $\hat{c}(x, f, x') = x - x' + \operatorname{ess\,sup}(Z)$. The economic interpretation is that the one-stage cost

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

now depends on the deviation from the maximal possible income instead of the actual income. However, note that this is not a change of the disturbance and the transition function still depends on the actual premium income, i.e. the current state still equals the current surplus. For brevity we write $\hat{z} = \operatorname{ess\,sup}(Z)$.

As in (6.3) we separate an inner and outer reinsurance problem. For a structural analysis we focus on the inner optimization problem

$$\inf_{\pi \in \Pi} \mathbb{E}\left[g\left(\sum_{k=0}^{N-1} \beta^k \Big(d_k(H_k^\pi)(Y_{k+1}) + \pi_R(d_k(H_k^\pi)) + \hat{z} - Z_{k+1} \Big) \right) \right] \tag{6.20}$$

with arbitrary $g \in \mathcal{G}$, cf. Lemma 6.4. On the extended state space $\widehat{E} = \mathbb{R} \times \mathbb{R}_+ \times (0, 1]$, the value of a policy $\pi \in \widehat{\Pi}$ is defined as

$$V_{N\pi}(h_N) = g(s_N),$$
$$V_{n\pi}(h_n) = \mathbb{E}_{nh_n}\left[g\left(s_n + t_n \sum_{k=n}^{N-1} \beta^{k-n} \Big(d_k(H_k^\pi)(Y_{k+1}) + \pi_R(d_k(H_k^\pi)) + \hat{z} - Z_{k+1} \Big) \right) \right],$$

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

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

Due to the real state space we want to apply Corollary 6.18 for solving the optimization problem. Let us verify the assumptions. The numbering is as in the corollary.

(i) The (original) state space is the real line $E = \mathbb{R}$.

(ii') The Continuity and Compactness Properties 3.1 with upper semicontinuous transition function have been verified in Section 3.2.1.

(iii') Monotonicity properties:

(iii' a) The set-valued mapping $\mathbb{R} \ni x \mapsto D(x) = \{f \in \mathcal{F} : \pi_R(f) \le x^+\}$ is increasing.

(iii' b) The transition function $T : \mathbb{R} \times \mathcal{F} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R},\ T(x, f, y, z) = x - f(y) - \pi_R(f) + z$ is increasing in x.

(iii' c) The composition $\mathbb{R} \ni x \mapsto \hat{c}(x, f, T(x, f, y, z)) = f(y) + \pi_R(f) + \hat{z} - z$ is independent of x and especially decreasing for all (f, y, z).

(iv) The modified one-stage cost $\hat{c}(x, f, T(x, f, Y, Z)) = f(Y) + \pi_R(f) + \hat{z} - Z$ is non-negative.

(v) It holds $0 \le f \le \mathrm{id}_{\mathbb{R}_+}$ for all $f \in \mathcal{F}$ and $\pi_R(f) = \pi_R(Y - f(Y)) \le \pi_R(Y)$ by the monotonicity of π_R. Thus,

$$\hat{c}(x, f, T(x, f, Y, Z)) = f(Y) + \pi_R(f) + \hat{z} - Z \le Y + \pi_R(Y) + \hat{z}.$$

Consequently, the conditions of Lemma 6.12 are satisfied with $\bar{c}(y) = y + \pi_R(Y) + \hat{z}$ since $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

(vi) The spectrum ϕ is bounded by assumption.

(vii) At least for an infinite planning horizon we require $\beta < 1$.

Hence, Corollary 6.18 yields that it is sufficient to minimize over all Markov policies, the value functions are in $\mathbb{B}$ and satisfy the Bellman equation

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

for $(x, s, t) \in \widehat{E}$ and $n = 0, \dots, N-1$. Moreover, there exists a Markov Decision rule $d_n^* : \widehat{E} \to \mathcal{F}$ minimizing J_{n+1} and every sequence $\pi = (d_0^*, \dots, d_{N-1}^*) \in \widehat{\Pi}^M$ of such minimizers is a solution to (6.20).

If the planning horizon is infinite and $\beta < 1$, we use forward indexation for the finite horizon value functions J_n. The infinite horizon value function J_∞ is the pointwise limit of the sequence $\{J_n\}_{n \in \mathbb{N}}$ and also characterized as the unique fixed point of the Bellman operator $\mathcal{T}$ in $\mathbb{B}$. Every minimizer d^* of J_∞ induces a stationary optimal policy $\pi^* = (d^*, d^*, \dots) \in \widehat{\Pi}^S$.

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 (6.19). 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. The latter is a necessary assumption to obtain a convex model since even for convex π_R, the set of admissible state-action combinations in case of a budget constraint $D = \{(x, f) \in$

$\mathbb{R} \times \mathcal{F} : \ \pi_R(f) \leq x^+\}$ is not convex. Moreover note that the absence of a budget constraint does not make the problem myopic as it is the case for the risk-sensitive recursive optimality criterion (see Remark 5.26) since the objective here does not become time-separable.

If π_R is convex and there is no budget constraint, we have indeed a convex model: D is trivially convex, the transition function $T(x, f, y, z) = x - f(y) - \pi_R(f) + z$ is concave in (x, f) as a sum of concave functions, the one-stage cost

$$(x, f) \mapsto \hat{c}(x, f, T(x, f, y, z)) = f(y) + \pi_R(f) + \hat{z} - z$$

is convex as a sum of convex functions and we have already verified the conditions of Corollary 6.18. Now, Corollary 6.20 yields that the value functions J_n and J_∞ are convex. Note that for a convex function $x \mapsto h(x)$ also $x \mapsto h(-x)$ is convex. Hence, we can infer from the Bellman equation

$$J_n(x, s, t) = \inf_{f \in \mathcal{F}} \mathbb{E}\Big[J_{n+1}\Big(x - f(Y) - \pi_R(f) + Z,\ s + t\big(f(Y) + \pi_R(f) + \hat{z} - Z\big),\ \beta t\Big)\Big]$$

that the reinsurance treaty f_1 is better than f_2 independently from time and state if

$$f_1(Y) + \pi_R(f_1) \leq_{cx} f_2(Y) + \pi_R(f_2) \tag{6.21}$$

I.e. a minimal element w.r.t. this order would be an optimal reinsurance treaty in every scenario. Even if such a minimal element does not exist, (6.21) can be used to reduce the optimization problem to a finite dimensional one in special cases.

Example 6.29. 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 treaties (i.e. retained loss functions) can be chosen from the class of *stop-loss* treaties

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

Due to (6.21) and the fact that $Y_1 \leq_{cx} Y_2$ implies $\mathbb{E}[Y_1] = \mathbb{E}[Y_2]$, it suffices to find an $a_f \in [0, \infty]$ such that

$$\min\{Y, a_f\} \leq_{cx} f(Y). \tag{6.22}$$

The mapping $[0, \infty] \to \mathbb{R}_+,\ a \mapsto \min\{Y(\omega), a\}$ 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 $[0, \infty] \to \mathbb{R}_+,\ a \mapsto \mathbb{E}[\min\{Y, a\}]$ is continuous. Furthermore,

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

Hence, by the intermediate value theorem there is an $a_f \in [0, \infty]$ such that $\mathbb{E}[f(Y)] =$

$\mathbb{E}[\min\{Y, a_f\}]$. Let us compare the survival functions:

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

The inequality holds since $f \leq \mathrm{id}_{\mathbb{R}_+}$. Hence, we have $S_{\min\{Y,a_f\}}(y) \geq S_{f(Y)}(y)$ for $y < a_f$ and $S_{\min\{Y,a_f\}}(y) \leq S_{f(Y)}(y)$ for $y \geq a_f$. The cut criterion (Müller and Stoyan; 2002, 1.5.17) implies $\min\{Y, a_f\} \leq_{icx} f(Y)$ and due to the equality in expectation follows (6.22), cf. Müller and Stoyan (2002, 1.5.3). So the inner optimization problem (6.20) is reduced to finding an optimal nonnegative parameter at every stage. If the claims (Y_n) are bounded, one can apply the algorithm in Section 6.2.2 and approximate the optimal reinsurance problem (6.19) by an entirely finite dimensional problem.

6.4. Outlook

Due to the subadditivity and positive homogeneity of spectral risk measures, a more conservative alternative to minimizing the risk capital for the total discounted cost (6.1) is to consider the total discounted risk capital for the one-stage costs

$$\inf_{\pi\in\Pi} \sum_{k=0}^{N-1} \beta^k \rho_\phi \left(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\right). \tag{6.23}$$

In order to simplify the exposition, we directly omit terminal costs here. Optimization problem (6.23) can be addressed with similar techniques than (6.1). Given that the one-stage costs are bounded below, one can apply Proposition 2.11, interchange infima and separate an inner and outer problem:

$$\begin{aligned} &\inf_{\pi\in\Pi} \sum_{k=0}^{N-1} \beta^k \rho_\phi \left(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\right) \\ &= \inf_{\pi\in\Pi} \sum_{k=0}^{N-1} \beta^k \left(\inf_{g_k\in G} \mathbb{E}\left[g_k\big(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\big)\right] + \int_0^1 g_k^*(\phi(u))\,\mathrm{d}\,u \right) \\ &= \inf_{\pi\in\Pi} \inf_{(g_0,\dots,g_{N-1})\in\mathcal{G}^N} \left(\mathbb{E}\left[\sum_{k=0}^{N-1} \beta^k g_k\big(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\big)\right] + \sum_{k=0}^{N-1} \beta^k \int_0^1 g_k^*(\phi(u))\,\mathrm{d}\,u \right) \\ &= \inf_{(g_0,\dots,g_{N-1})\in\mathcal{G}^N} \left(\inf_{\pi\in\Pi} \mathbb{E}\left[\sum_{k=0}^{N-1} \beta^k g_k\big(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\big)\right] + \sum_{k=0}^{N-1} \beta^k \int_0^1 g_k^*(\phi(u))\,\mathrm{d}\,u \right) \end{aligned}$$

For fixed functions $g = (g_0, \dots, g_{N-1}) \in \mathcal{G}^N$ the resulting inner optimization problem

$$V_0^g(x) = \inf_{\pi\in\Pi} \mathbb{E}\left[\sum_{k=0}^{N-1} \beta^k g_k\big(c_k(X_k^\pi, d_k(H_k^\pi), X_{k+1}^\pi)\big)\right]$$

is a risk-neutral MDP with one-stage cost functions $(x, a, x') \mapsto g_k(c_k(x, a, x'))$. I.e. it can be solved without a state space extension using standard methods. The outer optimization problem

$$\inf_{g=(g_0,\dots,g_{N-1})\in\mathcal{G}^N} V_0^g(x) + \sum_{k=0}^{N-1} \beta^k \int_0^1 g_k^*(\phi(u)) \,\mathrm{d}\, u$$

becomes more complicated than the one in Section 6.2 but has still a similar structure. I.e. existence of an optimal solution can be guaranteed with analogous arguments.

For an infinite planning horizon, however, suitable convergence conditions are needed to enable the separation into an inner and outer problem as above. Moreover, one is faced with additional technical difficulties to guarantee the existence of a solution for the outer problem. A comprehensive study remains open to further research.

APPENDIX A

Complements of Analysis

A.1. Semicontinuous Functions

Let (E, d) be a metric space.

Definition A.1. a) A function $v : E \to \bar{\mathbb{R}}$ is called *lower semicontinuous* (l.s.c.) if for all sequences $\{x_n\}_{n\in\mathbb{N}} \subseteq E$ with $\lim_{n\to\infty} x_n = x \in E$ it holds

$$\liminf_{n\to\infty} v(x_n) \geq v(x).$$

b) A function $v : E \to \bar{\mathbb{R}}$ is called *upper semicontinuous* (u.s.c.) if $-f$ is lower semicontinuous.

Lower and upper semicontinuity together imply continuity. The following considerations are restricted to lower semicontinuous functions. Mutatis mutandis, they apply to upper semicontinuous functions as well.

Lemma A.2 (Hernández-Lerma and Lasserre; 1996, A.1)**.** *The following are equivalent:*

a) $v : E \to \bar{\mathbb{R}}$ *is lower semicontinuous.*

b) For all $x \in E$ *and* $\epsilon > 0$ *there exists* $\delta > 0$ *such that* $v(y) \geq v(x) - \epsilon$ *for all* $y \in E$ *with* $d(x, y) \leq \delta$*.*

c) The epigraph $\operatorname{epi}(v) = \{(x, y) \in E \times \mathbb{R} : v(x) \leq y\}$ *is closed (w.r.t. the product topology).*

d) The sublevel sets $\operatorname{lev}_{\leq\alpha}(v) = \{x \in E : v(x) \leq \alpha\}$ *are closed for every* $\alpha \in \mathbb{R}$*.*

Corollary A.3. *a) Lower semicontinuous functions are Borel measurable.*

b) Let I be any index set and $v_i : E \to \bar{\mathbb{R}}$ be lower semicontinuous, $i \in I$. Then $E \ni x \mapsto \sup_{i\in I} v_i(x)$ is lower semicontinuous.

Proof. a) Follows directly from part d) of Lemma A.2.

b) Note that

$$\left\{x \in E : \sup_{i\in I} v_i(x) \leq \alpha\right\} = \bigcap_{i\in I}\{x \in E : v_i(x) \leq \alpha\}.$$

Now the claim follows again from part d) of Lemma A.2 since the intersection of an arbitrary number of closed sets is closed. □

Lemma A.4. *Let E, E' be metric spaces.*

a) If $u : E \to E'$ is continuous and $v : E \times E' \to \mathbb{R}$ is lower semicontinuous, then the composition $v(\,\cdot\,, u(\,\cdot\,)) : E \to \mathbb{R}$ is lower semicontinuous.

b) If $u : E \to \mathbb{R}$ is lower semicontinuous and $v : E \times \mathbb{R} \to \mathbb{R}$ is lower semicontinuous and increasing in the second argument, then the composition $v(\,\cdot\,, u(\,\cdot\,)) : E \to \mathbb{R}$ is lower semicontinuous.

Proof. Let $\{x_n\}_{n\in\mathbb{N}} \subseteq E$ be a sequence with $\lim_{n\to\infty} x_n = x \in E$.

a) It holds $(x_n, u(x_n)) \to (x, u(x)) \in E \times E'$ due to the continuity of u. Hence, the lower semicontinuity of v implies

$$\liminf_{n\to\infty} v(x_n, u(x_n)) \geq v(x, u(x)).$$

b) We define the increasing sequence $\{u_n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}$ by $u_n = \inf_{k\geq n} u(x_k)$. First, assume that $\{u_n\}_{n\in\mathbb{N}}$ is bounded from above. Then, the sequence is convergent with limit, say, $\hat{u}$ and it holds

$$\hat{u} = \lim_{n\to\infty} u_n = \lim_{n\to\infty} \inf_{k\geq n} u(x_k) = \liminf_{n\to\infty} u(x_n) \geq u(x),$$

since u is lower semicontinuous. Now the lower semicontinuity and monotonicity in the second argument of v imply

$$\begin{aligned}\liminf_{n\to\infty} v(x_n, u(x_n)) &\geq \liminf_{n\to\infty} v(x_n, \inf_{k\geq n} u(x_k))\\ &= \liminf_{n\to\infty} v(x_n, u_n)\\ &\geq v(x, \hat{u})\\ &\geq v(x, u(x)).\end{aligned}$$

If, however, $\{u_n\}_{n\in\mathbb{N}}$ is unbounded from above, then there is an $N \in \mathbb{N}$ such that $u_n \geq u(x)$ for all $n \geq N$. Consequently,

$$\liminf_{n\to\infty} v(x_n, u(x_n)) \geq \liminf_{n\to\infty} v(x_n, \inf_{k\geq n} u(x_n))$$

$$\begin{aligned} &= \liminf_{n\to\infty} v(x_n, u_n) \\ &\geq \liminf_{n\to\infty} v(x_n, u(x)) \\ &\geq v(x, u(x)), \end{aligned}$$

as v is increasing in the second argument and lower semicontinuous in the first. □

Remark A.5. If in Lemma A.4 b) the inner function u is upper semicontinuous and the outer function v is lower semicontinuous and decreasing in the second argument, the assertion holds, too.

Lemma A.6. *a) Let $v : \mathbb{R} \to \mathbb{R}$ be increasing. Then v is lower semicontinuous if and only if it is left-continuous.*

b) Let $v : \mathbb{R} \to \mathbb{R}$ be decreasing. Then v is lower semicontinuous if and only if it is right-continuous.

Proof. a) Let v be left-continuous and $x_n \to x$. If $x_n > x$ for almost all $n \in N$, then by monotonicity $v(x_n) \geq v(x)$ for these n and $\liminf_{n\to\infty} v(x_n) \geq v(x)$. Otherwise, $\{x_{n_k}\}_{k\in\mathbb{N}}$ with $\{n_k\}_{k\in\mathbb{N}} = \{n \in \mathbb{N} : \ x_n \leq x\}$ defines a subsequence. Since v is left-continuous and increasing it holds $\liminf_{n\to\infty} v(x_n) \geq \liminf_{k\to\infty} v(x_{n_k}) = \lim_{k\to\infty} v(x_{n_k}) = v(x)$. Hence, v is lower semicontinuous.
Now, let v be lower semicontinuous and $x_n \uparrow x$. As v is increasing, $v(x_n) \leq v(x)$ for all $n \in \mathbb{N}$. Together with the lower semicontinuity we get

$$v(x) \leq \liminf_{n\to\infty} v(x_n) \leq \limsup_{n\to\infty} v(x_n) \leq v(x),$$

i.e. v is left-continuous.

b) Let v be decreasing. Then $\tilde{v}(x) = v(-x)$ is increasing. Furthermore, $\tilde{v}$ is lower-semicontinuous if and only if v has this property and $\tilde{v}$ is left-continuous if and only if v is right-continuous. So the assertion follows from part a). □

Due to the following version of Weierstraß' extrem value theorem, semicontinuous functions play an important role in optimization.

Theorem A.7 (Bäuerle and Rieder; 2011, A.1.2)**.** *Let E be compact and $v : E \to \bar{\mathbb{R}}$ be lower semicontinuous. Then v attains its infimum.*

Definition A.8. A sequence $\{v_n\}_{n\in\mathbb{N}}$ of functions $v_n : E \to \mathbb{R}$ is called *weakly increasing* if there exists another sequence $\{\delta_n\}_{n\in\mathbb{N}}$ of functions $\delta_n : E \to \mathbb{R}_-$ with $\lim_{n\to\infty} \delta_n(x) = 0$ for all $x \in E$ such that

$$v_n(x) \geq v_m(x) + \delta_m(x) \qquad \text{for all } x \in E \text{ and } n \geq m.$$

Lemma A.9 (Bäuerle and Rieder; 2011, A.1.4, A.1.6)**.** *Let $\{v_n\}_{n\in\mathbb{N}}$ be a sequence of weakly increasing functions $v_n : E \to \mathbb{R}$ and $\{\delta_n\}_{n\in\mathbb{N}}$ as in the previous definition.*

a) The pointwise limit $v_\infty = \lim_{n\to\infty} v_n$ *exists.*

b) If the functions v_n *and* δ_n *are lower semicontinuous for all* $n \in \mathbb{N}$*, then so is* v_∞*.*

c) If δ_m *does not depend on* x *for all* $m \in \mathbb{N}$*, then*

$$\lim_{n\to\infty} \sup_{x\in E} v_n(x) = \sup_{x\in E} \lim_{n\to\infty} v_n(x) = \sup_{x\in E} v_\infty(x).$$

Let $b : E \to [1, \infty)$ and consider the set

$$\mathbb{B} = \{v : E \to \mathbb{R} \mid v \text{ lower semicontinuous with } \lambda \in \mathbb{R}_+ \text{ s.t. } |v(x)| \leq \lambda\, b(x) \text{ for all } x \in E\}$$

endowed with the weighted supremum norm

$$\|v\|_b = \sup_{x\in E} \frac{|v(x)|}{b(x)}.$$

Lemma A.10. *The set* $\mathbb{B}$ *is closed w.r.t.* $\|\cdot\|_b$*.*

Proof. We have to show the lower semicontinuity of the limit v of a sequence of lower semicontinuous functions $\{v_n\}_{n\in\mathbb{N}}$ that is convergent w.r.t $\|\cdot\|_b$. Let $\epsilon > 0$ and $x, y \in E$ s.t. $d(x, y) \leq \delta$ for sufficiently small $\delta > 0$. Due to convergence w.r.t. $\|\cdot\|_b$ there exists $N \in \mathbb{N}$ s.t.

$$\|v_n - v\|_b \leq \frac{\epsilon}{3\max\{b(x), b(y)\}} \qquad \text{for all } n \geq N,$$

i.e.

$$v(z) - \frac{\epsilon b(z)}{3\max\{b(x), b(y)\}} \leq v_n(z) \leq v(z) + \frac{\epsilon b(z)}{3\max\{b(x), b(y)\}} \qquad \text{for all } n \geq N \text{ and } z \in E.$$

Consequently, we have for $n \geq N$

$$\begin{aligned} v(y) &\geq v_n(y) - \frac{\epsilon b(y)}{3\max\{b(x), b(y)\}} \\ &\geq v_n(y) - \frac{\epsilon}{3} \\ &\geq v_n(x) - \frac{\epsilon}{3} - \frac{\epsilon}{3} \\ &\geq v(x) - \frac{\epsilon}{3} - \frac{\epsilon}{3} - \frac{\epsilon b(x)}{3\max\{b(x), b(y)\}} \\ &\geq v(x) - \epsilon. \end{aligned}$$

Here, the third inequality is by Lemma A.2 b) since v_n is lower semicontinuous and δ sufficiently small. Now, by the same lemma v is lower semicontinuous. □

A.2. Set-Valued Mappings

Let E and A be non-empty Borel spaces.

Definition A.11. A *set-valued mapping* (also known as multifunction or correspondence) $D(\cdot)$ from E to A is a function such that $D(x)$ is a non-empty subset of A for every $x \in E$.

By $D = \{(x, a) \in E \times A : a \in D(x)\}$ we denote the graph of a set-valued mapping $D(\cdot)$ from E to A. A set valued mapping $D(\cdot)$ is called *closed*, if its graph D is closed and *closed-valued* if $D(x)$ is closed for all $x \in E$.

We define for a subset $S \subseteq A$ the *upper inverse* $D^u(S) = \{x \in E : D(x) \subseteq S\}$ and the *lower inverse* $D^\ell(S) = \{x \in E : D(x) \cap S \neq \emptyset\}$. Based on these inverses, we define the following continuity properties of set-valued mappings.

Definition A.12. A set-valued mapping $D(\cdot)$ from E to A is called

a) *upper semicontinuous* if $D^u(S)$ is open for every open subset $S \subset A$.

b) *lower semicontinuous* if $D^\ell(S)$ is open for every open subset $S \subset A$.

c) *continuous* if it is both upper and lower semicontinuous.

Note that some authors refer to these properties as *hemicontinuity* in order to distinguish them from the respective properties of singleton-valued functions. Also note that singleton-valued mappings are upper semicontinuous if and only if they are lower semicontinuous if and only if they are continuous if and only if they are continuous viewed as a function. We have the following characterizations of semicontinuity.

Proposition A.13 (Aliprantis and Border; 2006, 17.20, 17.21)**.** *a) A set-valued mapping $D(\cdot)$ from E to A with compact values is upper semicontinuous if and only if it has the following property for every $x \in E$: If $x_n \to x$ and $a_n \in D(x_n)$ for every $n \in \mathbb{N}$, then $\{a_n\}_{n\in\mathbb{N}}$ has an accumulation point in $D(x)$.*

b) A set-valued mapping $D(\cdot)$ from E to A is lower semicontinuous if and only if it has the following property for every $x \in E$: If $x_n \to x$ then every $a \in D(x)$ is an accumulation point of a sequence of a sequence $\{a_n\}_{n\in\mathbb{N}}$ with $a_n \in D(x_n)$.

There are also different sufficient conditions for semicontinuity.

Lemma A.14 (Aliprantis and Border; 2006, 17.18)**.** *Let $D_1(\cdot)$ be an upper semicontinuous and compact-valued set-valued mapping from E to A. If $D_2(\cdot)$ is another set-valued map from E to A with a closed graph satisfying $D_2(x) \subseteq D_1(x)$ for all $x \in E$, then $D_2(\cdot)$ is upper semicontinuous, too.*

Lemma A.15 (Aliprantis and Border; 2006, 17.11)**.** *Let A be compact and $D(\cdot)$ a set-valued mapping from E to A.*

a) $D(\cdot)$ is upper semicontinuous and closed-valued if and only if D is closed.

b) If $D(x) = A$ for all $x \in E$ then $D(\cdot)$ is continuous.

The converse of Lemma A.15 a) does not require A to be compact.

Lemma A.16 (Hernández-Lerma and Lasserre; 1996, D.3)**.** *A closed-valued and upper semicontinuous set-valued mapping is closed.*

Semicontinuity is preserved by compositions.

Lemma A.17 (Aliprantis and Border; 2006, 17.23)**.** *Let A, B, C be non-empty Borel spaces, $D_1(\cdot)$ be an upper (lower) semicontinuous set-valued mapping from A to B and $D_2(\cdot)$ be an upper (lower) semicontinuous set-valued mapping from B to C. Then the composition*

$$(D_2 \circ D_1)(x) = \bigcup_{y \in D_1(x)} D_2(y)$$

is an upper (lower) semicontinuous set-valued mapping from A to C.

Definition A.18. A set-valued mapping $D(\cdot)$ from $\mathbb{R}$ to A is called *increasing* if $D(x) \subseteq D(y)$ for $x \leq y$ and *decreasing* if $D(x) \supseteq D(y)$ for $x \leq y$.

Lemma A.19. *Let $D(\cdot)$ a set-valued mapping from $\mathbb{R}$ to A and $\{v_a\}_{a \in A}$ a family of functions $v_a : \mathbb{R} \to \mathbb{R}$.*

a) If $D(\cdot)$ is increasing and the functions $\{v_a\}_{a \in A}$ are decreasing, then the function

$$\mathbb{R} \ni x \mapsto \inf_{a \in D(x)} v_a(x)$$

is decreasing.

b) If $D(\cdot)$ is decreasing and the functions $\{v_a\}_{a \in A}$ are increasing, then the function

$$\mathbb{R} \ni x \mapsto \inf_{a \in D(x)} v_a(x)$$

is increasing.

Proof. For part a) let $x \leq y$, then $v_a(x) \geq v_a(y)$ for every $a \in A$. Consequently,

$$\inf_{a \in D(x)} v_a(x) \geq \inf_{a \in D(x)} v_a(y) \geq \inf_{a \in D(y)} v_a(y).$$

Part b) follows analogously. □

A.3. Optimal Measurable Selection

The following results on optimal measurable selection are based on Rieder (1978) with the error outlined in Wagner (1980) taken into account. Let $(E, \mathcal{E})$ and $(A, \mathcal{A})$ be measurable spaces. For a subset $C \subseteq E \times A$ we denote by $pC = \{x \in E : (x, a) \in C\}$ the projection onto E and by $C(x) = \{a \in A : (x, a) \in C\}$ the x-section.

Definition A.20. A family $\mathcal{L}$ of subsets of $E \times A$ is called *selection class* for $(\mathcal{E}, \mathcal{A})$ if

(i) $C \in \mathcal{L}$ implies $pC \in \mathcal{E}$

(ii) and every non-empty $C \in \mathcal{L}$ admits a measurable selection, i.e. there is a measurable map $d : pC \to A$ with $d(x) \in C(x)$ for all $x \in pC$.

The map d is called *measurable selector*. The family $\{X \times Y : X \in \mathcal{E}, Y \subseteq A\}$ is always a selection class since it admits constant selectors. Henceforth, fix $D \subseteq E \times A$ and a selection class $\mathcal{L}$ for $(\mathcal{E}, \mathcal{A})$. Let $u : D \to \bar{\mathbb{R}}$ and define $v, w : pD \to \bar{\mathbb{R}}$ as

$$v(x) = \inf_{a \in D(x)} u(x,a) \qquad \text{and} \qquad w(x) = \sup_{a \in D(x)} u(x,a).$$

Definition A.21. Let $\epsilon > 0$. A measurable selector $d : pD \to A$ is called

a) ϵ-*minimizer* of u if for all $x \in pD$

$$u(x, d(x)) \leq \begin{cases} v(x) + \epsilon, & \text{if } v(x) > -\infty, \\ -\frac{1}{\epsilon}, & \text{if } v(x) = -\infty, \end{cases}$$

and ϵ-*maximizer* of u if for all $x \in pD$

$$u(x, d(x)) \geq \begin{cases} w(x) - \epsilon, & \text{if } v(x) < \infty, \\ \frac{1}{\epsilon}, & \text{if } v(x) = \infty. \end{cases}$$

b) *minimizer* or *maximizer* of u if for all $x \in pD$ it holds $u(x, d(x)) = v(x)$ or $u(x, d(x)) = w(x)$, respectively.

Theorem A.22 (Rieder; 1978, 3.2)**.** *If $D \in \mathcal{L}$ and*

$$\{(x,a) \in D : u(x,a) \leq c\} \in \mathcal{L} \qquad \textit{for all } c \in \mathbb{R},$$

then v is measurable and for every $\epsilon > 0$ there exist an ϵ-minimizer. Replacing $\leq$ by $\geq$ yields the existence of an ϵ-maximizer.

Theorem A.23 (Rieder; 1978, 3.7)**.** *Let A be a separable metric space and $\mathcal{A} = \mathcal{B}(A)$ the Borel σ-algebra. If*

(i) $D \in \mathcal{L}$,

(ii) $\{(x,a) \in D : u(x,a) \leq c\} \in \mathcal{L}$ *for all* $c \in \mathbb{R}$,

(iii) $\{a \in D(x) : u(x,a) \leq c\}$ *is compact for all* $c \in \mathbb{R},\ x \in pD$,

then v is measurable and there exist a minimizer. Replacing $\leq$ by $\geq$ yields the existence of a maximizer.

The following result is a special case.

Proposition A.24 (Hernández-Lerma and Lasserre; 1996, D.5)**.** *Let E and A be Borel spaces, u measurable, $u(x, \cdot)$ lower semicontinuous for each $x \in E$ and $E \ni x \mapsto D(x)$ compact-valued. Then v is measurable and there exists a measurable minimizer. Alternatively, if $u(x, \cdot)$ is upper semicontinuous for all $x \in E$, then w is measurable and there exists a measurable maximizer.*

The next Proposition makes stronger conclusions if certain continuity properties are fulfilled.

Proposition A.25 (Bäuerle and Rieder; 2011, 2.4.3)**.** *Let E and A be Borel spaces, u lower semicontinuous and $E \ni x \mapsto D(x)$ upper semicontinuous and compact-valued. Then v is lower semicontinuous and there exists a minimizer. Alternatively, if u is upper semicontinuous on D, then w is upper semicontinuous and there exists a maximizer.*

A.4. Minimax Theorem

An extended-real-valued function defined on the product of two convex subsets of real vector spaces is called *concave-convex* if it is concave in the fist argument and convex in the second. This property can be generalized to domains without linear structure. Let X, Y be non-empty sets.

Definition A.26. A function $f : X \times Y \to \bar{\mathbb{R}}$ is called *concave-convex-like* if it is concavelike in the first argument and convexlike in the second, i.e. if for every $x_1, x_2 \in X$ and $\lambda \in [0,1]$ there is an $x_3 \in X$ such that

$$\lambda f(x_1, y) + (1-\lambda) f(x_2, y) \le f(x_3, y) \quad \text{for all } y \in Y,$$

and for all $y_1, y_2 \in Y$ and $\lambda \in [0,1]$ there is an $y_3 \in Y$ such that

$$\lambda f(x, y_1) + (1-\lambda) f(x, y_2) \ge f(x, y_3) \quad \text{for all } x \in X.$$

Note that a concave–convex function is concave–convex-like. In the next theorem, we implicitly require for semicontinuous functions that their domain is the subset of a metric space.

Theorem A.27 (Sion; 1958, 4.1, 4.2)**.** *a) Let X be any set, Y compact and $f : X \times Y \to \bar{\mathbb{R}}$ concave-convex-like and lower semicontinuous in the second argument, then*

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

b) Let X be compact, Y any set and $f : X \times Y \to \bar{\mathbb{R}}$ concave-convex-like and upper semicontinuous in the first argument, then

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

For a detailed account of different minimax theorems see Chapter 2.3 in Barbu and Precupanu (2012).

Remark A.28. The assumptions of Theorem A.27 a) imply that the infimum on both sides is attained, since the supremum of a collection of lower semicontinuous functions is lower semicontinuous (Lemma A.3). Likewise, in part b) the suprema are attained. If both the infima and the suprema are attained, e.g. if both the conditions of parts a) and b) are fulfilled, the function f is said to satisfy the *minimax equality*

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

The minimax equality is related to the notion of a saddle point.

Definition A.29. A pair $(\tilde{x},\tilde{y}) \in X \times Y$ is a *saddle point* of the function $f : X \times Y \to \bar{\mathbb{R}}$, if

$$f(x,\tilde{y}) \le f(\tilde{x},\tilde{y}) \le f(\tilde{x},y) \qquad \text{for all } (x,y) \in X \times Y.$$

Lemma A.30 (Barbu and Precupanu; 2012, 2.105)**.** *A function $f : X \times Y \to \bar{\mathbb{R}}$ satisfies the minimax equality if and only if it has a saddle point.*

A.5. Miscellaneous

Lemma A.31. *Let E be a non-empty set and $u, v : E \to \mathbb{R}$ bounded functions. Then*

a)

$$\inf_{x\in E} u(x) - \inf_{x\in E} v(x) \le \sup_{x\in E} u(x) - v(x),$$

b)

$$\left|\inf_{x\in E} u(x) - \inf_{x\in E} v(x)\right| \le \sup_{x\in E} \left|u(x) - v(x)\right|.$$

c)

$$\sup_{x\in E} u(x) - \sup_{x\in E} v(x) \le \sup_{x\in E} u(x) - v(x),$$

d)

$$\left|\sup_{x\in E} u(x) - \sup_{x\in E} v(x)\right| \le \sup_{x\in E} \left|u(x) - v(x)\right|.$$

Proof. We only proof parts a) and b). The rest follows analogously.

a) From $v = v - u + u$ it follows

$$\begin{aligned}\inf_{x\in E} v(x) &= \inf_{x\in E} v(x) - u(x) + u(x)\\ &\ge \inf_{x\in E} v(x) - u(x) + \inf_{x\in E} u(x)\end{aligned}$$

and by subtraction

$$\inf_{x\in E} v(x) - \inf_{x\in E} u(x) \ge \inf_{x\in E} v(x) - u(x).$$

Multiplying this inequality with (-1) yields the assertion:

$$\begin{aligned}\inf_{x\in E} u(x) - \inf_{x\in E} v(x) &\le -\inf_{x\in E} v(x) - u(x)\\ &= \sup_{x\in E} u(x) - v(x).\end{aligned}$$

b) From part a) we have

$$\begin{aligned}\inf_{x\in E} u(x) - \inf_{x\in E} v(x) &\leq \sup_{x\in E} u(x) - v(x)\\ &\leq \sup_{x\in E} |u(x) - v(x)|\,.\end{aligned}$$

Interchanging the roles of u and v completes the proof. □

Consider the space $C(\mathbb{R},\mathbb{R})$ of continuous real valued functions on $\mathbb{R}$ endowed with the metric

$$m(f,g) = \sum_{j=1}^{\infty} 2^{-j} \frac{\max_{|t|\leq j} |f(t) - g(t)|}{1 + \max_{|t|\leq j} |f(t) - g(t)|}.$$

m metricizes the topology of compact convergence on $C(\mathbb{R},\mathbb{R})$, i.e. $m(f_n, f) \to 0$ if and only if f_n converges uniformly to f on every compact subset of $\mathbb{R}$. We need the following version of the *Arzelà-Ascoli Theorem*.

Theorem A.32 (Pugh; 2015, 4.18)**.** *A subset $\mathcal{F} \subseteq C(\mathbb{R},\mathbb{R})$ is relatively compact if and only if*

(i) $\{f(x) : f \in \mathcal{F}\}$ *is bounded for all* $x \in \mathbb{R}$ *and*

(ii) $\lim_{\delta\to 0} \sup_{f\in\mathcal{F}} \sup_{x,y\in\{[-j,j]:\ |x-y|<\delta\}} |f(x) - f(y)| = 0$ *for all* $j \in \mathbb{N}$.

Remark A.33. a) Theorem A.32 remains true in $C(I,\mathbb{R})$, where $I \subseteq \mathbb{R}$ is any closed interval. Regarding the metric m, the sequence $\{[-j,j]\}_{j\in\mathbb{N}}$ can then be replaced by any ascending sequence of subsets of I whose union is I.

b) The second condition of the theorem means that the familiy $\mathcal{F}$ is uniformly equicontinuous. A common Lipschitz constant is sufficient.

APPENDIX B

Complements of Measure and Probability Theory

B.1. Integration

First, we state the most suitable version of *Fatou's Lemma* for our purposes.

Theorem B.1 (Klenke; 2014, 4.21)**.** *Let* $(\Omega, \mathcal{A}, \mu)$ *be a measure space,* $v \in L^1(\Omega, \mathcal{A}, \mu)$*. For measurable functions* $v_1, v_2, \ldots$ *with* $v_n \geq v$ μ*-a.e.* ($n \in \mathbb{N}$) *it holds*

$$\int \left(\liminf_{n\to\infty} v_n \right) \mathrm{d}\,\mu \leq \liminf_{n\to\infty} \int v_n \,\mathrm{d}\,\mu.$$

Note that if μ is a probability measure, a constant lower bound is sufficient. In a similar way, the classical setting of *Tonelli's Theorem* with non-negative random variables can be extended to quasi-integrable ones.

Theorem B.2. *Let* $(\Omega_i, \mathcal{A}_i, \mu_i)$ *be* σ*-finite measure spaces,* $i = 1, 2$*. Further, let* $u, v : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ *be measurable with respect to* $\mathcal{A}_1 \otimes \mathcal{A}_2$ *and* $u \in L^1$*. If* $v \geq u$ μ*-a.e., then*

$$\begin{aligned} \omega_1 &\mapsto \int_{\Omega_2} v(\omega_1, \omega_2)\ \mu_2(\mathrm{d}\,\omega_2) \qquad \text{is } \mathcal{A}_1\text{-measurable,} \\ \omega_2 &\mapsto \int_{\Omega_1} v(\omega_1, \omega_2)\ \mu_1(\mathrm{d}\,\omega_1) \qquad \text{is } \mathcal{A}_2\text{-measurable,} \end{aligned}$$

and it holds

$$\int_{\Omega_1\times\Omega_2} v\ \mu_1 \otimes \mu_2(\mathrm{d}\,\omega) = \int_{\Omega_1} \left(\int_{\Omega_2} v(\omega_1, \omega_2)\ \mu_2(\mathrm{d}\,\omega_2) \right) \mu_1(\mathrm{d}\,\omega_1)$$

$$= \int_{\Omega_2} \left(\int_{\Omega_1} v(\omega_1, \omega_2)\ \mu_1(\mathrm{d}\,\omega_1) \right) \mu_2(\mathrm{d}\,\omega_2).$$

Proof. Apply Tonelli's/ Fubini's theorem (Klenke; 2014, 14.16) to the non-negative function $v - u$ and add the finite value

$$\int_{\Omega_1} \left(\int_{\Omega_2} u(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1)$$

on both sides. □

Again, for probability measures constant lower bounds are sufficient. Finally, we need a result to interchange series and integrals.

Theorem B.3 (Hinderer; 1970, A3)**.** *Let $X_1, X_2, \ldots$ be extended real-valued random variables such that $\sum_{n=1}^{\infty} \mathbb{E}X_n^+ < \infty$ or $\sum_{n=1}^{\infty} \mathbb{E}X_n^- < \infty$. Then*

a) $\sum_{n=1}^{N} X_n$ converges a.s. to a quasi-integrable random variable.

b) $\sum_{n=1}^{N} \mathbb{E}X_n$ converges to $\sum_{n=1}^{\infty} \mathbb{E}X_n^+ - \sum_{n=1}^{\infty} \mathbb{E}X_n^-$ as $N \to \infty$ and

$$\sum_{n=1}^{\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right].$$

B.2. Separability of Lebesgue Spaces

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space with finite measure and define

$$d(A, B) = \mu(A \triangle B) = \mu((A \cup B) \setminus (A \cap B)), \qquad A, B \in \mathcal{A}.$$

Sets $A, B \in \mathcal{A}$ with differ only by a null set if and only if $d(A, B) = 0$. The collection of equivalence classes $\mathcal{A}/\mu$ endowed with d forms a metric space, the so called *metric Boolean algebra* generated by $(\mathcal{A}, \mu)$, cf. Bogachev (2007, 1.12(iii)).

Definition B.4. A finite measure μ is called separable if the metric space $(\mathcal{A}/\mu, d)$ is separable.

Lemma B.5 (Bogachev; 2007, 1.12.102)**.** *Let $\mathbb{P}$ be a probability measure on a measurable space $(\Omega, \mathcal{A})$. If the σ-algebra $\mathcal{A}$ is countably generated, the probability measure $\mathbb{P}$ is separable.*

Lemma B.6 (Bogachev; 2007, 4.7.63)**.** *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space with finite measure. Then the Banach space $L^p(\Omega, \mathcal{A}, \mu)$, $1 \le p < \infty$, is separable if and only if μ is separable.*

B.3. Quantiles

Definition B.7. Let F_X be the distribution function of a real-valued random variable X.

a) The *(lower) quantile function* of X is the left-continuous generalized inverse of F_X

$$F_X^{-1}(\alpha) = q_X^-(\alpha) = \inf\{x \in \mathbb{R} : \; F_X(x) \geq \alpha\}, \qquad \alpha \in (0,1).$$

b) The *upper quantile function* of X is the right-continuous generalized inverse inverse of F_X

$$q_X^+(\alpha) = \inf\{x \in \mathbb{R} : \; F_X(x) > \alpha\}, \qquad \alpha \in (0,1).$$

It is standard in the literature that the short form quantile function refers to the lower quantile function. The following properties can be found in Embrechts and Hofert (2013, Prop. 1) and Föllmer and Schied (2016, A.3).

Lemma B.8. *Let X be a real-valued random variable with distribution function F_X, $x \in \mathbb{R}$ and $\alpha \in (0,1)$.*

a) It holds $q_x^-(\alpha) \leq q_x^+(\alpha)$ and they coincide Lebesgue a.e.
b) q_X^- is increasing, left-continuous and admits limits from the right.
c) q_X^+ is increasing, right-continuous and admits limits from the left.
d) It holds $q_X^-(\alpha) \leq x \Leftrightarrow \alpha \leq F_X(x)$.
e) It holds $q_X^+(\alpha) \geq x \Leftrightarrow \alpha \geq F_X(x)$.
f) $q_X^-(F_X(x)) \leq x \leq q_X^+(F_X(x))$.
g) $\alpha \leq F_X(q_X^-(\alpha)) \leq F_X(q_X^+(\alpha))$. If F_X is continuous, then $\alpha = F_X(q_X^-(\alpha)) = F_X(q_X^+(\alpha))$.
h) F_X is strictly increasing if and only if both q_X^- and q_X^+ are continuous. In this case, F_X is invertible in the usual sense and $q_X^- = q_X^+$ is the inverse function.
i) F_X is continuous if and only if both q_X^- and q_X^+ are strictly increasing.

Lemma B.9 (Dhaene et al.; 2002, Theorem 1)**.** *Let $h : \mathbb{R} \to \mathbb{R}$ be increasing and left-continuous. Then*

$$F_{h(X)}^{-1}(\alpha) = h(F_X^{-1}(\alpha)), \qquad \alpha \in (0,1).$$

It is easy to see that $q_X^-(U) \sim q_X^+(U) \sim X$ for any $U \sim \mathcal{U}(0,1)$. The following result shows that one can find a suitable $U = U_X$ such that equality in distribution strengthens to almost sure equality. The random variable U_X is referred to as *(generalized) distributional transform* of X.

Lemma B.10 (Rüschendorf; 2009, 2.1)**.** *For any random variable X on an atomless probability space there exists a random variable $U_X \sim \mathcal{U}(0,1)$ such that*

$$q_X^-(U_X) = q_X^+(U_X) = X \quad \mathbb{P}\text{-a.s.}$$

Corollary B.11. *Let X be a random variable and $h : \mathbb{R} \to \mathbb{R}$ increasing and left-continuous. Then X and $h(X)$ have the same distributional transform.*

The distributional transform is related to the following dependence concept for random vectors.

Definition B.12. Let $X = (X_1, \ldots, X_n)$ be a random vector and let $F_1^{-1}, \ldots, F_n^{-1}$ be the quantile functions of its components. X is called *comonotonic* if there exists a random variable $U \sim \mathcal{U}(0,1)$ such that

$$(X_1, \ldots, X_n) = \left(F_1^{-1}(U), \ldots, F_n^{-1}(U)\right) \quad \mathbb{P}\text{-a.s.}$$

Proposition B.13 (Rüschendorf; 2013, 2.14)**.** *Let $X = (X_1, \ldots, X_n)$ be a random vector with distribution function F. Then the following are equivalent.*

(i) *X is comonotonic.*

(ii) *F is the* upper Fréchet-Hoeffding bound

$$F(x) = \min\{F_1(x_1), \ldots, F_n(x_n)\}.$$

(iii) *There is a random variable Z and increasing functions $f_1, \ldots, f_n$ such that*

$$X = (f_1(Z), \ldots, f_n(Z)) \quad \mathbb{P}\text{-}a.s.$$

(iv) *For all $i, j = 1, \ldots, n$ and almost all $\omega, \omega' \in \Omega$ it holds*

$$X_i(\omega) \leq X_i(\omega') \Rightarrow X_j(\omega) \leq X_j(\omega').$$

Bibliography

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking & Finance* **26**(7): 1505–1518.

Albrecher, H., Beirlant, J. and Teugels, J. L. (2017). *Reinsurance: Actuarial and Statistical Aspects*, John Wiley & Sons, Hoboken, NJ.

Albrecher, H. and Thonhauser, S. (2009). Optimality results for dividend problems in insurance, *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas* **103**(2): 295–320.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd edn, Springer-Verlag, Berlin Heidelberg.

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

Asienkiewicz, H. and Jaśkiewicz, A. (2017). A note on a new class of recursive utilities in Markov decision processes, *Applicationes Mathematicae* **44**(2): 149–161.

Barbu, V. and Precupanu, T. (2012). *Convexity and Optimization in Banach Spaces*, 4th edn, Springer Netherlands, Dordrecht.

Bäuerle, N. and Glauner, A. (2018). Optimal risk allocation in reinsurance networks, *Insurance: Mathematics and Economics* **82**: 37–47.

Bäuerle, N. and Glauner, A. (2020). Distributionally robust Markov decision processes and their connection to risk measures. arXiv:2007.13103.

Bäuerle, N. and Jaśkiewicz, A. (2017). Optimal dividend payout model with risk sensitive preferences, *Insurance: Mathematics and Economics* **73**: 82–93.

Bäuerle, N. and Jaśkiewicz, A. (2018). Stochastic optimal growth model with risk sensitive preferences, *Journal of Economic Theory* **173**: 181–200.

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures: Consistency and bounds, *Insurance: Mathematics and Economics* **38**(1): 132–148.

Bäuerle, N. and Mundt, A. (2009). Dynamic mean-risk optimization in a binomial model, *Mathematical Methods of Operations Research* **70**(2): 219–239.

Bäuerle, N. and Ott, J. (2011). Markov decision processes with Average-Value-at-Risk criteria, *Mathematical Methods of Operations Research* **74**(3): 361–379.

Bäuerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*, Springer-Verlag, Berlin Heidelberg.

Bäuerle, N. and Rieder, U. (2014). More risk-sensitive Markov decision processes, *Mathematics of Operations Research* **39**(1): 105–120.

Bäuerle, N. and Rieder, U. (2019). Markov decision processes under ambiguity. arXiv:1907.02347.

Bellman, R. (1954). The theory of dynamic programming, *Bulletin of the American Mathematical Society* **60**(6): 503–515.

Bellman, R. (1957). *Dynamic Programming*, Princeton University Press, Princeton, NJ.

Bertsekas, D. P. (2012). *Dynamic Programming and Optimal Control*, Vol. 2, 4th edn, Athena Scientific, Belmont, Mass.

Bertsekas, D. P. (2017). *Dynamic Programming and Optimal Control*, Vol. 1, 4th edn, Athena Scientific, Belmont, Mass.

Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York.

Bielecki, T. R., Chen, T., Cialenco, I., Cousin, A. and Jeanblanc, M. (2019). Adaptive robust control under model uncertainty, *SIAM Journal on Control and Optimization* **57**(2): 925–946.

Birghila, C. and Pflug, G. C. (2019). Optimal XL-insurance under Wasserstein-type ambiguity, *Insurance: Mathematics and Economics* **88**: 30–43.

Blackwell, D. (1965). Discounted dynamic programming, *The Annals of Mathematical Statistics* **36**(1): 226–235.

Bogachev, V. I. (2007). *Measure Theory*, Vol. 1, Springer-Verlag, Berlin Heidelberg.

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

Cai, J. and Tan, K. S. (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures, *ASTIN Bulletin* **37**(1): 93–112.

Chen, Y. and Assa, H. (2019). Dynamic set-up for designing optimal reinsurance contracts. https://ssrn.com/abstract=3420090.

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

Corrias, L. (1996). Fast Legendre-Fenchel transform and applications to Hamilton-Jacobi equations and conservation laws, *SIAM Journal on Numerical Analysis* **33**(4): 1534–1558.

Cui, W., Yang, J. and Wu, L. (2013). Optimal reinsurance minimizing the distortion risk measure under general reinsurance premium principles, *Insurance: Mathematics and Economics* **53**(1): 74–85.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, John Wiley & Sons, Chichester.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: theory, *Insurance: Mathematics and Economics* **31**(1): 3–33.

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

Dhaene, J. and Wang, S. (1998). Comonotonicity, correlation order and premium principles, *Insurance: Mathematics and Economics* **22**(3): 235–242.

Dhaene, J., Wang, S., Young, V. and Goovaerts, M. J. (2000). Comonotonicity and maximal stop-loss premiums, *Bulletin of the Swiss Association of Actuaries* **2000**(2): 99–113.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms, *The Quarterly Journal of Economics* **75**(4): 643–669.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses, *Mathematical Methods of Operations Research* **77**(3): 423–432.

Epstein, L. G. and Schneider, M. (2003). Recursive multiple-priors, *Journal of Economic Theory* **113**(1): 1–31.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* **57**(4): 937–969.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance: An Introduction in Discrete Time*, 4th edn, De Gruyter, Berlin Boston.

González-Trejo, J. I., Hernández-Lerma, O. and Hoyos-Reyes, L. F. (2002). Minimax control of discrete-time stochastic systems, *SIAM Journal on Control and Optimization* **41**(5): 1626–1659.

Hansen, L. P. and Sargent, T. J. (1995). Discounted linear exponential quadratic Gaussian control, *IEEE Transactions on Automatic Control* **40**(5): 968–971.

Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York.

Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*, Springer-Verlag, New York.

Hinderer, K. (1970). *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter*, Springer-Verlag, Berlin Heidelberg.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*, Technology Press of Massachusetts Institute of Technology, Cambridge, Mass.

Howard, R. A. and Matheson, J. E. (1972). Risk-sensitive Markov decision processes, *Management Science* **18**(7): 356–369.

Iyengar, G. N. (2005). Robust dynamic programming, *Mathematics of Operations Research* **30**(2): 257–280.

Jaśkiewicz, A. and Nowak, A. S. (2011). Stochastic games with unbounded payoffs: Applications to robust control in economics, *Dynamic Games and Applications* **1**(2): 253–279.

Jaśkiewicz, A. and Nowak, A. S. (2014). Robust Markov control processes, *Journal of Mathematical Analysis and Applications* **420**(2): 1337–1353.

Jouini, E., Schachermayer, W. and Touzi, N. (2006). Law invariant risk measures have the Fatou property, *in* S. Kusuoka and A. Yamazaki (eds), *Advances in Mathematical Economics*, Vol. 9, Springer-Verlag, Tokyo, pp. 49–71.

Klenke, A. (2014). *Probability Theory: A Comprehensive Course*, 2nd edn, Springer-Verlag, London.

Kreps, D. M. (1977a). Decision problems with expected utility critera, I: Upper and lower convergent utility, *Mathematics of Operations Research* **2**(1): 45–53.

Kreps, D. M. (1977b). Decision problems with expected utility criteria, II: Stationarity, *Mathematics of Operations Research* **2**(3): 266–274.

Kreps, D. M. and Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* **46**(1): 185–200.

Leskelä, L. and Vihola, M. (2013). Stochastic order characterization of uniform integrability and tightness, *Statistics & Probability Letters* **83**(1): 382–389.

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

Maskin, E. and Tirole, J. (1988). A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs, *Econometrica* **56**(3): 549–569.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, revised edn, Princeton University Press.

Miao, J. (2014). *Economic Dynamics in Discrete Time*, MIT Press, Cambridge, Mass.

Morrison, T. J. (2001). *Functional Analysis: An Introduction to Banach Space Theory*, John Wiley & Sons, New York.

Müller, A. (1997). How does the value function of a Markov decision process depend on the transition probabilities?, *Mathematics of Operations Research* **22**(4): 872–885.

Müller, A. (2007). Certainty equivalents as risk measures, *Brazilian Journal of Probability and Statistics* **21**(1): 1–12.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, John Wiley & Sons, Chichester.

Nilim, A. and El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices, *Operations Research* **53**(5): 780–798.

Ó Searcóid, M. (2007). *Metric Spaces*, Springer-Verlag, London.

Osogami, T. (2012). Robustness and risk-sensitivity in Markov decision processes, *in* F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger (eds), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pp. 233–241.

Pichler, A. (2013). The natural Banach space for version independent risk measures, *Insurance: Mathematics and Economics* **53**(2): 405–415.

Pichler, A. (2015). Premiums and reserves, adjusted by distortions, *Scandinavian Actuarial Journal* **2015**(4): 332–351.

Pugh, C. C. (2015). *Real Mathematical Analysis*, 2nd edn, Springer, Cham.

Rieder, U. (1978). Measurable selection theorems for optimization problems, *Manuscripta Mathematica* **24**(1): 115–131.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of Conditional Value-at-Risk, *Journal of Risk* **2**(3): 21–41.

Rüschendorf, L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process, *Journal of Statistical Planning and Inference* **139**(11): 3921–3927.

Rüschendorf, L. (2013). *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios*, Springer-Verlag, Berlin Heidelberg.

Ruszczyński, A. (2010). Risk-averse dynamic programming for Markov decision processes, *Mathematical Programming* **125**(2): 235–261.

Satia, J. K. and Lave, R. E. (1973). Markovian decision processes with uncertain transition probabilities, *Operations Research* **21**(3): 728–740.

Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games, *International Journal of Game Theory* **4**(1): 25–55.

Sereda, E. N., Bronshtein, E. M., Rachev, S. T., Fabozzi, F. J., Sun, W. and Stoyanov, S. V. (2010). Distortion risk measures in portfolio optimization, *in* J. B. Guerard (ed.), *Handbook of Portfolio Construction*, Springer, New York, chapter 25, pp. 649–673.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*, Springer, New York.

Shapiro, A. (2012). Minimax and risk averse multistage stochastic programming, *European Journal of Operational Research* **219**(3): 719–726.

Shapiro, A. (2013). On Kusuoka representation of law invariant risk measures, *Mathematics of Operations Research* **38**(1): 142–152.

Shapiro, A. (2016). Rectangular sets of probability measures, *Operations Research* **64**(2): 528–541.

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

Tsanakas, A. and Desli, E. (2003). Risk measures and theories of choice, *British Actuarial Journal* **9**(4): 959–991.

Wagner, D. H. (1980). Survey of measurable selection theorems: An update, *in* D. Kölzow (ed.), *Measure Theory Oberwolfach 1979: Proceedings of the Conference Held at Oberwolfach, Germany, July 1–7, 1979*, Springer-Verlag, Berlin Heidelberg, pp. 176–219.

Wang, S. (1996). Premium calculation by transforming the layer premium density, *ASTIN Bulletin* **26**(1): 71–92.

White, C. C. and Eldeib, H. K. (1994). Markov decision processes with imprecise transition probabilities, *Operations Research* **42**(4): 739–749.

Wiesemann, W., Kuhn, D. and Rustem, B. (2012). Robust Markov decision processes, *Mathematics of Operations Research* **38**(1): 153–183.