MARKOV DECISION PROCESSES WITH AVERAGE-VALUE-AT-RISK CRITERIA

NICOLE BÄUERLE* AND JONATHAN OTT‡

Abstract. $\mathbb{P}$ We investigate the problem of minimizing the Average-Value-at-Risk ($AVaR_\tau$) of the discounted cost over a finite and an infinite horizon which is generated by a Markov Decision Process (MDP). We show that this problem can be reduced to an ordinary MDP with extended state space and give conditions under which an optimal policy exists. We also give a time-consistent interpretation of the $AVaR_\tau$. At the end we consider a numerical example which is a simple repeated casino game. It is used to discuss the influence of the risk aversion parameter τ of the $AVaR_\tau$-criterion.

Key words: Markov Decision Problem, Average-Value-at-Risk, Time-consistency, Risk aversion.

AMS subject classifications: 90C40, 91B06.

1. Introduction

Risk-sensitive optimality criteria for Markov Decision Processes (MDPs) have been considered by various authors over the years. In contrast to risk neutral optimality criteria which simply minimize expected discounted cost, risk-sensitive criteria often lead to non-standard MDPs which cannot be solved in a straightforward way by using the Bellman equation. This property is often called *time-inconsistency*. For example Howard & Matheson (1972) introduced the notion of *risk-sensitive* MDPs by using an exponential utility function. Jaquette (1973) considers moments of total discounted cost as an optimality criterion. Later e.g. Wu & Lin (1999) investigated the *target level criterion* where the aim is to maximize the probability that the total discounted reward exceeds a given target value. The related *target hitting criterion* is studied in Boda et al. (2004) where the aim is to minimize the probability that the total discounted reward does not exceed a given target value. A quite general problem is investigated in Collins & McNamara (1998). There the authors deal with a finite horizon problem which looks like an MDP, however the classical terminal reward is replaced by a strictly concave functional of the terminal distribution. Other probabilistic criteria, mostly in combination with long-run performance measures, can be found in the survey of White (1988).

Another quite popular risk-sensitive criterion is the *mean-variance* criterion, where the aim is to minimize the variance, given the expected reward exceeds a certain target. Since it is not possible to write down a straightforward Bellman equation it took some time until Li & Ng (2000) managed to solve these kind of problems in a multiperiod setting using MDP methods. In the last decade risk measures have become popular and the simple variance has been replaced by more complicated risk measures like Value-at-Risk (VaR_τ) or Average-Value-at-Risk ($AVaR_\tau$). Clearly when risk measures are used as optimization criteria, we cannot expect multiperiod problems to become time-consistent. In Bäuerle & Mundt (2009) a mean-$AVaR_\tau$ problem has been solved for an investment problem in a binomial financial market.

‡ The underlying projects have been funded by the Bundesministerium für Bildung und Forschung of Germany under promotional reference 03BAPAC1. The authors are responsible for the content of this article.

* Institute for Stochastics, Karlsruhe Institute of Technology, D-76128 Karlsruhe, Germany, e-mail: nicole.baeuerle@kit.edu.

Some authors now tackled the problem of formulating time-consistent risk-sensitive multi-period optimization problems. For example in Boda & Filar (2006) a time-consistent $AVaR_\tau$-problem has been given by restricting the class of admissible policies. Björk & Murgoci (2010) tackle the general problem of defining time-consistent controls using game theoretic considerations. A different notion of time-consistency has been discussed in Shapiro (2009). He calls a policy time-consistent if the current optimal action does not depend on paths which are known cannot happen in the future. In Shapiro (2009) it is shown that the $AVaR_\tau$ is not time-consistent w.r.t. this definition but an alternative formulation of a time-consistent criterion is given. Further time-consistency considerations for risk measures can e.g. be found in Artzner et al. (2007) or Bion-Nadal (2008).

In this paper we investigate the problem of minimizing the $AVaR_\tau$ of the discounted cost over a finite and an infinite horizon which is generated by a Markov Decision Process. We show that this problem can be reduced to an ordinary MDP with extended state space and give conditions under which an optimal policy exists. In particular it is seen that the optimal policy depends on the history only through a certain kind of 'sufficient statistic'. In the case of an infinite horizon we show that the minimal value can be characterized as the unique fixed point of a minimal cost operator. Further we give a time-consistent interpretation of the $AVaR_\tau$. At the end we also consider a numerical example which is a simple repeated casino game. It is used to discuss the influence of the risk aversion parameter τ of the $AVaR_\tau$. For $\tau \to 0$ the $AVaR_\tau$ coincides with the risk neutral optimization problem and for $\tau \to 1$ it coincides with the Worst-Case risk measure. We see that with increasing τ the distribution of the final capital narrows and the probability of getting ruined is decreasing.

The paper is organized as follows: In Section 2 we explain the joint state-cost process and the admissible policies. In Section 3 we solve the finite horizon $AVaR_\tau$ problem and give a time-consistent interpretation. Next, in Section 4 we consider and solve the infinite horizon problem and Section 5 contains the numerical example.

2. A Markov Decision Process with Average-Value-at-Risk Criteria

We suppose that a controlled Markov state process (X_n) in discrete time is given with values in a Borel set E, together with a non-negative cost process (C_n). All random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The evolution of the system is as follows: suppose that we are in state $X_n = x$ at time n. Then we are allowed to choose an action a from an action space A which is an arbitrary Borel space. In general we assume that not all actions from the set A are admissible. We denote by $D \subset E \times A$, the set of all admissible state-action combinations. The set $D(x) := \{a \in A : (x,a) \in D\}$ gives the admissible actions in state x for all states $x \in E$. When we choose the action $a \in D(x)$ at time n, a random cost $C_n \geq 0$ is incurred and a transition to the next state X_{n+1} takes place. The distribution of C_n and X_{n+1} is given by a transition kernel $\mathbb{Q}$ (see below). When we denote by A_n the (random) action which is chosen at time n, then we assume that A_n is $\mathcal{F}_n = \sigma(X_0, A_0, C_0, \ldots, X_n)$-measurable, i.e. at time n we are allowed to use the complete history of the state process for our decision. Thus we introduce recursively the sets of histories:

$$H_0 := E, \quad H_{k+1} := H_k \times A \times \mathbb{R} \times E$$

where $h_k = (x_0, a_0, c_0, x_1, \ldots, a_{k-1}, c_{k-1}, x_k) \in H_k$ gives a history up to time k. A *history-dependent policy* $\pi = (g_k)_{k \in \mathbb{N}_0}$ is given by a sequence of mappings $g_k : H_k \to A$ such that $g_k(h_k) \in D(x_k)$. We denote the set of all such policies by Π. A policy $\pi \in \Pi$ induces a probability measure $\mathbb{P}^\pi$ on $(\Omega, \mathcal{F})$. We suppose that there is a joint (stationary) transition kernel $\mathbb{Q}$ from $E \times A$ to $E \times \mathbb{R}$ such that

$$\begin{aligned}
& \mathbb{P}^\pi(X_{n+1} \in B_x, C_n \in B_c \mid X_0, g_0(X_0), C_0, \ldots, X_n, g_n(X_0, A_0, C_0, \ldots, X_n)) \\
= \;& \mathbb{P}^\pi(X_{n+1} \in B_x, C_n \in B_c \mid X_n, g_n(X_0, A_0, C_0, \ldots, X_n)) \\
= \;& \mathbb{Q}(B_x \times B_c \mid X_n, g_n(X_0, A_0, C_0, \ldots, X_n))
\end{aligned}$$

for measurable sets $B_x \subset E$ and $B_c \subset \mathbb{R}$. There is a discount factor $\beta \in [0,1]$ and we will either consider a finite planning horizon $N \in \mathbb{N}_0$ or an infinite planning horizon. Thus we will either consider the cost

$$C^N := \sum_{k=0}^{N} \beta^k C_k \quad \text{or} \quad C^\infty := \sum_{k=0}^{\infty} \beta^k C_k.$$

We will always assume that the random variables C_k are non-negative and bounded from above by a constant $\bar{C}$. Instead of minimizing the expected cost we will now use the non-standard criterion of minimizing the so-called *Average-Value-at-Risk* which is defined as follows (note that we assume here that large values of X are bad and small values of X are good):

Definition 2.1. Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ be a real-valued random variable and let $\tau \in (0,1)$.

a) The *Value-at-Risk* of X at level τ, denoted by $VaR_\tau(X)$ is defined by

$$VaR_\tau(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq \tau\}.$$

b) The *Average-Value-at-Risk* of X at level τ, denoted by $AVaR_\tau(X)$ is defined by

$$AVaR_\tau(X) = \frac{1}{1-\tau} \int_\tau^1 VaR_t(X)dt.$$

Note that, if X has a continuous distribution, then the $AVaR_\tau(X)$ can be written in the more intuitive form:

$$AVaR_\tau(X) = \mathbb{E}[X | X \geq VaR_\tau(X)],$$

see e.g. Acerbi & Tasche (2002). The aim now is to find for fixed $\tau \in (0,1)$:

$$\inf_{\pi \in \Pi} AVaR_\tau^\pi(C^N | X_0 = x), \tag{2.1}$$

$$\inf_{\pi \in \Pi} AVaR_\tau^\pi(C^\infty | X_0 = x), \tag{2.2}$$

where $AVaR_\tau^\pi$ indicates that the $AVaR_\tau$ is taken w.r.t. the probability measure $\mathbb{P}^\pi$. A policy π^* is called *optimal* for the finite horizon problem if

$$\inf_{\pi \in \Pi} AVaR_\tau^\pi(C^N | X_0 = x) = AVaR_\tau^{\pi^*}(C^N | X_0 = x)$$

and a policy π^* is called *optimal* for the infinite horizon problem if

$$\inf_{\pi \in \Pi} AVaR_\tau^\pi(C^\infty | X_0 = x) = AVaR_\tau^{\pi^*}(C^\infty | X_0 = x).$$

Note that this problem is no longer a standard Markov Decision Problem since the Average-Value-at-Risk is a convex risk measure. However, if we let $\tau \to 0$ then we obtain the usual expectation, i.e.

$$\lim_{\tau \to 0} AVaR_\tau^\pi(C^N | X_0 = x) = \mathbb{E}_x^\pi[C^N]$$

where $\mathbb{E}_x^\pi$ is the expectation with respect to the probability measure $\mathbb{P}_x^\pi$ which is induced by policy π and conditioned on $X_0 = x$. On the other hand, if we let $\tau \to 1$, then we obtain in the limit the *Worst-Case risk measure* which is defined by

$$WC(C^N) := \sup_\omega C^N(\omega).$$

Hence the parameter τ can be seen as a kind of degree of risk aversion. For a discussion of the task of minimizing the Average-Value-at-Risk of the average cost $\limsup_{N\to\infty} \frac{1}{N+1} \sum_{k=0}^N C_k$, see Ott (2010), Chapter 8.

3. Solution of the finite Horizon Problem

For the solution of the problem it is important to note that the Average-Value-at-Risk can be represented as the solution of a convex optimization problem. More precisely, the following lemma is given in Rockafellar & Uryasev (2002).

Lemma 3.1. *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ be a real-valued random variable and let $\tau \in (0,1)$. Then it holds:*

$$AVaR_\tau(X) = \min_{s\in\mathbb{R}} \Big\{ s + \frac{1}{1-\tau}\, \mathbb{E}[(X-s)^+] \Big\}.$$

and the minimum-point is given by $s^* = VaR_\tau(X)$.

Hence we obtain for the problem with finite time horizon:

$$\begin{aligned}\inf_{\pi\in\Pi} AVaR_\tau^\pi(C^N|X_0=x) &= \inf_{\pi\in\Pi}\inf_{s\in\mathbb{R}} \Big\{ s + \frac{1}{1-\tau}\, \mathbb{E}_x^\pi[(C^N-s)^+] \Big\} \\ &= \inf_{s\in\mathbb{R}}\inf_{\pi\in\Pi} \Big\{ s + \frac{1}{1-\tau}\, \mathbb{E}_x^\pi[(C^N-s)^+] \Big\} \\ &= \inf_{s\in\mathbb{R}} \Big\{ s + \frac{1}{1-\tau} \inf_{\pi\in\Pi} \mathbb{E}_x^\pi[(C^N-s)^+] \Big\}.\end{aligned}$$

In what follows we will investigate the inner optimization problem and show that it can be solved with the help of a suitably defined Markov Decision Problem. For this purpose let us denote for $n = 0, 1, \ldots, N$

$$\begin{aligned} w_{n\pi}(x,s) &:= \mathbb{E}_x^\pi[(C^n-s)^+], \quad x\in E, s\in\mathbb{R}, \pi\in\Pi, \\ w_n(x,s) &:= \inf_{\pi\in\Pi} w_{n\pi}(x,s), \quad x\in E, s\in\mathbb{R}. \end{aligned} \tag{3.1}$$

We consider a Markov Decision Model which is given by a 2-dimensional state space $\tilde{E} := E \times \mathbb{R}$, action space A and admissible actions in D. The interpretation of the second component of the state $(x,s) \in \tilde{E}$ will become clear later. It captures the relevant information of the history of the process (see Remark 3.3). Further, there are disturbance variables $Z_n = (Z_n^1, Z_n^2) = (X_n, C_{n-1})$ with values in $E \times \mathbb{R}_+$ which influence the transition. If the state of the Markov Decision Process is (x,s) at time n and action a is chosen, then the distribution of Z_{n+1} is given by the transition kernel $\mathbb{Q}(\cdot \mid x, a)$. The transition function $F : \tilde{E} \times A \times E \times \mathbb{R}_+ \to \tilde{E}$ which determines the new state, is given by

$$F\big((x,s), a, (z_1,z_2)\big) = \big(z_1, \frac{s-z_2}{\beta}\big).$$

The first component of the right-hand side is simply the new state of our original state process and the necessary information update takes place in the second component. There is no running cost and the terminal cost function is given by $V_{-1\pi}(x,s) := V_{-1}(x,s) := s^-$. We consider here decision rules $f : \tilde{E} \to A$ such that $f(x,s) \in D(x)$ and denote by Π^M the set of Markov policies $\sigma = (f_0, f_1, \ldots)$ where f_n are decision rules. Note that 'Markov' refers here to the fact that the decision at time n depends only on x and s. For convenience we denote for $v \in \mathbb{M}(\tilde{E}) := \{v : \tilde{E} \to \mathbb{R}_+ : v \text{ is measurable }\}$ the operators

$$Lv(x,s,a) := \beta \int v\big(x', \frac{s-c}{\beta}\big)\, \mathbb{Q}\big(dx' \times dc|x,a\big), \quad (x,s)\in\tilde{E}, a\in D(x)$$

and

$$T_f v(x,s) := \beta \int v\big(x', \frac{s-c}{\beta}\big)\, \mathbb{Q}\big(dx' \times dc|x,f(x,s)\big), \quad (x,s)\in\tilde{E}.$$

The minimal cost operator of this Markov Decision Model is given by

$$Tv(x,s) = \inf_{a\in D(x)} Lv(x,s,a).$$

For a policy $\sigma = (f_0, f_1, f_2, \dots) \in \Pi^M$ we will denote by $\vec{\sigma} = (f_1, f_2, \dots)$ the shifted policy. We define for $\sigma \in \Pi^M$ and $n = -1, 0, 1, \dots N$:

$$\begin{aligned} V_{n+1\sigma} &:= T_{f_0} V_{n\vec{\sigma}}, \\ V_{n+1} &:= \inf_{\sigma} V_{n+1\sigma} = TV_n. \end{aligned}$$

A decision rule f_n^* with the property that $V_n = T_{f_n^*} V_{n-1}$ is called minimizer of V_n. Next note that we have $\Pi^M \subset \Pi$ in the following sense: For every $\sigma = (f_0, f_1, \dots) \in \Pi^M$ we find a $\pi = (g_0, g_1, \dots) \in \Pi$ such that (the variable s is considered as a global variable)

$$\begin{aligned} g_0(x_0) &:= f_0(x_0, s) \\ g_1(x_0, a_0, c_0, x_1) &:= f_1\big(x_1, \frac{s - c_0}{\beta}\big) \\ \vdots &:= \vdots \end{aligned}$$

With this interpretation $w_{n\sigma}$ is also defined for $\sigma \in \Pi^M$. Note that a policy $\sigma = (f_0, f_1, \dots) \in \Pi^M$ also depends on the history of our process, however in a weak sense. The only necessary information at time n of the history $h_n = (x_0, a_0, c_0, x_1, \dots, a_{n-1}, c_{n-1}, x_n)$ is x_n and the value $\frac{s-c_0}{\beta^n} - \frac{c_1}{\beta^{n-1}} - \dots - \frac{c_{n-1}}{\beta}$. Also note that Π is strictly larger than Π^M: There are history-dependent policies π which cannot be represented as a Markov policy $\sigma \in \Pi^M$. However, it will be shown in Theorem 3.2 that indeed the optimal policy π^* of problem (3.1) (if it exists) can be found among the smaller class Π^M.

The connection of the MDP to the optimization problem in (3.1) is stated in the next theorem.

Theorem 3.2. *It holds for $n = 0, 1, \dots, N$ that*

a) $w_{n\sigma} = V_{n\sigma}$ *for* $\sigma \in \Pi^M$.

b) $w_n = V_n$.

If there exist minimizers f_n^* of V_n on all stages, then the Markov policy $\sigma^* = (f_N^*, \dots, f_0^*)$ is optimal for problem (3.1).

Proof. We first prove that $w_{n\sigma} = V_{n\sigma}$ for all $\sigma \in \Pi^M$. This is done by induction on n. For $n = 0$ we obtain

$$\begin{aligned} V_{0\sigma}(x, s) &= T_{f_0} V_{-1}(x, s) \\ &= \beta \int V_{-1}\big(x', \frac{s - c}{\beta}\big)\, \mathbb{Q}\big(dx' \times dc | x, f_0(x, s)\big) \\ &= \beta \int \big(\frac{s - c}{\beta}\big)^- \, \mathbb{Q}\big(dx' \times dc | x, f_0(x, s)\big) \\ &= \int (c - s)^+ \, \mathbb{Q}\big(dx' \times dc | x, f_0(x, s)\big) \\ &= \mathbb{E}_x^\pi[(C^0 - s)^+] = w_{0\sigma}(x, s). \end{aligned}$$

Next we assume that the statement is true for n and show that it also holds for $n+1$. We obtain

$$\begin{aligned} V_{n+1\sigma}(x, s) &= T_{f_0} V_{n\vec{\sigma}}(x, s) \\ &= \beta \int V_{n\vec{\sigma}}\big(x', \frac{s - c}{\beta}\big)\, \mathbb{Q}\big(dx' \times dc | x, f_0(x, s)\big) \\ &= \beta \int \mathbb{E}_{x'}^{\vec{\sigma}}\, \big[\big(C^n - \frac{s - c}{\beta}\big)^+\big]\, \mathbb{Q}\big(dx' \times dc | x, f_0(x, s)\big) \\ &= \int \mathbb{E}_{x'}^{\vec{\sigma}}\, \big[\big(c + \beta C^n - s\big)^+\big]\, \mathbb{Q}\big(dx' \times dc | x, f_0(x, s)\big) \\ &= \mathbb{E}_x^\sigma[(C^{n+1} - s)^+] = w_{n+1\sigma}(x, s). \end{aligned}$$

Histories of the Markov Decision Process $\tilde{h}_n = (x_0, s_0, a_0, c_0, x_1, s_1, a_1, \dots, x_n, s_n)$ contain the history $h_n = (x_0, a_0, c_0, x_1, a_1, \dots, x_n,)$. We denote by $\tilde{\Pi}$ the history dependent policies of the

Markov Decision Process. Now it is well-known (see e.g. Bäuerle & Rieder (2011) Theorem 2.2.3) that

$$\inf_{\sigma\in\Pi^M} V_{n\sigma}(x,s) = \inf_{\tilde\pi\in\tilde\Pi} V_{n\tilde\pi}(x,s).$$

Thus we obtain by part a)

$$\inf_{\sigma\in\Pi^M} w_{n\sigma} \quad\geq\quad \inf_{\pi\in\Pi} w_{n\pi} \geq \inf_{\tilde\pi\in\tilde\Pi} V_{n\tilde\pi} = \inf_{\sigma\in\Pi^M} V_{n\sigma} = \inf_{\sigma\in\Pi^M} w_{n\sigma}$$

and equality holds which implies the remaining statements. □

Remark 3.3. Note that the optimal policy π^* (if it exists) is Markov. The term 'Markov' refers here to the two-dimensional Markov Decision Process which consists of the system state and the quantity s which is the current threshold beyond which costs matter, i.e. the decision at time point n depends only on the system state at time n and s_n. Recall that s_n is updated in a transition step by $s_{n+1} = \frac{s_n - c_n}{\beta}$. The quantity s_n thus contains the information of the history which is necessary to take a decision and hence can be seen as a 'sufficient statistic'.

Next we impose some assumptions on the model data of the general Markov Decision Process which guarantee that an optimal policy for problem (3.1) exists. Besides the fact that the non-negative cost C_k is bounded from above by a constant $\bar C$ we impose the following assumption.

Assumption (C):

(i) $D(x)$ is compact for all $x \in E$,

(ii) $x \mapsto D(x)$ is upper semicontinuous, i.e. it has the following property for all $x \in E$: If $x_n \to x$ and $a_n \in D(x_n)$ for all $n \in \mathbb{N}$, then (a_n) has an accumulation point in $D(x)$.

(iii) $(x,a) \mapsto \int v\big(x', \frac{s-c}{\beta}\big)\, \mathbb{Q}\left(dx' \times dc | x,a\right)$ is lower semicontinuous for all lower semicontinuous functions $v \geq 0$.

Then the next theorem can be shown.

Theorem 3.4. *Under Assumption (C) there exists an optimal Markov policy σ^* for problem* (3.1).

Proof. In view of Theorem 3.2 we have to show that there exist minimizers for the value functions V_n. But this follows directly from our assumptions and Theorem 2.4.6 in Bäuerle & Rieder (2011). Note that since the cost variables are non-negative we can use $b(x,s) \equiv 1$ as a lower bounding function. □

It is now possible to show some more properties of the value functions V_n. For this purpose, let us define the set

$$\begin{aligned}\mathbb{M} \;:=\; & \Big\{ v : \tilde E \to \mathbb{R}_+ \mid v(x,\cdot) \text{ is non-increasing for } x\in E;\ |v(x,s)-v(x,t)| \leq |s-t|; \\ & \exists\, \tilde c : E \to \mathbb{R}\ s.t.\ v(x,s) = \tilde c(x) - s, \text{ for } s<0 \text{ and } v(x,s)=0 \text{ for } s \text{ large enough}\Big\}.\end{aligned}$$

It is possible to show the following result.

Theorem 3.5. *It holds that:*

a) $T : \mathbb{M} \to \mathbb{M}$.

b) $w_N \in \mathbb{M}$.

Proof. We first prove part a) by showing that if $v \in \mathbb{M}$, the function Tv has the four stated properties. Recall that for $v \in \mathbb{M}$ we have

$$Tv(x,s) = \beta \inf_{a\in D(x)} \int v\big(x', \frac{s-c}{\beta}\big)\, \mathbb{Q}(dx' \times dc | x,a).$$

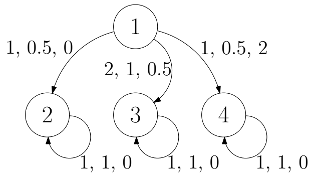

Figure 1. MDP model of Example 3.7.

This definition directly implies that $Tv(x,\cdot)$ is non-increasing if $v(x,\cdot)$ is non-increasing. The Lipschitz property is satisfied since for $s,t\in\mathbb{R}$ and $v\in\mathbb{M}$:

$$\begin{aligned}|Tv(x,s)-Tv(x,t)| &\le \beta \sup_{a\in D(x)} \int \Big| v\big(x',\frac{s-c}{\beta}\big)-v\big(x',\frac{t-c}{\beta}\big)\Big| \mathbb{Q}(dx'\times dc|x,a)\\ &\le \beta \sup_{a\in D(x)} \int \Big| \frac{s-c}{\beta}-\frac{t-c}{\beta}\Big| \mathbb{Q}(dx'\times dc|x,a) = |s-t|.\end{aligned}$$

For the next property note that if $s<0$, then $\frac{s-c}{\beta}<0$. This implies that for $s<0$:

$$\begin{aligned}Tv(x,s) &= \beta \inf_{a\in D(x)} \int \Big(\tilde{c}(x')-\frac{s-c}{\beta}\Big) \mathbb{Q}(dx'\times dc|x,a)\\ &= \inf_{a\in D(x)} \int (\beta\tilde{c}(x')+c)\, \mathbb{Q}(dx'\times dc|x,a)-s.\end{aligned}$$

The last property is obvious since the cost are assumed to be bounded.

Now for part b) note that by Theorem 3.2 and the fact that $V_{n+1}=TV_n$ (see e.g. Bertsekas & Shreve (1978)) it is enough to show that $V_{-1}\in\mathbb{M}$. Since $V_{-1}(x,s)=s^-$ this can be seen directly from the definition of $\mathbb{M}$. □

With the help of these properties it follows now that there is a 'Markov' optimal policy for the $AVaR_\tau$-problem with finite horizon. Consider the problem

$$\inf_{s\in\mathbb{R}}\Big(s+\frac{1}{1-\tau}w_N(x,s)\Big). \tag{3.2}$$

We obtain our next statement.

Theorem 3.6. *There exists a solution s^* of problem* (3.2) *and the optimal policy of problem* (3.1) *with initial state (x,s^*) solves problem* (2.1)*.*

Proof. It is not difficult to see from Theorem 3.5 part b) and the definition of the set $\mathbb{M}$ that for all $x\in E$:

$$\lim_{s\to\infty}\Big(s+\frac{1}{1-\tau}w_N(x,s)\Big)=\infty \quad\text{and}\quad \lim_{s\to-\infty}\Big(s+\frac{1}{1-\tau}w_N(x,s)\Big)=\infty.$$

Hence there exists a number $R(x)\in\mathbb{R}$ such that $K:=\{s\in\mathbb{R}: s+\frac{1}{1-\tau}w_N(x,s)\le R(x)\}\neq\emptyset$. Since $s\mapsto w_N(x,s)$ is continuous (see Theorem 3.5 part b)) it follows that K is compact and problem (3.2) has a solution. The remaining statement follows from the considerations at the beginning of this section. □

Example 3.7. Here, we briefly illustrate that a general $AVaR_\tau$-optimal policy might not be VaR_τ-optimal. The Markov Decision Model is the following: $S=\{1,2,3,4\}$, $A=\{1,2\}$, $D(1)=A$, $D(2)=D(3)=D(4)=\{1\}$. The cost C_n can take the possible values $\{0,0.5,2\}$ and the transition kernel is given by

$$\mathbb{Q}(\{2\}\times\{0\}|1,1)=0.5,\quad \mathbb{Q}(\{4\}\times\{2\}|1,1)=0.5,\quad \mathbb{Q}(\{3\}\times\{0.5\}|1,2)=1.$$

A sketch of this model can be found in Figure 1 where the numbers on the arrows denote the action, the transition probability and the cost respectively. Let $\tau=0.5$, β be arbitrary, and let the initial state be $x_0=1$. Consider the 0-horizon problem, i.e., the decision maker has to

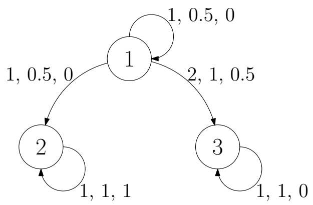

Figure 2. MDP model of Example 3.8.

make exactly one decision. As we have shown, there is a Markov optimal policy to the $AVaR_\tau$-criterion. Consider the two possible policies σ_1 and σ_2, which are defined by the first decision rules $f_0^1(1,s^*) := 1$ and $f_0^2(1,s^*) := 2$. Then we have

$$AVaR_{0.5}^{\sigma_1}(C^0\,|\,X_0=1)=2 \quad\text{and}\quad AVaR_{0.5}^{\sigma_2}(C^0\,|\,X_0=1)=0.5.$$

So, σ_2 is $AVaR_{0.5}$-optimal. But for the Value-at-Risk at level 0.5 of C^0 under σ_1 and under σ_2 respectively we have

$$VaR_{0.5}^{\sigma_1}(C^0\,|\,X_0=1)=0 \quad\text{and}\quad VaR_{0.5}^{\sigma_2}(C^0\,|\,X_0=1)=0.5.$$

Example 3.8. In this example, we demonstrate that the principle of optimality does not hold for the Average-Value-at-Risk criterion when we consider policies which depend only on the current state. We call these policies 'simple'. Let $S=\{1,2,3\}$, $A=\{1,2\}$, $D(1)=A$, $D(2)=D(3)=\{1\}$. The set $\{0,0.5,1\}$ are the possible values of C_n and the transition kernel is given by

$$\begin{aligned}&\mathbb{Q}(\{1\}\times\{0\}|1,1)=0.5,\quad \mathbb{Q}(\{2\}\times\{0\}|1,1)=0.5,\quad \mathbb{Q}(\{3\}\times\{0.5\}|1,2)=1,\\&\mathbb{Q}(\{2\}\times\{1\}|2,1)=1,\quad \mathbb{Q}(\{3\}\times\{0\}|3,1)=1.\end{aligned}$$

A sketch of this model can be found in Figure 2 where the numbers on the arrows denote the action, the transition probability and the cost respectively. Further, let $\tau=0.5$ and $\beta=0.4$. Let us again consider the 0- and the 1-horizon problem for the initial state 1. There are three possible simple policies since there is nothing to decide in states 2 and 3. Define the policies $\sigma^1=(f_0^1,f_1^1,\dots)$, $\sigma^2=(f_0^2,f_1^2,\dots)$ and $\sigma^3=(f_0^3,f_1^3,\dots)$ by

$$\begin{aligned}&f_0^1(1)=1,\quad f_1^1(1)=1,\\&f_0^2(1)=1,\quad f_1^2(1)=2,\\&f_0^3(1)=2.\end{aligned}$$

Then we obtain

$$\begin{aligned}AVaR_{0.5}^{\sigma^1}\left(C^1\;\middle|\;X_0=1\right)&=0.4,\\AVaR_{0.5}^{\sigma^2}\left(C^1\;\middle|\;X_0=1\right)&=0.4,\\AVaR_{0.5}^{\sigma^3}\left(C^1\;\middle|\;X_0=1\right)&=0.5.\end{aligned}$$

Hence, the two policies σ^1 and σ^2 are optimal within the class of simple policies in the 1-horizon case. But for the shifted policies $\vec{\sigma}^1=(f_1^1,\dots)$ and $\vec{\sigma}^2=(f_1^2,\dots)$, we have

$$\begin{aligned}AVaR_{0.5}^{\vec{\sigma}^1}\left(C^0\;\middle|\;X_0=1\right)&=0,\\AVaR_{0.5}^{\vec{\sigma}^2}\left(C^0\;\middle|\;X_0=1\right)&=0.5\end{aligned}$$

and $\vec{\sigma}^2$ is not optimal in the 0-horizon case, which shows that the principle of optimality does not hold for the Average Value-at-Risk criterion within the class of simple policies. This example also shows that the Average Value-at-Risk is not a *time-consistent* optimization criterion (see also the next remark).

Remark 3.9 (Discussion of time-inconsistency of the $AVaR_\tau$-criterion)**.** Risk- sensitive criteria like the $AVaR_\tau$ or mean-variance (see e.g. Li & Ng (2000)) are known to lack the property of time-consistency. This has been discussed among others in Björk & Murgoci (2010), Boda & Filar (2006), Shapiro (2009). However, one has to be careful with the notion of time-consistency, because there are various ways to interpret it.

Here we indeed present a time-consistent interpretation of the $AVaR_\tau$-criterion: First note that choosing the risk level τ corresponds to choosing the parameter s in the representation

$$AVaR_\tau^\pi(C^N|X_0=x) = \min_{s\in\mathbb{R}} \Big\{ s + \frac{1}{1-\tau}\, \mathbb{E}_x^\pi[(C^N-s)^+] \Big\}$$

because the minimum point is given by $s^*(\tau) = VaR_\tau^\pi(C^N|X_0=x)$. Hence as an approximation, our decision maker may fix s instead of τ to choose her risk aversion and simply solve $\inf_\pi \mathbb{E}_x^\pi[(C^N-s)^+]$. The function $x\mapsto (x-s)^+$ may be interpreted as a *disutility function* with a certain parameter s which represents the risk aversion of the decision maker. For this s we compute the optimal policy π^* as in (3.1). The shifted policy $\vec{\pi}^*$ is then optimal for the problem $\inf_\pi \mathbb{E}_{x'}^\pi[(C^{N-1}-\frac{s-C_0}{\beta})^+]$ with new state x' and adapted disutility function $du_{N-1}(x) = (x-\frac{s-C_0}{\beta})^+$. It is next possible to choose (under some assumptions) τ^* such that π^* is optimal for the $AVaR_{\tau^*}$-criterion. Adopting this point of view the optimal policy is time-consistent w.r.t. to the adapted, recursively defined sequence of disutility functions. Also in the sense that optimal decisions do not depend on scenarios which we already know cannot happen in the future. The difference to the point of view in Shapiro (2009) is that our investor chooses the risk aversion parameter s instead of τ which implies that the outer optimization problem can be skipped.

4. Solution of the infinite Horizon Problem

Here we assume that $\beta<1$ and consider problem (2.2). Note that $C^\infty \le \frac{\bar{C}}{1-\beta}$. We can apply the same trick as for the finite horizon problem and obtain

$$\begin{aligned} \inf_{\pi\in\Pi} AVaR_\tau^\pi(C^\infty|X_0=x) &= \inf_{\pi\in\Pi}\inf_{s\in\mathbb{R}} \Big\{ s + \frac{1}{1-\tau}\, \mathbb{E}_x^\pi[(C^\infty-s)^+] \Big\} \\ &= \inf_{s\in\mathbb{R}}\inf_{\pi\in\Pi} \Big\{ s + \frac{1}{1-\tau}\, \mathbb{E}_x^\pi[(C^\infty-s)^+] \Big\} \\ &= \inf_{s\in\mathbb{R}} \Big\{ s + \frac{1}{1-\tau} \inf_{\pi\in\Pi} \mathbb{E}_x^\pi[(C^\infty-s)^+] \Big\}. \end{aligned}$$

Now we define for $\pi\in\Pi$ and $(x,s)\in\tilde{E}$:

$$\begin{aligned} w_{\infty\pi}(x,s) &:= \mathbb{E}_x^\pi[(C^\infty-s)^+], \quad (x,s)\in\tilde{E}, \pi\in\Pi \\ w_\infty(x,s) &:= \inf_{\pi\in\Pi} w_{\infty\pi}(x,s), \quad (x,s)\in\tilde{E}. \end{aligned} \tag{4.1}$$

Since $C^n \le C^{n+1} \le \frac{\bar{C}}{1-\beta}$ $\mathbb{P}^\pi$-a.s. it is not difficult to see that the value functions w_n of the previous section are increasing in n. Thus, the following limit is well-defined

$$w^*(x,s) = \lim_{n\to\infty} w_n(x,s), \quad (x,s)\in\tilde{E}.$$

A first result tells us that we can obtain w_∞ as the limit of the functions w_n.

Theorem 4.1. *It holds that $w^*=w_\infty$.*

Proof. Since costs are non-negative we obtain

$$w_n(x,s) = \inf_{\pi\in\Pi} \mathbb{E}_x^\pi[(C^n-s)^+] \le \inf_{\pi\in\Pi} \mathbb{E}_x^\pi[(C^\infty-s)^+] = w_\infty(x,s).$$

On the other hand it holds for arbitrary $\pi \in \Pi$ (note that $\beta < 1$)

$$\begin{aligned} w_{\infty\pi}(x,s) &= \mathbb{E}_x^\pi[(C^\infty - s)^+] = \mathbb{E}_x^\pi\Big[(C^n + \beta^{n+1}\sum_{k=0}^\infty \beta^k C_{n+k+1} - s)^+\Big] \\ &\le \mathbb{E}_x^\pi\Big[(C^n + \beta^{n+1}\frac{\bar{C}}{1-\beta} - s)^+\Big]. \end{aligned}$$

Since $(a+b)^+ \le a^+ + b$ if $b \ge 0$ this implies

$$w_{\infty\pi}(x,s) \le \mathbb{E}_x^\pi\Big[(C^n - s)^+\Big] + \beta^{n+1}\frac{\bar{C}}{1-\beta}.$$

Taking the infimum over all $\pi \in \Pi$ yields

$$w_\infty \le w_n + \beta^{n+1}\frac{\bar{C}}{1-\beta}.$$

Altogether we have

$$w_n \le w_\infty \le w_n + \beta^{n+1}\frac{\bar{C}}{1-\beta}.$$

Letting $n \to \infty$ yields the statement. □

Next we consider the operator T more closely. First we define the the set $\mathbb{M}^\circ \subset \mathbb{M}$ by setting:

$$\mathbb{M}^\circ := \Big\{v \in \mathbb{M} \mid v(x,s) = 0 \text{ for } s \ge \frac{\bar{C}}{1-\beta}\Big\}.$$

On $\mathbb{M}^\circ$ we define the metric d by

$$d(u,v) := \sup_{x,s} |u(x,s) - v(x,s)|, \quad \text{for } u,v \in \mathbb{M}^\circ.$$

The following properties of d and T hold.

Theorem 4.2. a) *The metric space* $(\mathbb{M}^\circ, d)$ *is complete.*

b) $T : \mathbb{M}^\circ \to \mathbb{M}^\circ$.

c) $d(Tu,Tv) \le \beta d(u,v)$ *for* $u,v \in \mathbb{M}^\circ$.

d) *For an arbitrary decision rule* f*, the operator* T_f *is monotone, i.e* $u \le v$ *for* $u,v \in \mathbb{M}^\circ$ *implies* $T_f u \le T_f v$.

Proof. a) We have to show that every Cauchy sequence in $\mathbb{M}^\circ$ convergence towards an element of $\mathbb{M}^\circ$ w.r.t. the metric d. Now if $(v_n) \subset \mathbb{M}^\circ$ is a Cauchy sequence we can define a limit pointwise by setting $v(x,s) := \lim_{n\to\infty} v_n(x,s)$ for $(x,s) \in \tilde{E}$. Obviously $\lim_{n\to\infty} d(v_n, v) = 0$. Moreover, it is easy to see that v inherits the properties of the sequence (v_n), thus $v \in \mathbb{M}^\circ$.

b) From Theorem 3.5 we already know that $T : \mathbb{M} \to \mathbb{M}$. It remains to show that $v \in \mathbb{M}^\circ$ implies that $Tv(x,s) = 0$ for $s \ge \frac{\bar{C}}{1-\beta}$. Note that we have for all $c \le \bar{C}$:

$$\frac{\bar{C}}{1-\beta} \le \frac{\frac{\bar{C}}{1-\beta} - c}{\beta}.$$

This implies for $v \in \mathbb{M}^\circ$ and $s \ge \frac{\bar{C}}{1-\beta}$ that for all $x' \in E$:

$$0 \le v\Big(x', \frac{s-c}{\beta}\Big) \le v\Big(x', \frac{\frac{\bar{C}}{1-\beta} - c}{\beta}\Big) \le v\Big(x', \frac{\bar{C}}{1-\beta}\Big) = 0.$$

Thus we obtain

$$Tv(x,s) = \beta \inf_{a\in D(x)} \int v\Big(x', \frac{s-c}{\beta}\Big)\, \mathbb{Q}(dx' \times dc|x,a) = 0$$

and the statement is shown.

c) For $v, w \in \mathbb{M}^\circ$ and fixed $(x,s) \in \tilde{E}$ we obtain

$$\begin{aligned} |Tu(x,s) - Tv(x,s)| &\leq \beta \sup_{a \in D(x)} \int \Big| u\big(x', \frac{s-c}{\beta}\big) - v\big(x', \frac{s-c}{\beta}\big) \Big| \, \mathbb{Q}(dx' \times dc | x, a) \\ &\leq \beta \sup_{a \in D(x)} \int d(u,v) \, \mathbb{Q}(dx' \times dc | x,a) = \beta d(u,v). \end{aligned}$$

Taking the supremum over all $(x,s) \in \tilde{E}$ yields the statement.

d) Follows directly from the definition of T_f.

□

Finally we can give a solution of the inner optimization problem (4.1) in the next theorem.

Theorem 4.3. *The value function w_∞ is the unique fixed point of T in $\mathbb{M}^\circ$ and if there exists a decision rule f^* such that $w_\infty = T_{f^*} w_\infty$, then the stationary policy $(f^*, f^*, \dots)$ is optimal for problem* (4.1)*.*

Proof. Since by Theorem 4.1 $w_\infty = \lim_{n \to \infty} T^n V_{-1}$ and since $V_{-1} \in \mathbb{M}^\circ$ it follows directly from Theorem 4.2 and Banach's fixed point theorem that $w_\infty \in \mathbb{M}^\circ$ and w_∞ is the unique fixed point of T. Next note that for all $\pi \in \Pi$ and $(x,s) \in \tilde{E}$:

$$w_{\infty\pi}(x,s) = \mathbb{E}_x^\pi[(C^\infty - s)^+] \geq s^- = V_{-1}(s).$$

Thus we obtain by iterating $w_\infty = T_{f^*} w_\infty$ with Theorem 4.2 part d) that

$$w_\infty = \lim_{n \to \infty} T_{f^*}^n w_\infty \geq \lim_{n \to \infty} T_{f^*}^n V_{-1} \geq \lim_{n \to \infty} T^n V_{-1} = w_\infty.$$

Using monotone convergence we get

$$w_\infty(x,s) = \lim_{n \to \infty} T_{f^*}^n V_{-1}(x,s) = \mathbb{E}_x^{(f^*, f^*, \dots)} \Big[(C^\infty - s)^+ \Big]$$

which yields the optimality of the stationary policy $(f^*, f^*, \dots)$. □

As in the previous section, Assumption (C) implies that a decision rule f^* with the property $w_\infty = T_{f^*} w_\infty$ exists, i.e. f^* is a minimizer of w_∞. The proof is analogous to the proof of Theorem 3.4.

Theorem 4.4. *Under Assumption (C) there exists a decision rule f^* with the property $w_\infty = T_{f^*} w_\infty$.*

Consider now the problem

$$\min_{s \in \mathbb{R}} \Big(s + \frac{1}{1-\tau} w_\infty(x,s) \Big). \tag{4.2}$$

The proof of the following theorem is analogous to the proof of Theorem 3.6.

Theorem 4.5. *There exists a solution s^* of problem* (4.2) *and the optimal stationary policy of problem* (4.1) *with initial state (x, s^*) solves problem* (2.2)*.*

Remark 4.6. The results of the previous sections hold true when the cost C_n can get negative, but are bounded from below by a constant $\underline{C} < 0$. In this case, considering the cost $\tilde{C} := C_n - \underline{C}$ and using the fact that the $AVaR_\tau$ is translation-invariant, i.e.

$$AVaR_\tau(\tilde{C}^N) = AVaR_\tau\Big(C^N - \sum_{k=0}^N \beta^k \underline{C}\Big) = AVaR_\tau(C^N) - \sum_{k=0}^N \beta^k \underline{C}$$

transforms the problem into the one considered here. In the general (unbounded) case suitably integrability conditions have to be imposed.

5. Numerical Example

In this section, we are going to illustrate the results of Section 3 and the influence of the risk aversion parameter τ of the $AVaR_\tau$-criterion by means of a numerical example. We consider the undiscounted case $\beta = 1$. For the given horizon $N \in \mathbb{N}$, we consider N independent identically distributed games. The probability of winning one game is given by $p \in (0, 1)$. We assume that the gambler starts with the certain capital $X_0 \in \mathbb{N}$. Further, let X_{k-1}, $k = 1, \ldots, N$, be the capital of the gambler right before the k-th game. The final capital is denoted by X_N. Before each game, the gambler has to decide how much capital she wants to bet in the following game in order to maximize her risk-adjusted profit.

The formal description of the repeated game follows. The state space is $S = \mathbb{N}_0$. The action space is $A = \mathbb{N}_0$ with the restriction set $D(x) = \{0, 1, \ldots, x\}$, $x \in \mathbb{N}_0$. For $a_k \in D(X_k)$, we have $X_{k+1} = X_k + a_k \cdot Z_{k+1}$, $k = 0, \ldots, N-1$, where $Z_{k+1} = 1$ if the $(k+1)$-th game is won and $Z_{k+1} = -1$ if the $(k+1)$-th game is lost and the $Z_1, \ldots, Z_N$ are independent. The probability of winning one game is $p \in (0, 1)$. The problem formulation is in terms of maximizing the profit. In order to correspond with the previous sections, we reformulate the original problem in terms of minimizing a certain cost. For $x \in \mathbb{N}_0$, the transition kernel takes the following form:

$$\mathbb{Q}(\{x+a\} \times \{c_o - a\} \,|\, x, a) := p \quad \text{and} \quad \mathbb{Q}(\{x-a\} \times \{c_o + a\} \,|\, x, a) := 1 - p, \quad a \in D(x),$$

where $c_o := 2^{N-1} X_0$ such that the one-stage costs remain non-negative for all admissible state-action pairs since c_o is the maximal reward the gambler might receive when she always bets the entire capital. In this manner, the gambler incurs the total cost $N 2^{N-1} X_0 - X_N$, which is essentially the negative of the gambler's final capital. So, we are seeking for policies π^*_τ such that they minimize $AVaR^\pi_\tau(N 2^{N-1} X_0 - X_N \,|\, X_0)$ for $\tau \in (0, 1)$.

Let us assume that $p > 1/2$, i.e. we have a 'superfair' game and $\tau = 0$ so that we have the case of the expected cost criterion. Then it is known that the 'bold' strategy is optimal for any time horizon $N \in \mathbb{N}$, i. e., it is optimal to bet the entire capital at each game.

For our specific numerical example, the probability of winning one game is $p = 0.8$, the starting capital is $X_0 = 5$ and the horizon is $N = 5$ games. In order to derive an optimal policy with respect to the $AVaR_\tau$-criterion, we proceed as proposed in section 3. At first, we compute the functions $w_k(x, \cdot)$ for all $x = 0, \ldots, 2^{N-k} X_0$, $k = 1, \ldots, N$. Then we pick some $s^*(\tau)$ such that it is a minimum point of the function $s \mapsto s + 1/(1-\tau) w_N(X_0, s)$. The $AVaR_\tau$-optimal policy is then given by an optimal policy for problem (3.1) with initial state $(X_0, s^*(\tau))$ and horizon N.

The functions $s \mapsto s + 1/(1-\tau) w_5(5, s)$ are illustrated in Figure 3 for several $\tau \in (0, 1)$. Note that the merely differentiable looking functions $s \mapsto s + 1/(1-\tau) w_5(5, s)$ are indeed piecewise linear, in general non-convex, functions. From Figure 3, we obtain that the minimum point $s^*(\tau)$ increases as τ increases.

Furthermore, we simulated each $AVaR_\tau$-optimal policy 100,000 times where the histograms of the respective final capital can be found in Figure 4. For $\tau = 0.1230$, we obtain that the bold strategy is optimal which is very risky and which can only end up with capital 0 or 160. On the other hand we obtain that it is optimal to never bet anything for $\tau = 0.9750$ so that the final capital is surely 5. The remaining policies are somewhere in between. We observe that the range of possible outcomes decreases as τ increases. Moreover, the probability of ending up with no capital diminishes with increasing τ.

The mean of the $(1-\tau) \cdot 100\,\%$ lowest outcomes of the simulation runs of the final capital, which is an estimator for the AVaR at level τ of the profit, is presented in Table 1 where π^*_τ denotes an $AVaR_\tau$-optimal policy. From Table 1, we obtain that the respective policy is optimal for the τ which it is supposed to be optimal for.

Remark 5.1. Note that the practical computation of the $AVaR_\tau$-optimal policy is quite hard. Following our derivation it is easy to see that the minimum point of $h(s) := s + \frac{1}{1-\tau} w_N(x, s)$ is within the interval $[0, \sup_\omega C^N(\omega)]$ where an evaluation of h at point s means solving one MDP. In our example we have $\sup_\omega C^N(\omega) = N 2^{N-1} X_0 = 400$. The function h we have to minimize is

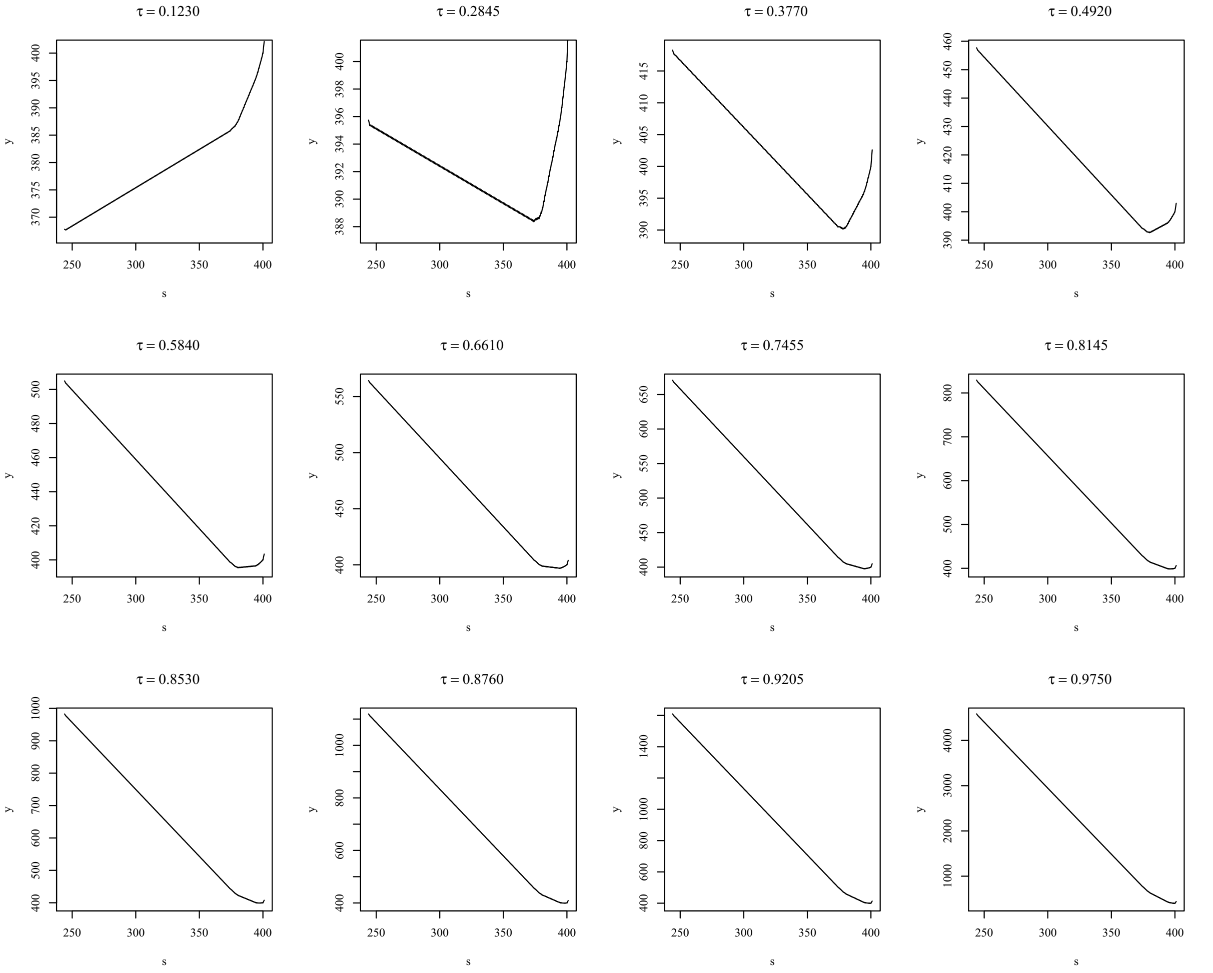

FIGURE 3. Functions $s \mapsto s + 1/(1-\tau)w_5(5,s)$.

τ	0.1230	0.2845	0.3770	0.4920	0.5840	0.6610	0.7455	0.8145	0.8530	0.8760	0.9205	0.9750
$\pi^*_{0.1230}$	36.91	9.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\pi^*_{0.2845}$	19.35	16.72	14.60	11.05	7.41	3.76	0.00	0.00	0.00	0.00	0.00	0.00
$\pi^*_{0.3770}$	18.23	16.25	14.66	11.87	8.81	5.36	0.20	0.00	0.00	0.00	0.00	0.00
$\pi^*_{0.4920}$	17.65	15.99	14.65	12.31	9.50	6.00	0.67	0.13	0.00	0.00	0.00	0.00
$\pi^*_{0.5840}$	17.18	15.65	14.41	12.23	9.63	6.37	1.04	0.00	0.00	0.00	0.00	0.00
$\pi^*_{0.6610}$	9.91	9.67	9.47	9.13	8.71	8.19	7.26	5.95	4.89	4.18	2.03	0.00
$\pi^*_{0.7455}$	9.23	9.06	8.92	8.68	8.39	8.02	7.36	6.38	5.43	4.58	2.61	0.00
$\pi^*_{0.8145}$	8.47	8.35	8.26	8.09	7.88	7.63	7.18	6.50	5.84	5.26	3.17	0.08
$\pi^*_{0.8530}$	7.66	7.58	7.52	7.41	7.28	7.12	6.82	6.38	5.96	5.58	4.23	0.00
$\pi^*_{0.8760}$	6.82	6.78	6.74	6.68	6.61	6.53	6.37	6.13	5.91	5.70	4.98	2.16
$\pi^*_{0.9205}$	5.94	5.93	5.92	5.90	5.88	5.85	5.80	5.72	5.65	5.59	5.36	3.96
$\pi^*_{0.9750}$	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00

TABLE 1. Estimated AVaR of the final capital for the simulated policies.

in general not convex (see Ott (2010), Chapter 7). In our example h is piecewise linear, but this may not be the case in general. On the positive side, we know that h is Lipschitz-continuous with constant $\frac{2-\tau}{1-\tau}$. Hence it is possible to find the minimum point by a suitable bisection procedure.

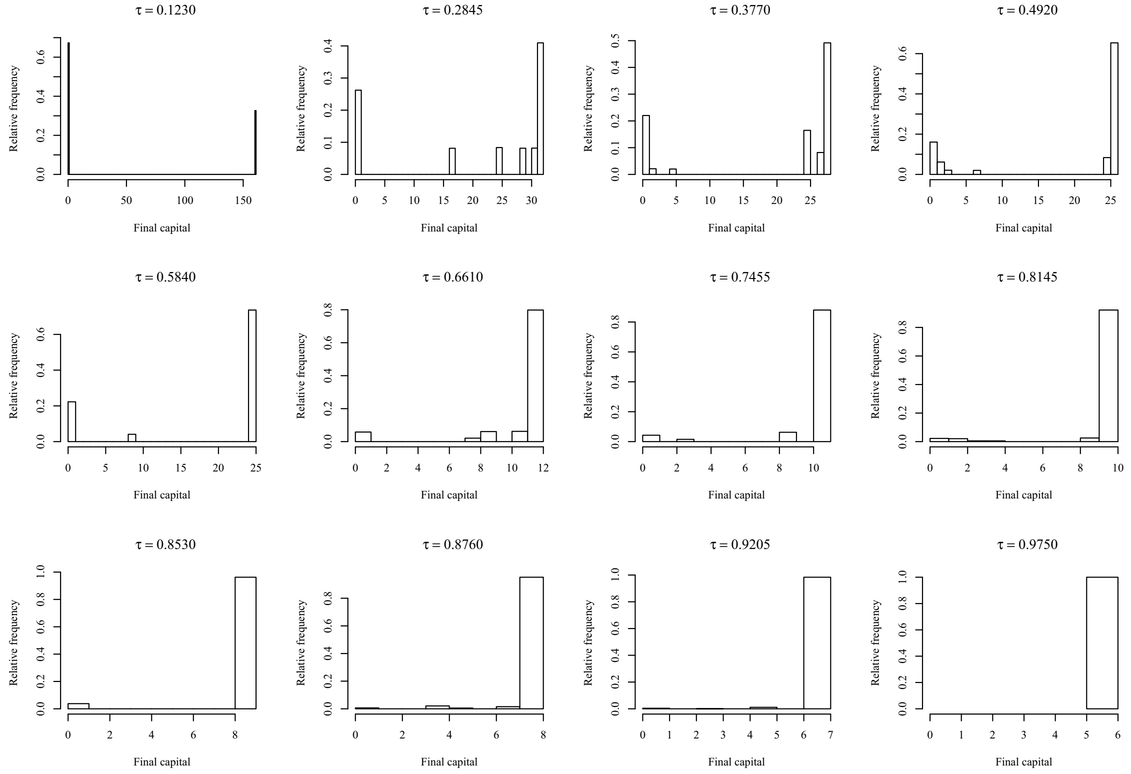

Figure 4. Histograms of the final capital for $AVaR_\tau$-optimal policies.

References

Acerbi, C. & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance* **26**, 1487–1503.

Artzner, P., Delbaen, F., Eber, J., Heath, D. & Ku, H. (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Oper. Res.* **152**, 5–22.

Bäuerle, N. & Mundt, A. (2009). Dynamic mean-risk optimization in a binomial model. *Math. Methods Oper. Res.* **70**, 219–239.

Bäuerle, N. & Rieder, U. (2011). *Markov Decision Processes with applications to finance*. Springer.

Bertsekas, D. P. & Shreve, S. E. (1978). *Stochastic optimal control*. Academic Press, New York.

Bion-Nadal, J. (2008). Dynamic risk measures: Time consistency and risk measures from BMO martingales. *Finance and Stochastics* **12**, 219–244.

Björk, T. & Murgoci, A. (2010). A general theory of Markovian time inconsistent stochastic control problems. *Available at SSRN: http://ssrn.com/abstract=1694759* 1–39.

Boda, K. & Filar, J. (2006). Time consistent dynamic risk measures. *Mathematical Methods of Operations Research* **63**, 169–186.

Boda, K., Filar, J. A., Lin, Y. & Spanjers, L. (2004). Stochastic target hitting time and the problem of early retirement. *IEEE Trans. Automat. Control* **49**, 409–419.

Collins, E. & McNamara, J. (1998). Finite-horizon dynamic optimisation when the terminal reward is a concave functional of the distribution of the final state. *Advances in Applied Probability* **30**, 122–136.

Howard, R. & Matheson, J. (1972). Risk-sensitive Markov Decision Processes. *Management Science* **18**, 356–369.

Jaquette, S. (1973). Markov Decision Processes with a new optimality criterion: discrete time. *Ann. Statist.* **1**, 496–505.

Li, D. & Ng, W.-L. (2000). Optimal dynamic portfolio selection: multiperiod mean-variance formulation. *Math. Finance* **10**, 387–406.

Ott, J. (2010). *A Markov decision model for a surveillance application and risk-sensistive Markov decision processes*. Ph.D. thesis, Karlsruhe Institute of Technology, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000020835.

Rockafellar, R. T. & Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance* **26**, 1443–1471.

Shapiro, A. (2009). On a time consistency concept in risk averse multistage stochastic programming. *Operations Research Letters* **37**, 143–147.

White, D. J. (1988). Mean, variance, and probabilistic criteria in finite Markov Decision Processes: a review. *J. Optim. Theory Appl.* **56**, 1–29.

Wu, C. & Lin, Y. (1999). Minimizing risk models in Markov Decision Processes with policies depending on target values. *J. Math. Anal. Appl.* **231**, 47–67.

(N. Bäuerle) Institute for Stochastics, Karlsruhe Institute of Technology, D-76128 Karlsruhe, Germany

E-mail address: nicole.baeuerle@kit.edu

(J. Ott) Institute for Stochastics, Karlsruhe Institute of Technology, D-76128 Karlsruhe, Germany

E-mail address: jonathan.ott@kit.edu