Math Meth Oper Res (2018) 88:161–184
https://doi.org/10.1007/s00186-018-0633-5

ORIGINAL ARTICLE

Risk measurement and risk-averse control of partially observable discrete-time Markov systems

Jingnan Fan[1] · **Andrzej Ruszczyński**[2]

Received: 22 March 2017 / Published online: 23 February 2018

Abstract We consider risk measurement in controlled partially observable Markov processes in discrete time. We introduce a new concept of conditional stochastic time consistency and we derive the structure of risk measures enjoying this property. We prove that they can be represented by a collection of static law invariant risk measures on the space of function of the observable part of the state. We also derive the corresponding dynamic programming equations. Finally we illustrate the results on a machine deterioration problem.

Keywords Partially observable Markov processes · Dynamic risk measures · Time consistency · Dynamic programming

1 Introduction

The main objective of this paper is to provide theoretical foundations of the theory of dynamic risk evaluation for Markov processes with only partial observation of the state and to derive the corresponding dynamic programming equations.

The theory of dynamic risk measures in discrete time has been intensively developed in the last 10 years (see Scandolo 2003; Riedel 2004; Roorda et al. 2005; Föllmer and Penner 2006; Cheridito et al. 2006; Ruszczyński and Shapiro 2006b; Artzner et al.

✉ Andrzej Ruszczyński
rusz@rutgers.edu

Jingnan Fan
jf546@rutgers.edu

[1] RUTCOR, Rutgers University, Piscataway, NJ 08854, USA

[2] Department of Management Science and Information Systems, Rutgers University, Piscataway, NJ 08854, USA

2007; Pflug and Römisch 2007; Klöppel and Schweizer 2007; Jobert and Rogers 2008; Cheridito and Kupper 2011 and the references therein). The basic setting is the following: we have a probability space $(\Omega, \mathscr{F}, P)$, a filtration $\{\mathscr{F}_t\}_{t=1,\dots,T}$ with a trivial $\mathscr{F}_1$, and we define appropriate spaces $\mathscr{Z}_t$ of $\mathscr{F}_t$-measurable random variables, $t = 1, \dots, T$. For each $t = 1, \dots, T$, a mapping $\rho_{t,T} : \mathscr{Z}_T \to \mathscr{Z}_t$ is called a *conditional risk measure*. The central role in the theory is played by the concept of *time consistency*, which regulates relations between the mappings $\rho_{t,T}$ and $\rho_{s,T}$ for different s and t. One definition employed in the literature (Cheridito et al. 2006; Ruszczyński 2010) is the following: *for all* $Z, W \in \mathscr{Z}_T$, *if* $\rho_{t,T}(Z) \le \rho_{t,T}(W)$ *then* $\rho_{s,T}(Z) \le \rho_{s,T}(W)$ *for all* $s < t$. This can be used to derive recursive relations $\rho_{t,T}(Z) = \rho_t\big(\rho_{t+1,T}(Z)\big)$, with simpler *one-step conditional risk mappings* $\rho_t : \mathscr{Z}_{t+1} \to \mathscr{Z}_t$, $t = 1, \dots, T-1$.

When applied to processes described by controlled kernels, like controlled Markov processes, the theory of dynamic measures of risk encounters difficulties. The domain and range spaces of the one-step mappings ρ_t change, when t increases. With $\mathscr{Z}_t$ containing all $\mathscr{F}_t$-measurable random variables, arbitrary dependence of ρ_t on the past is allowed. These difficulties are compounded by the fact that a control policy changes the probability measure on the space of paths of the process. Risk measurement of the entire family of processes defined by control policies is needed.

In the extant literature, three basic approaches to introduce risk aversion in Markov decision processes have been employed: utility functions (see, e.g., Jaquette 1973; Denardo and Rothblum 1979; Bäuerle and Rieder 2013; Jaśkiewicz et al. 2013), mean-variance models (see, e.g., White 1988; Filar et al. 1989; Mannor and Tsitsiklis 2013; Arlotto et al. 2014), and entropic (exponential) models (see, e.g., Howard and Matheson 1971/72; Marcus et al. 1997; Coraluppi and Marcus 1999; Di Masi and Stettner 1999; Levitt and Ben-Israel 2001). Our approach complements the utility and exponential models; the mean-variance models do not satisfy, in general, the monotonicity and time consistency conditions, except the version of Chen et al. (2014).

In Ruszczyński (2010), we introduced *Markov risk measures*, in which the one-step conditional risk mappings ρ_t have a special form, which allowed for the development of dynamic programming equations and corresponding solution methods. This is related to the expected utility models with an aggregator, considered in Jaśkiewicz et al. (2013), but considers models which are nonlinear in probability (do not have an expected value representation). The aggegator is replaced by a more general *transition risk mapping*. Our ideas were successfully extended to transient models in Çavus and Ruszczyński (2014a, b) and Lin and Marcus (2013) and to problems with unbounded costs in Chu and Zhang (2014) and Shen et al. (2013). These works were further generalized in Fan and Ruszczyński (2016), where we introduced so-called *process-based* measures and described them by a sequence of transition risk mappings: static law invariant risk measures on a space $\mathscr{V}$ of measurable functions on the state space $\mathscr{X}$. In the special case of controlled Markov processes, we derived the structure postulated in Ruszczyński (2010).

In this paper, we develop risk theory for *partially observable* controlled Markov processes. In the expected-value case, this classical topic is covered in many monographs (see, e.g., Hinderer 1970; Bertsekas and Shreve 1978; Bäuerle and Rieder 2011 and the references therein). The standard approach is to consider the *belief state*: the conditional probability distribution of the unobserved part of the state. The recent article (Feinberg et al. 2016) provides the state-of-the-art setting. The risk-averse case has been dealt, so far, with the use of the entropic risk measure (James et al. 1994; Fernández-Gaucherand and Marcus 1997). A more general partially-observable utility model was recently analyzed in Bäuerle and Rieder (2017).

Our main result is that in partially observable systems the dynamic risk measures can be equivalently modeled by special forms of transition risk mappings: static risk measures on the space of functions defined on the observable part of the state only. We also derive dynamic programming equations for risk-averse partially observable Markov models. In these equations, the state space comprises belief states and observable states, as in the expected value model, but the conditional expectation is replaced by a transition risk mapping.

The paper is organized as follows. In Sect. 2, we briefly describe the partially observable Markov process and introduce relevant notation. In Sect. 3, we recall the concept of a dynamic risk measure and review its properties. We introduce the key property of stochastic dynamic time consistency and use it to derive a special structure of risk measures. In Sect. 4 we specialize these results to Markov systems with costs dependent on both components of the current state: observable and unobservable. In Sect. 5, we prove dynamic programming equations for risk-averse partially observable models. Finally, in Sect. 6, we illustrate our approach on a machine replacement problem.

2 Partially observable Markov decision processes

We consider a partially observable Markov process $\{X_t, Y_t\}_{t=1,\dots,T}$, in which $\{X_t\}$ is observable and $\{Y_t\}$ is not. We use the term "partially observable Markov decision process" (POMDP) in a more general way than the extant literature, because we consider dynamic risk measures of the cost sequence, rather than just the expected value.

In order to develop our subsequent theory, it is essential to define the model in a clear and rigorous way (*cf.* Bäuerle and Rieder 2011, Ch. 5). The state space of the model is defined as $\mathcal{X} \times \mathcal{Y}$ where $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ are two Borel spaces (Borel subsets of Polish spaces). From the modeling perspective, $x \in \mathcal{X}$ is the part of the state that we can observe at each step, while $y \in \mathcal{Y}$ is unobservable. The measurable space that we will work with is then given by $\Omega = (\mathcal{X} \times \mathcal{Y})^T$ endowed with the canonical product σ-field $\mathcal{F}$, and we use x_t and y_t to denote the canonical projections at time t.

Let $\{\mathcal{F}_t^{X,Y}\}_{t=1,\dots,T}$ denote the natural filtration generated by the process (X, Y) and $\{\mathcal{F}_t^{X}\}_{t=1,\dots,T}$ be the filtration generated by the process X.

The control space is given by a Borel space $\mathcal{U}$, and since only $\mathcal{X}$ is observable, the set of admissible controls at step t is given by a measurable multifunction $\mathcal{U}_t : \mathcal{X} \rightrightarrows \mathcal{U}$ with nonempty values. The transition kernel at time t is $K_t : \mathcal{X} \times \mathcal{Y} \times \mathcal{U} \to \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, where $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is the space of probability measures on $\mathcal{X} \times \mathcal{Y}$. In other words, if the state at time t is (x, y) and we apply control u, the distribution of the next state is $K_t(x, y, u)$.

At time t, the history of observed states is $h_t = (x_1, x_2, \dots, x_t)$, while all the information available for making a decision is $g_t = (x_1, u_1, x_2, u_2, \dots, x_t)$. We use $\mathscr{H}_t = \mathscr{X}^t = \underbrace{\mathscr{X} \times \cdots \times \mathscr{X}}_{t \text{ times}}$ to denote the spaces of possible histories h_t. We make distinction of g_t and h_t because we should make decision of u_t based on g_t as the past controls $u_1, \dots, u_{t-1}$ are also taken into consideration when estimating the conditional distribution of Y_t (see Sect. 4). We write H_t for $(X_1, \dots, X_t)$.

For this controlled process, a (deterministic) *history-dependent admissible policy* $\pi = (\pi_1, \dots, \pi_T)$ is a sequence of measurable functions $\pi_t(g_t)$ such that $\pi_t(g_t) \in \mathscr{U}_t(x_t)$ for all possible g_t [such a policy exists, due to the measurable selector theorem of Kuratowski and Ryll-Nardzewski (1965)]. We can easily prove by induction on t that for such an admissible policy π, each π_t reduces to a measurable function of $h_t = (x_1, x_2, \dots, x_t)$, as $u_s = \pi_s(x_1, \dots, x_s)$ for all $s = 1, \dots, t-1$. We are still using π_s to denote the decision rule; it will not lead to any misunderstanding. Therefore the set of admissible policies is

$$\Pi = \left\{ \pi = (\pi_1, \dots, \pi_T) \mid \pi_t(x_1, \dots, x_t) \in \mathscr{U}_t(x_t),\ t = 1, \dots, T \right\}.$$

For a random Y_1, any policy $\pi \in \Pi$ defines a process $\{X_t, Y_t, U_t\}_{t=1,\dots,T}$ on the probability space $(\Omega, \mathscr{F}, P^\pi)$, with $U_t = \pi_t(X_1, \dots, X_t)$.

We assume that the cost process Z_t^π, $t = 1, \dots, T$, is bounded and adapted to $\{\mathscr{F}_t^X\}$, i.e., $Z_t^\pi \in \mathscr{Z}_t$ for all π and t, where

$$\mathscr{Z}_t = \left\{ Z : \Omega \to \mathbb{R} \,\middle|\, Z \text{ is } \mathscr{F}_t^X\text{-measurable and bounded} \right\}, \quad t = 1, \dots, T.$$

We allow the cost process to depend on the policy π in order to cover the case of control-dependent costs, such as $Z_t^\pi = c_t(X_t, \pi_t(X_t))$, or more general cases discussed in section. For any $Z \in \mathscr{Z}_t$, a measurable and bounded functional $\overline{Z} : \mathscr{X}^t \to \mathbb{R}$ exists such that $Z = \overline{Z}(X_1, \dots, X_t)$. With an abuse of notation, we still use Z to denote this function.

3 Risk measures for partially observable systems

3.1 Dynamic risk measures

For any policy $\pi \in \Pi$, our objective is to evaluate at each time t the sequence of costs $Z_t^\pi, \dots, Z_T^\pi$ in such a way that the evaluation is $\mathscr{F}_t^X$-measurable. We denote $\mathscr{Z}_{t,T} = \mathscr{Z}_t \times \cdots \times \mathscr{Z}_T$, $t = 1, \dots, T$.

In what follows, all equality and inequality relations between random variables are understood in the "everywhere" sense.

Definition 1 A mapping $\rho_{t,T} : \mathscr{Z}_{t,T} \to \mathscr{Z}_t$, where $1 \le t \le T$, is called a *conditional risk measure*, if it satisfies the *monotonicity property*: for all $(Z_t, \dots, Z_T)$ and $(W_t, \dots, W_T)$ in $\mathscr{Z}_{t,T}$, if $Z_s \le W_s$ for all $s = t, \dots, T$, then $\rho_{t,T}(Z_t, \dots, Z_T) \le \rho_{t,T}(W_t, \dots, W_T)$.

Definition 2 A conditional risk measure $\rho_{t,T} : \mathscr{Z}_{t,T} \to \mathscr{Z}_t$

(i) is *normalized* if $\rho_{t,T}(0, \dots, 0) = 0$;
(ii) is *translation invariant* if $\forall (Z_t, \dots, Z_T) \in \mathscr{Z}_{t,T}$,
$\rho_{t,T}(Z_t, \dots, Z_T) = Z_t + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T)$;
(iii) has the *local property* if for any event $A \in \mathscr{F}_t^X$ and all $(Z_t, \dots, Z_T) \in \mathscr{Z}_{t,T}$ we have $\mathbb{1}_A \rho_{t,T}(Z_t, \dots, Z_T) = \rho_{t,T}(\mathbb{1}_A Z_t, \dots, \mathbb{1}_A Z_T)$.

From now on, we assume all conditional risk measures to be at least normalized.

Definition 3 A *dynamic risk measure* $\big\{\rho_{t,T}\big\}_{t=1,\dots,T}$ is a sequence of conditional risk measures $\rho_{t,T} : \mathscr{Z}_{t,T} \to \mathscr{Z}_t$. We say that it is normalized, translation-invariant, decomposable, or has the local property, if all $\rho_{t,T}$, $t = 1, \dots, T$, satisfy the respective conditions of Definition 2.

3.2 Stochastic conditional time consistency

For a partially observable controlled process defined in Sect. 2, we have to use a family of risk measures $\big\{\rho^{\pi}_{t,T}\big\}^{\pi\in\Pi}_{t=1,\dots,T}$, because the policy affects the probability measure on the space Ω. When two policies π and π' are compared, even if the resulting costs were pointwise equal, $\rho^{\pi}_{t,T}(Z_t, \dots, Z_T)$ and $\rho^{\pi'}_{t,T}(Z_t, \dots, Z_T)$ may differ, because the probability measures P^{π} and $P^{\pi'}$ could differ. The concept of stochastic conditional time consistency allows us to relate the whole family of risk measures.

Definition 4 A family of dynamic risk measures $\big\{\rho^{\pi}_{t,T}\big\}^{\pi\in\Pi}_{t=1,\dots,T}$ is *stochastically conditionally time consistent* if for any $\pi, \pi' \in \Pi$, for any $1 \le t < T$, for all $h_t \in \mathscr{X}^t$, all $(Z_{t+1}, \dots, Z_T) \in \mathscr{Z}_{t+1,T}$ and all $(W_{t+1}, \dots, W_T) \in \mathscr{Z}_{t+1,T}$, the condition

$$\big(\rho^{\pi}_{t+1,T}(Z_{t+1}, \dots, Z_T) \mid H^{\pi}_t = h_t\big) \preceq_{\text{st}} \big(\rho^{\pi'}_{t+1,T}(W_{t+1}, \dots, W_T) \mid H^{\pi'}_t = h_t\big),$$

implies

$$\rho^{\pi}_{t,T}(0, Z_{t+1}, \dots, Z_T)(h_t) \le \rho^{\pi'}_{t,T}(0, W_{t+1}, \dots, W_T)(h_t).$$

Remark 1 The conditional stochastic order "$\preceq_{\text{st}}$" means that for all $\eta \in \mathbb{R}$ we have

$$\begin{aligned}&\mathbb{P}^{\pi}\big[\rho^{\pi}_{t+1,T}(Z_{t+1}, \dots, Z_T)(H_{t+1}) \le \eta \big| H_t = h_t\big]\\&\quad\le \mathbb{P}^{\pi'}\big[\rho^{\pi'}_{t+1,T}(W_{t+1}, \dots, W_T)(H_{t+1}) \le \eta \big| H_t = h_t\big].\end{aligned}$$

Proposition 1 *If a family of dynamic risk measures* $\big\{\rho^{\pi}_{t,T}\big\}^{\pi\in\Pi}_{t=1,\dots,T}$ *is normalized, has the translation property, and is stochastically conditionally time consistent, then it has the local property.*

Proof We use induction on t from T down to 1. At the final time, for all $A \in \mathscr{F}_T$ and all $Z_T \in \mathscr{Z}_T$ we have $\rho^{\pi}_{T,T}(\mathbb{1}_A Z_T) = \mathbb{1}_A Z_T = \mathbb{1}_A \rho^{\pi}_{T,T}(Z_T)$.

Suppose $\rho^{\pi}_{t+1,T}$ has the local property. Then for all $A \in \mathscr{F}_t$, by translation,

$$\rho^{\pi}_{t,T}(\mathbb{1}_A Z_t, \dots, \mathbb{1}_A Z_T) = \mathbb{1}_A Z_t + \rho^{\pi}_{t,T}(0, \mathbb{1}_A Z_{t+1}, \dots, \mathbb{1}_A Z_T)$$

and

$$\mathbb{1}_A \rho^{\pi}_{t,T}(Z_t, \dots, Z_T) = \mathbb{1}_A Z_t + \mathbb{1}_A \rho^{\pi}_{t,T}(0, Z_{t+1}, \dots, Z_T).$$

To verify the local property of $\rho^{\pi}_{t,T}$ we need to show that both right hand sides are equal. For any $h_t \in \mathscr{X}^t$ we have

$$[(\mathbb{1}_A \rho^{\pi}_{t,T}(0, Z_{t+1}, \dots, Z_T))](h_t) = \mathbb{1}_A(h_t)\rho^{\pi}_{t,T}(0, Z_{t+1}, \dots, Z_T))(h_t).$$

The local property of $\rho^{\pi}_{t+1,T}$ yields

$$\rho^{\pi}_{t+1,T}(\mathbb{1}_A Z_{t+1}, \mathbb{1}_A Z_{t+2}, \dots, \mathbb{1}_A Z_T)(h_t, \cdot) = \mathbb{1}_A(h_t)\rho^{\pi}_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, \cdot),$$

so by stochastic conditional time consistency,

$$\rho^{\pi}_{t,T}(0, \mathbb{1}_A Z_{t+1}, \dots, \mathbb{1}_A Z_T)(h_t) = \begin{cases} 0 & \text{if } \mathbb{1}_A(h_t) = 0, \\ \rho^{\pi}_{t,T}(0, Z_{t+1}, \dots, Z_T)(h_t) & \text{if } \mathbb{1}_A(h_t) = 1. \end{cases}$$

Thus,

$$\rho^{\pi}_{t,T}(0, \mathbb{1}_A Z_{t+1}, \dots, \mathbb{1}_A Z_T)(h_t) = \mathbb{1}_A \rho^{\pi}_{t,T}(0, Z_{t+1}, \dots, Z_T)(h_t), \ \forall h_t \in \mathscr{X}^t,$$

which proves the local property of $\rho^{\pi}_{t,T}$. □

The following theorem shows that the stochastic conditional time consistency implies that one-step risk mappings can be represented by static law-invariant risk measures on $\mathscr{V}$, the set of all bounded measurable functions on $\mathscr{X}$. We first recall the definition of a risk measure and slightly refine the standard concept of law invariance.

Definition 5 A measurable functional $r : \mathscr{V} \to \mathbb{R}$ is called a risk measure.

(i) It is *monotonic*, if $V \le W$ implies $r(V) \le r(W)$;
(ii) It is *normalized* if $r(0) = 0$;
(iii) It is *translation invariant* if for all $V \in \mathscr{V}$ and all $a \in \mathbb{R}$, $r(a + V) = a + r(V)$;
(iv) It is *law invariant with respect to the probability measure* q on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, if $V \overset{q}{\sim} W \Rightarrow r(V) = r(W)$, where $V \overset{q}{\sim} W$ means that $q\{V \le \eta\} = q\{W \le \eta\}$ for all $\eta \in \mathbb{R}$.

The conditional distribution of $\rho^{\pi}_{t+1,T}(Z_{t+1}, \dots, Z_T)(H_{t+1})$ given $H_t = h_t$ under P^{π} plays an important role in the stochastic conditional time consistency, so does the conditional distribution of X_{t+1}, given h_t. We denote the latter by $Q^{\pi}_t(h_t) \in \mathscr{P}(\mathscr{X})$:

$$Q^{\pi}_t(h_t)(C) = \mathbb{P}^{\pi}[X_{t+1} \in C \mid H_t = h_t], \quad \forall C \in \mathscr{B}(\mathscr{X}). \tag{1}$$

Later in Sect. 4 we show that Q_t^π can be computed in a recursive way with the help of belief states and Bayes operators. We can now state the main result of this section.

Theorem 1 *A family of dynamic risk measures $\left\{\rho_{t,T}^\pi\right\}_{t=1,\dots,T}^{\pi\in\Pi}$ is normalized, translation invariant, and stochastically conditionally time consistent if and only if transition risk mappings*

$$\sigma_t : \left\{ \bigcup_{\pi\in\Pi} \operatorname{graph}(Q_t^\pi) \right\} \times \mathscr{V} \to \mathbb{R}, \quad t = 1\dots T-1,$$

exist, such that

(i) *For all $t = 1\dots T-1$ and all $h_t \in \mathscr{X}^t$, $\sigma_t(h_t, \cdot, \cdot)$ is normalized and has the following property of strong monotonicity with respect to stochastic dominance:*

$$\begin{aligned}&\forall q^1, q^2 \in \left\{ Q_t^\pi(h_t) : \pi \in \Pi \right\},\ \forall V^1, V^2 \in \mathscr{V},\\ &(V^1; q^1) \preceq_{\text{st}} (V^2; q^2) \implies \sigma_t(h_t, q^1, V^1) \le \sigma_t(h_t, q^2, V^2),\end{aligned}$$

where $(V; q) = q \circ V^{-1}$ means "the distribution of V under q;"

(ii) *For all $\pi \in \Pi$, for all $t = 1\dots T-1$, for all $(Z_t, \dots, Z_T) \in \mathscr{Z}_{t,T}$, and for all $h_t \in \mathscr{X}^t$,*

$$\rho_{t,T}^\pi(Z_t, Z_{t+1}, \dots, Z_T)(h_t) = Z_t + \sigma_t(h_t, Q_t^\pi(h_t), \rho_{t+1,T}^\pi(Z_{t+1}, \dots, Z_T)(h_t, \cdot)). \tag{2}$$

Moreover, for all $t = 1\dots T-1$, σ_t is uniquely determined by $\{\rho_{t,T}^\pi\}^{\pi\in\Pi}$ as follows: for every $h_t \in \mathscr{X}^t$, for every $q \in \left\{ Q_t^\pi(h_t) : \pi \in \Pi \right\}$, and for every $V \in \mathscr{V}$,

$$\sigma_t(h_t, q, V) = \rho_{t,T}^\pi(0, v, 0, \dots, 0)(h_t),$$

where π is any admissible policy such that $q = Q_t^\pi(h_t)$, and $v \in \mathscr{S}_{t+1}$ satisfies the equation $v(h_t, \cdot) = V(\cdot)$, and can be arbitrary elsewhere.

Proof Assume $\left\{\rho_{t,T}^\pi\right\}_{t=1,\dots,T}^{\pi\in\Pi}$ is translation invariant and stochastically conditionally time consistent. For any $V \in \mathscr{V}$ and any $h_t \in \mathscr{X}^t$ we define $v(h_t, \cdot) = V(\cdot)$. The function v is an element of $\mathscr{S}_{t+1}$. Then the formula $\sigma_t^\pi(h_t, q, V) = \rho_{t,T}^\pi(0, v, 0, \dots, 0)(h_t)$, defines for each π a normalized and monotonic risk measure on the space $\mathscr{V}$. For any $(Z_t, \dots, Z_T) \in \mathscr{Z}_{t,T}$, setting

$$\begin{aligned}W(x) &= \rho_{t+1,T}^\pi(Z_{t+1}, \dots, Z_T)(h_t, x), \quad \forall x \in \mathscr{X},\\ w(h_{t+1}) &= \begin{cases} W(x), & \text{if } h_{t+1} = (h_t, x),\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

we obtain, by translation invariance and normalization,

$$\rho_{t+1,T}^\pi(w, 0, \dots, 0)(h_t, \cdot) = w(h_t, \cdot) = \rho_{t+1,T}^\pi(Z_{t+1}, \dots, Z_T)(h_t, \cdot).$$

Thus, by translation invariance and stochastic conditional time consistency,

$$\begin{aligned}\rho^{\pi}_{t,T}(Z_t,\dots,Z_T)(h_t) &= Z_t(h_t) + \rho^{\pi}_{t,T}(0, Z_{t+1},\dots,Z_T)(h_t)\\ &= Z_t(h_t) + \rho^{\pi}_{t,T}(0, w, 0,\dots,0)(h_t) = Z_t(h_t) + \sigma^{\pi}_t(h_t, q, W).\end{aligned} \tag{3}$$

This chain of relations proves also the uniqueness of σ^{π}_t for each π.

We can now verify the strong monotonicity of $\sigma^{\pi}_t(h_t,\cdot,\cdot)$ with respect to stochastic dominance. Suppose

$$(V^1; Q^{\pi_1}_t(h_t)) \preceq_{\text{st}} (V^2; Q^{\pi_2}_t(h_t)), \tag{4}$$

where $V^1, V^2 \in \mathscr{V}$ and $h_t \in \mathscr{X}^t$. Define $v^1(h_t,\cdot) = V^1(\cdot)$ and $v^2(h_t,\cdot) = V^2(\cdot)$. Then Definition 4 implies that $\rho^{\pi_1}_{t,T}(0, v^1, 0,\dots,0)(h_t) \le \rho^{\pi_2}_{t,T}(0, v^2, 0,\dots,0)(h_t)$. This combined with (3) yields

$$\sigma^{\pi_1}_t(h_t, Q^{\pi_1}_t(h_t), V^1) \le \sigma^{\pi_2}_t(h_t, Q^{\pi_2}_t(h_t), V^2). \tag{5}$$

Suppose $Q^{\pi_1}_t(h_t) = Q^{\pi_2}_t(h_t)$ and $V^1 = V^2$. Then both $\preceq_{\text{st}}$ and $\succeq_{\text{st}}$ are true in (4) and thus (5) becomes an equality. This proves that in fact σ^{π}_t does not depend on π, and all dependence on π is carried by the controlled kernel Q^{π}_t. Moreover, the function $\sigma_t(h_t,\cdot,\cdot)$ is indeed strongly monotonic with respect to stochastic dominance.

On the other hand, if such transition risk mappings σ_t exist, then $\big\{\rho_{t,T}\big\}^{\pi\in\Pi}_{t=1,\dots,T}$ is stochastically conditionally time consistent by the monotonicity and law invariance of $\sigma_t(h_t,\cdot,\cdot)$. We can now use (2) to obtain for any $t = 1,\dots,T-1$, and for all $h_t \in \mathscr{X}^t$ the translation invariance of $\rho^{\pi}_{t,T}$. □

The following transition risk mappings satisfy the condition of Theorem 1 and correspond to stochastically conditionally time consistent risk measures.

Example 1 The entropic transition risk mapping,

$$\sigma_t(h_t, q, v) = \frac{1}{\gamma}\ln\big(\mathbb{E}_q[e^{\gamma v}]\big) = \frac{1}{\gamma}\ln\bigg(\int_{\mathscr{X}} e^{\gamma v(x)}\, q(dx)\bigg), \quad \gamma > 0.$$

We could make γ dependent on the time t, the current state x_t, or the entire history h_t, and still obtain a stochastically conditionally time consistent risk measure (with $q = Q_t(h_t)$).

Example 2 The mean-semideviation transition risk mapping:

$$\sigma_t(h_t, q, v) = \int_{\mathscr{X}} v(s)\, q(ds) + \varkappa_t(h_t)\bigg(\int_{\mathscr{X}} \Big[\Big(v(s) - \int_{\mathscr{X}} v(s')\, q(ds')\Big)_+\Big]^p q(ds)\bigg)^{1/p},$$

where $\varkappa_t : \mathscr{X}^t \to [0,1]$ is a measurable function, and $p \in [1,+\infty)$. It is an analogue of the static mean-semideviation measure of risk, whose consistency with stochastic dominance is well-known (Ogryczak and Ruszczyński 1999, 2001). In the risk measure, we use $q = Q_t(h_t)$.

In fact, all coherent or convex law invariant risk measures may be used to construct σ_t; just the dependence on the probability measure q must be explicit.

4 Markov models

4.1 Bayes operator

At each time t, the conditional distribution of the next observable state $Q_t^\pi(h_t)$ defined in (1) can be easily computed if we know the conditional distribution of the current unobservable state, called the *belief state*:

$$\Xi_t^\pi(h_t) \in \mathscr{P}(\mathscr{Y}) : \ \Xi_t^\pi(h_t)(D) = \mathbb{P}^\pi[Y_t \in D \mid H_t = h_t], \quad \forall D \in \mathscr{B}(\mathscr{Y}), \tag{6}$$

as we have

$$Q_t^\pi(h_t) = \int_{\mathscr{Y}} K_t^X(x_t, y, \pi_t(h_t)) \ \Xi_t^\pi(h_t)(dy), \tag{7}$$

where $K_t^X(x_t, y, \pi_t(h_t))$ is the marginal distribution of $K_t(x_t, y, \pi_t(h_t))$ on $\mathscr{X}$.

In a POMDP, the Bayes operator provides a way to update from prior belief to posterior belief. Suppose the current state observation is x, the action is u, and the conditional distribution of the unobservable state, given the history of the process, is ξ. After a new observation x' of the observable part of the state, we can find a formula to determine the posterior distribution of the unobservable state.

Let us start with a fairly general construction of the Bayes operator. Assuming the above setup, for given $(x, \xi, u) \in \mathscr{X} \times \mathscr{P}(\mathscr{Y}) \times \mathscr{U}$, define a new measure $m_t(x, \xi, u)$ on $\mathscr{X} \times \mathscr{Y}$, initially on all measurable rectangles $A \times B$, as

$$m_t(x, \xi, u)(A \times B) = \int_{\mathscr{Y}} K_t(A \times B \mid x, y, u) \ \xi(dy).$$

We verify readily that this uniquely defines a probability measure on $\mathscr{X} \times \mathscr{Y}$. If the measurable space $(\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ is standard Borel, i.e., isomorphic to a Borel subspace of $\mathbb{R}$, we can disintegrate $m_t(x, \xi, u)$ into its marginal $\lambda_t(x, \xi, u)(dx')$ on $\mathscr{X}$ and a transition kernel $\Gamma_t(x, \xi, u)(x', dy')$ from $\mathscr{X}$ to $\mathscr{Y}$:

$$m_t(x, \xi, u)(dx', dy') = \lambda_t(x, \xi, u)(dx') \ \Gamma_t(x, \xi, u)(x', dy').$$

For all $C \in \mathscr{B}(\mathscr{Y})$, we define the *Bayes operator* of the POMDP as follows:

$$\Phi_t(x, \xi, u, x')(C) = \Gamma_t(x, \xi, u)(x', C).$$

The above argument shows that the Bayes operator exists and is unique as long as the space $\mathscr{Y}$ is standard Borel, which is almost always the case in applications of POMDP. In the following, we always assume the existence of the Bayes operator.

Example 3 Assume that each transition kernel $K_t(x, y, u)$ has a density $q_t(\cdot, \cdot \mid x, y, u)$ with respect to a finite product measure $\mu_X \otimes \mu_Y$ on $\mathscr{X} \times \mathscr{Y}$. Then the Bayes operator has the form

$$\big[\Phi_t(x, \xi, u, x')\big](A) = \frac{\int_A \int_{\mathscr{Y}} q_t(x', y' \mid x, y, u) \ \xi(dy) \ \mu_Y(dy')}{\int_{\mathscr{Y}} \int_{\mathscr{Y}} q_t(x', y' \mid x, y, u) \ \xi(dy) \ \mu_Y(dy')}, \quad \forall A \in \mathscr{B}(\mathscr{Y}).$$

If the formula above has a zero denominator for some (x, ξ, u, x'), we can formally define $\Phi_t(x, \xi, u, x')$ to be an arbitrarily selected distribution on $\mathscr{Y}$.

Thus, we can calculate Q_t and Ξ_t (defined in (1) and (6)) recursively with the help of Bayes operators:

- The initial belief $\Xi_1(x_1)$ is the conditional distribution of Y_1 given $X_1 = x_1$;
- The Bayes operator provides us the following formula to update the belief states:

$$\Xi^\pi_{t+1}(h_{t+1}) = \Phi_t\big(x_t, \Xi^\pi_t(h_t), \pi_t(h_t), x_{t+1}\big),$$

  and, by induction on t, $\Xi^\pi_t(h_t) = \Xi_t\big(x_1, \pi_1(x_1), \dots, x_{t-1}, \pi_{t-1}(h_{t-1}), x_t\big)$;
- The conditional distribution of X_{t+1} can be calculated by (7), and

$$Q^\pi_t(h_t) = Q_t\big(x_1, \pi_1(x_1), \dots, x_t, \pi_t(h_t)\big).$$

4.2 Markov risk measures

We make the following additional assumptions:

Assumption 1 The costs $Z_1, \dots, Z_T$ are only dependent on the current observable state, the current belief state, and the current control, that is,

$$Z^\pi_t(h_t) = r_t(x_t, \Xi^\pi_t(h_t), \pi_t(h_t)), \quad t = 1, \dots, T. \tag{8}$$

For example, in expected value models, we have $r_t(x, \xi, u) = \int_{\mathscr{Y}} c_t(x, y, u)\, \xi(dy)$, where $c_t : \mathscr{X} \times \mathscr{Y} \times \mathscr{U} \to \mathbb{R}$ is the running cost function, but more general functionals can be used here instead of the expectation with respect to the belief state.

Definition 6 In POMDP, a policy $\pi \in \Pi$ is *Markov* if $\pi_t(h_t) = \pi_t(h'_t)$ for all $t = 1, \dots, T$ and all $h_t,\ h'_t \in \mathscr{X}^t$ such that $x_t = x'_t$ and $\Xi^\pi_t(h_t) = \Xi^\pi_t(h'_t)$.

For a fixed Markov policy π, the future evolution of the process $\{(X_\tau, \Xi^\pi_\tau)\}_{\tau=t,\dots,T}$ is solely dependent on the current $(x_t, \Xi^\pi_t(h_t))$, and so is the distribution of the future risk functions $r_\tau(X_\tau, \Xi^\pi_\tau, \pi_\tau(X_\tau, \Xi^\pi_\tau))$, $\tau = t, \dots, T$. Therefore, we can define the Markov property of risk measures for POMDP. To alleviate notation, for all $\pi \in \Pi$ and for a measurable and bounded $r = (r_1, \dots, r_T)$, we write

$$v^\pi_t(h_t) := \rho^\pi_{t,T}\big(r_t(X_t, \Xi^\pi_t, \pi_t(H_t)), \dots, r_T(X_T, \Xi^\pi_T, \pi_T(H_T))\big)(h_t). \tag{9}$$

Definition 7 A family of dynamic risk measures $\big\{\rho^\pi_{t,T}\big\}^{\pi\in\Pi}_{t=1,\dots,T}$ for a POMDP is *Markov* if for all Markov policies $\pi \in \Pi$, for all bounded measurable $r = (r_1, \dots, r_T)$, and for all $h_t = (x_1, \dots, x_t)$ and $h'_t = (x'_1, \dots, x'_t)$ in $\mathscr{X}^t$ such that $x_t = x'_t$ and $\Xi^\pi_t(h_t) = \Xi^\pi_t(h'_t)$, we have

$$v^\pi_t(h_t) = v^\pi_t(h'_t).$$

Proposition 2 *A normalized, translation invariant, and stochastically conditionally time consistent family of risk measures $\{\rho_{t,T}^{\pi}\}_{t=1,\dots,T}^{\pi\in\Pi}$ is Markov if and only if the dependence of σ_t on h_t is carried by $(x_t, \Xi_t^{\pi}(h_t))$ only, for all $t = 1, \dots, T-1$.*

Proof Fix $t = 1, \dots, T-1$ and $W \in \mathcal{V}$. Let $\pi \in \Pi$ be an arbitrary policy. Consider $h_t, h_t' \in \mathcal{X}^t$ such that $x_t = x_t'$, $\Xi_t^{\pi}(h_t) = \Xi_t^{\pi}(h_t') = \xi_t$, and $Q_t^{\pi}(h_t) = Q_t^{\pi}(h_t')$. We construct a Markov policy $\lambda \in \Pi$, such that $\pi_t(h_t) = \lambda_t(x_t, \xi_t)$ for this particular t and h_t. By construction, $Q_t^{\pi}(h_t) = Q_t^{\lambda}(x_t, \xi_t)$. We also construct a sequence of costs $r = (0, \dots, 0, r_{t+1}, 0, \dots, 0)$ with $r_{t+1}(x', \xi', u') \equiv W(x')$. We obtain the following chain of equations, in which we use the construction of λ, the construction of r, the Markov property of Definition 7, and the assumed equality of $Q_t^{\pi}(h_t)$ and $Q_t^{\pi}(h_t')$:

$$\begin{aligned}\sigma_t(h_t, Q_t^{\pi}(h_t), W) &= \sigma_t(h_t, Q_t^{\lambda}(x_t, \xi_t), W) = v_t^{\lambda}(h_t) = v_t^{\lambda}(h_t')\\ &= \sigma_t(h_t', Q_t^{\lambda}(x_t, \xi_t), W) = \sigma_t(h_t', Q_t^{\pi}(h_t), W)\\ &= \sigma_t(h_t', Q_t^{\pi}(h_t'), W).\end{aligned}$$

Therefore, σ_t is its direct dependence on h_t is carried by (x_t, ξ_t) only.

If $\sigma_t,\ t = 1, \dots, T-1$, are all memoryless, we can prove by induction backward in time that for all $t = T, \dots, 1$, $v_t^{\pi}(h_t) = v_t^{\pi}(h_t')$ for all Markov π and all $h_t, h_t' \in \mathcal{X}^t$ such that $x_t = x_t'$ and $\xi_t = \xi_t'$. □

The following theorem summarizes our observations.

Theorem 2 *A family of dynamic risk measures $\{\rho_{t,T}^{\pi}\}_{t=1,\dots,T}^{\pi\in\Pi}$ for a POMDP is normalized, translation-invariant, stochastically conditionally time consistent, and Markov if and only if transition risk mappings*

$$\sigma_t : \left\{\left(x_t, \Xi_t^{\pi}(h_t), Q_t^{\pi}(h_t)\right) : \pi \in \Pi,\ h_t \in \mathcal{X}^t\right\} \times \mathcal{V} \to \mathbb{R}, \quad t = 1 \dots T-1,$$

exist, such that

(i) *for all $t = 1, \dots, T-1$ and all $(x, \xi) \in \left\{\left(x_t, \Xi_t^{\pi}(h_t)\right) : \pi \in \Pi,\ h_t \in \mathcal{X}^t\right\}$, $\sigma_t(x, \xi, \cdot, \cdot)$ is normalized and strongly monotonic with respect to stochastic dominance on $\left\{ Q_t^{\pi}(h_t) : \pi \in \Pi,\ h_t \in \mathcal{X}^t \text{ such that } x_t = x,\ \Xi_t^{\pi}(h_t) = \xi \right\}$;*

(ii) *for all $\pi \in \Pi$, for all measurable bounded r, for all $t = 1, \dots, T-1$, and for all $h_t \in \mathcal{X}^t$,*

$$v_t^{\pi}(h_t) = r_t(x_t, \Xi_t^{\pi}(h_t), \pi_t(h_t)) + \sigma_t\left(x_t, \Xi_t^{\pi}(h_t), Q_t^{\pi}(h_t), v_{t+1}^{\pi}(h_t, \cdot)\right). \quad (10)$$

This allows us to evaluate risk of Markov policies in a recursive way.

Corollary 1 *Under the conditions of Theorem 2, for any Markov policy π, the function (9) depends on r, π_t,...,π_T, and (x_t, ξ_t) only, and the following relation is true:*

$$\begin{aligned}v_t^{\pi_t,\dots,\pi_T}(x_t, \xi_t) = r_t(x_t, \xi_t, \pi_t(x_t, \xi_t))\\ +\sigma_t\Big(x_t, \xi_t, \int_{\mathcal{Y}} K_t^X(x_t, y, \pi_t(x_t, \xi_t))\, \xi_t(dy), x'\\ \mapsto v_{t+1}^{\pi_{t+1},\dots,\pi_T}(x', \Phi_t(x_t, \xi_t, \pi_t(x_t, \xi_t), x'))\Big). \quad (11)\end{aligned}$$

Proof We use induction backward in time. For $t = T$ we have $v_T^\pi(h_T) = r_T(x_T, \xi_T, \pi_T(x_T, \xi_T))$ and our assertion is true. If it is true for $t+1$, formula (10) reads

$$\begin{aligned} v_t^\pi(h_t) = r_t(x_t, \xi_t, \pi_t(x_t, \xi_t)) \\ &+\sigma_t\big(x_t, \xi_t, Q_t(x_t, \xi_t, \pi_t(x_t, \xi_t)), \\ &x' \mapsto v_{t+1}^{\pi_{t+1},\dots,\pi_T}(x', \Phi_t(x_t, \xi_t, \pi_t(x_t, \xi_t), x'))\big). \end{aligned}$$

Substitution of (7) proves our assertion. □

5 Dynamic programming

We consider a family of dynamic risk measures $\{\rho_{t,T}^\pi\}_{t=1,\dots,T}^{\pi\in\Pi}$ which is normalized, translation-invariant, stochastically conditionally time consistent, and Markov. Our objective is to analyze the risk minimization problem:

$$\min_{\pi\in\Pi} v_1^\pi(x_1, \Xi_1(x_1)), \quad x_1 \in \mathscr{X}.$$

For this purpose, we introduce the family of *value functions*:

$$v_t^*(h_t) = \inf_{\pi\in\Pi_{t,T}} v_t^\pi(h_t), \quad t = 1, \dots, T, \quad h_t \in \mathscr{X}^t, \tag{12}$$

where $\Pi_{t,T}$ is the set of feasible deterministic policies $\pi = \{\pi_t, \dots, \pi_T\}$. By Theorem 2, transition risk mappings $\{\sigma_t\}_{t=1,\dots,T-1}$ exist, such that Eq. (10) hold.

We assume that the spaces $\mathscr{P}(\mathscr{X})$ and $\mathscr{P}(\mathscr{Y})$ are equipped with the topology of weak convergence, and the space $\mathscr{V}$ is equipped with the topology of pointwise convergence. All continuity statements are made with respect to the said topologies.

We also assume that the kernels $K_t(x, y, u)$ have densities $q_t(\cdot, \cdot \mid x, y, u)$ with respect to a finite product measure $\mu_X \otimes \mu_Y$ on $\mathscr{X} \times \mathscr{Y}$, as in Example 3. In this case,

$$\left[\int_{\mathscr{Y}} K_t^X(x, y, u)\, \xi(dy)\right](dx') = \left[\int_{\mathscr{Y}} \int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\, \xi(dy)\, \mu_Y(dy')\right] \mu_X(dx'). \tag{13}$$

Our main result is that the value functions (12) are *memoryless*, that is, they depend on (x_t, ξ_t) only, and that they satisfy a generalized form of a dynamic programming equation. The equation also allows us to identify the optimal policy.

Theorem 3 *We assume the following conditions:*

(i) *The functions $(x, u) \mapsto q_t(x', y'|x, y, u)$ are continuous at all (x', y', x, y, u), uniformly over (x', y', y);*

(ii) *The transition risk mappings $\sigma_t(\cdot, \cdot, \cdot, \cdot)$, $t = 1, \dots, T$, are lower semicontinuous;*

(iii) *The functions $r_t(\cdot, \cdot, \cdot)$, $t = 1, \dots, T$, are lower semicontinuous;*

(iv) *The multifunctions $\mathscr{U}_t(\cdot)$, $t = 1, \ldots, T$, are compact-valued and upper-semicontinuous.*

Then the functions v_t^*, $t = 1, \ldots, T$ are memoryless, lower semicontinuous, and satisfy the following dynamic programming equations:

$$\begin{aligned}
v_T^*(x, \xi) &= \min_{u \in \mathscr{U}_T(x)} r_T(x, \xi, u), \quad x \in \mathscr{X}, \quad \xi \in \mathscr{P}(\mathscr{X}),\\
v_t^*(x, \xi) &= \min_{u \in \mathscr{U}_t(x)} \Big\{ r_t(x, \xi, u)\\
&\quad + \sigma_t\Big(x, \xi, \int_{\mathscr{Y}} K_t^X(x, y, u)\, \xi(dy), x' \mapsto v_{t+1}^*\big(x', \Phi_t(x, \xi, u, x')\big)\Big)\Big\},\\
&\quad x \in \mathscr{X}, \quad \xi \in \mathscr{P}(\mathscr{Y}), \quad t = T-1, \ldots, 1.
\end{aligned}$$

Moreover, an optimal Markov policy $\hat{\pi}$ exists and satisfies the equations:

$$\begin{aligned}
\hat{\pi}_T(x, \xi) &\in \operatorname*{argmin}_{u \in \mathscr{U}_T(x)} r_T(x, \xi, u), \quad x \in \mathscr{X}, \quad \xi \in \mathscr{P}(\mathscr{Y}),\\
\hat{\pi}_t(x, \xi) &\in \operatorname*{argmin}_{u \in \mathscr{U}_t(x)} \Big\{ r_t(x, \xi, u)\\
&\quad + \sigma_t\Big(x, \xi, \int_{\mathscr{Y}} K_t^X(x, y, u)\, \xi(dy), x' \mapsto v_{t+1}^*\big(x', \Phi_t(x, \xi, u, x')\big)\Big)\Big\},\\
&\quad x \in \mathscr{X}, \quad \xi \in \mathscr{P}(\mathscr{Y}), \quad t = T-1, \ldots, 1.
\end{aligned}$$

Proof For all $h_T \in \mathscr{X}^T$ we have

$$v_T^*(h_T) = \inf_{\pi_T} r_T(x_T, \xi_T, \pi_T(h_T)) = \inf_{u \in \mathscr{U}_T(x_T)} r_T(x_T, \xi_T, u). \tag{14}$$

By assumptions (iii) and (iv), owing to the Berge theorem [see (Aubin and Frankowska 2009, Theorem 1.4.16)], the infimum in (14) is attained and is a lower semicontinuous function of (x_T, ξ_T). Hence, v_T^* is memoryless. Moreover, the optimal solution mapping $\Psi_T(x, \xi) = \big\{u \in \mathscr{U}_T(x) : r_T(x, \xi, u) = v_T^*(x, \xi)\big\}$ has nonempty and closed values and is measurable. Therefore, a measurable selector $\hat{\pi}_T$ of Ψ_T exists (see, Kuratowski and Ryll-Nardzewski 1965; Aubin and Frankowska 2009, Thm. 8.1.3), and

$$v_T^*(h_T) = v_T^*(x_T, \xi_T) = v_T^{\hat{\pi}_T}(x_T, \xi_T).$$

We prove the theorem by induction backward in time. Suppose $v_{t+1}^*(\cdot)$ is memoryless, lower semicontinuous, and Markov decision rules $\{\hat{\pi}_{t+1}, \ldots, \hat{\pi}_T\}$ exist such that

$$v_{t+1}^*(h_{t+1}) = v_{t+1}^*(x_{t+1}, \xi_{t+1}) = v_{t+1}^{\{\hat{\pi}_{t+1}, \ldots, \hat{\pi}_T\}}(x_{t+1}, \xi_{t+1}), \quad \forall h_{t+1} \in \mathscr{X}^{t+1}.$$

Then for any $h_t \in \mathscr{X}^t$ formula (10), after substituting (7), yields

$$\begin{aligned} v_t^*(h_t) &= \inf_{\pi \in \Pi_{t,T}} v_t^\pi(h_t) \\ &= \inf_{\pi \in \Pi_{t,T}} \{r_t(x_t, \xi_t, \pi_t(h_t)) \\ &\quad + \sigma_t\Big(x_t, \xi_t, \int_{\mathscr{Y}} K_t^X(x_t, y, \pi_t(h_t))\, \xi_t(dy), v_{t+1}^\pi(h_t, \cdot)\Big)\Big\}. \end{aligned}$$

Since $v_{t+1}^\pi(h_t, x') \geq v_{t+1}^*\big(x', \Phi_t\big(x_t, \xi_t, \pi_t(h_t), x'\big)\big)$ for all $x' \in \mathscr{X}$, and σ_t is non-decreasing with respect to the last argument, we obtain

$$\begin{aligned} v_t^*(h_t) &\geq \inf_{\pi \in \Pi_{t,T}} \Big\{r_t(x_t, \xi_t, \pi_t(h_t)) \\ &\quad + \sigma_t\Big(x_t, \xi_t, \int_{\mathscr{Y}} K_t^X(x_t, y, \pi_t(h_t))\, \xi_t(dy), x' \mapsto v_{t+1}^*(x', \Phi_t(x_t, \xi_t, \pi_t(h_t), x'))\Big)\Big\} \\ &= \inf_{u \in \mathscr{U}_t(x_t)} \Big\{r_t(x_t, \xi_t, u) \\ &\quad + \sigma_t\Big(x_t, \xi_t, \int_{\mathscr{Y}} K_t^X(x_t, y, u)\, \xi_t(dy), x' \mapsto v_{t+1}^*(x', \Phi_t(x_t, \xi_t, u, x'))\Big)\Big\}. \end{aligned} \tag{15}$$

In order to complete the induction step, we need to establish lower semicontinuity of the mapping

$$(x, \xi, u) \mapsto \sigma_t\Big(x, \xi, \int_{\mathscr{Y}} K_t^X(x, y, u)\, \xi(dy), x' \mapsto v_{t+1}^*\big(x', \Phi_t(x, \xi, u, x')\big)\Big). \tag{16}$$

To this end, suppose $x^{(k)} \to x$, $\xi^{(k)} \to \xi$ (weakly), $u^{(k)} \to u$, as $k \to \infty$.

First, we verify that the mapping $(x, \xi, u) \mapsto \int_{\mathscr{Y}} K_t^X(x, y, u)\, \xi(dy)$ appearing in the third argument of σ_t is weakly continuous. By formula (13), for any bounded continuous function $f : \mathscr{X} \to \mathbb{R}$ we have

$$\begin{aligned} &\int_{\mathscr{X}} f(x') \left[\int_{\mathscr{Y}} K_t^X(x^{(k)}, y, u^{(k)})\, \xi^{(k)}(dy)\right](dx') \\ &\quad = \int_{\mathscr{X}} f(x') \left[\int_{\mathscr{Y}} \int_{\mathscr{Y}} q_t(x', y' \mid x^{(k)}, y, u^{(k)})\, \xi^{(k)}(dy)\, \mu_Y(dy')\right] \mu_X(dx') \end{aligned} \tag{17}$$

By assumption (i),

$$\lim_{k \to \infty} \int_{\mathscr{Y}} \big[q_t(x', y' \mid x^{(k)}, y, u^{(k)}) - q_t(x', y' \mid x, y, u)\big] \xi^{(k)}(dy) = 0, \tag{18}$$

uniformly over x', y'. Moreover, by Lebesgue theorem, the function

$$y \mapsto \int_{\mathscr{X}} f(x') \int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\, \mu_Y(dy')\, \mu_X(dx') \tag{19}$$

is continuous. Therefore, combining (17) and (18), we obtain the chain of equations:

$$\begin{aligned}
&\lim_{k\to\infty}\int_{\mathscr{X}} f(x')\left[\int_{\mathscr{Y}} K_t^X(x^{(k)}, y, u^{(k)})\;\xi^{(k)}(dy)\right](dx')\\
&\quad=\lim_{k\to\infty}\int_{\mathscr{X}} f(x')\left[\int_{\mathscr{Y}}\int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\,\xi^{(k)}(dy)\,\mu_Y(dy')\right]\mu_X(dx')\\
&\quad=\lim_{k\to\infty}\int_{\mathscr{Y}}\left[\int_{\mathscr{X}} f(x')\int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\;\mu_Y(dy')\;\mu_X(dx')\right]\xi^{(k)}(dy)\\
&\quad=\int_{\mathscr{Y}}\left[\int_{\mathscr{X}} f(x')\int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\;\mu_Y(dy')\;\mu_X(dx')\right]\xi(dy)\\
&\quad=\int_{\mathscr{X}} f(x')\left[\int_{\mathscr{Y}} K_t^X(x, y, u)\;\xi(dy)\right](dx').
\end{aligned}$$

The last by one equation follows from the weak convergence of $\xi^{(k)}$ to ξ and from the continuity of the function (19). Thus, the third argument of σ_t in (16) is continuous with respect to (x, ξ, u).

Let us examine the last argument of σ_t in (16). By (18), for every continuous bounded function $f(\cdot)$ on $\mathscr{Y}$, and for each fixed $x' \in \mathscr{X}$,

$$\begin{aligned}
&\lim_{k\to\infty}\int_{\mathscr{Y}} f(y')\;\Phi_t(x^{(k)}, \xi^{(k)}, u^{(k)}, x')(dy')\\
&\quad=\lim_{k\to\infty}\frac{\int_{\mathscr{Y}} f(y')\int_{\mathscr{Y}} q_t(x', y' \mid x^{(k)}, y, u^{(k)})\;\xi^{(k)}(dy)\;\mu_Y(dy')}{\int_{\mathscr{Y}}\int_{\mathscr{Y}} q_t(x', y' \mid x^{(k)}, y, u^{(k)})\;\xi^{(k)}(dy)\;\mu_Y(dy')}\\
&\quad=\frac{\int_{\mathscr{Y}} f(y')\int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\;\xi(dy)\;\mu_Y(dy')}{\int_{\mathscr{Y}}\int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\;\xi(dy)\;\mu_Y(dy')},
\end{aligned}$$

provided that (x, ξ, u, x') is such that

$$\int_{\mathscr{Y}}\int_{\mathscr{Y}} q_t(x', y' \mid x, y, u)\;\xi(dy)\;\mu_Y(dy') > 0. \tag{20}$$

Therefore, the operator $\Phi_t(\cdot, \cdot, \cdot, x')$ is weakly continuous at these points. Let x, ξ, u be fixed. Consider the sequence of functions $V^{(k)} : \mathscr{X} \to \mathbb{R}$, $k = 1, 2, \ldots$, and the function $V : \mathscr{X} \to \mathbb{R}$, defined as follows:

$$\begin{aligned}
V^{(k)}(x') &= v^*_{t+1}\big(x', \Phi_t(x^{(k)}, \xi^{(k)}, u^{(k)}, x')\big),\\
V(x') &= v^*_{t+1}\big(x', \Phi_t(x, \xi, u, x')\big).
\end{aligned}$$

Since $v^*_{t+1}(\cdot, \cdot)$ is lower-semicontinuous and $\Phi_t(\cdot, \cdot, \cdot, x')$ is continuous, whenever condition (20) is satisfied, we infer that $V(x') \le \liminf_{k\to\infty} V^{(k)}(x')$, at all $x' \in \mathscr{X}$ at which (20) holds. As v^*_{t+1} and Φ_t are measurable, both V and $\liminf_{k\to\infty} V^{(k)}$ are measurable as well.

By Theorem 2, the mapping σ_t is preserving the stochastic order $\preceq_{\text{st}}$ of the last argument with respect to the measure $\int_{\mathscr{Y}} K_t^X(x,y,u)\,\xi(dy)$. Since

$$\Big(\int_{\mathscr{Y}} K_t^X(x,y,u)\,\xi(dy)\Big)\Big\{x'\in\mathscr{X} : \int_{\mathscr{Y}}\int_{\mathscr{Y}} q_t(x',y'\mid x,y,u)\,\xi(dy)\,\mu_Y(dy') = 0\Big\} = 0,$$

the value of $\liminf_{k\to\infty}\ V^{(k)}(x')$ at the set of x' at which (20) is violated, is irrelevant. Consequently, by assumption (ii), with the view at the already established continuity of the third argument, we obtain the following chain of relations:

$$\begin{aligned}\sigma_t\Big(x,\xi,\int_{\mathscr{Y}} K_t^X(x,y,u)\,\xi(dy),V\Big) &\le \sigma_t\Big(x,\xi,\int_{\mathscr{Y}} K_t^X(x,y,u)\,\xi(dy),\liminf_{k\to\infty}\ V^{(k)}\Big)\\ &= \sigma_t\Big(x,\xi,\lim_{k\to\infty}\int_{\mathscr{Y}} K_t^X(x^{(k)},y,u^{(k)})\,\xi^{(k)}(dy),\liminf_{k\to\infty}\ V^{(k)}\Big)\\ &\le \liminf_{k\to\infty}\sigma_t\Big(x^{(k)},\xi^{(k)},\int_{\mathscr{Y}} K_t^X(x^{(k)},y,u^{(k)})\,\xi^{(k)}(dy),V^{(k)}\Big).\end{aligned}$$

Consequently, the mapping (16) is lower semicontinuous.

Using assumptions (ii) and (iv) and invoking the Berge theorem again (see, e.g., Aubin and Frankowska 2009, Theorem 1.4.16), we deduce that the infimum in (15) is attained and is a lower semicontinuous function of (x_t,ξ_t). Moreover, the optimal solution mapping, that is, the set of $u\in\mathscr{U}_T(x)$ at which the infimum in (15) is attained, is nonempty, closed-valued, and measurable. Therefore, a minimizer $\hat{\pi}_t$ in (15) exists and is a measurable function of (x_t,ξ_t) (see, e.g., Kuratowski and Ryll-Nardzewski 1965; Aubin and Frankowska 2009, Thm. 8.1.3). Substituting this minimizer into (15), we obtain

$$\begin{aligned}v_t^*(h_t) \ge r_t\big(x_t,\xi_t,\hat{\pi}_t(x_t,\xi_t)\big)&\\ &+\sigma_t\Big(x_t,\xi_t,\int_{\mathscr{Y}} K_t^X\big(x_t,y,\hat{\pi}_t(x_t,\xi_t)\big)\,\xi_t(dy),\\ &x'\mapsto v_{t+1}^*\big(x',\Phi_t(x,\xi,\hat{\pi}_t(x_t,\xi_t),x')\big)\Big)\\ &= v_t^{\{\hat{\pi}_t,\dots,\hat{\pi}_T\}}(x_t,\xi_t).\end{aligned}$$

In the last equation, we used Corollary 1. On the other hand, we have

$$v_t^*(h_t) = \inf_{\pi\in\Pi_{t,T}} v_t^{\pi}(h_t) \le v_t^{\{\hat{\pi}_t,\dots,\hat{\pi}_T\}}(x_t,\xi_t).$$

Therefore $v_t^*(h_t) = v_t^{\{\hat{\pi}_t,\dots,\hat{\pi}_T\}}(x_t,\xi_t)$ is memoryless, lower semicontinuous, and

$$\begin{aligned}v_t^*(x_t,\xi_t) = \min_{u\in\mathscr{U}_t(x_t)}\Big\{&r_t(x_t,\xi_t,u)\\ &+\sigma_t\Big(x_t,\xi_t,\int_{\mathscr{Y}} K_t^X(x_t,y,u)\,\xi_t(dy),x'\mapsto v_{t+1}^*\big(x',\Phi_t(x_t,\xi_t,u,x')\big)\Big)\Big\}\end{aligned}$$

$$
\begin{aligned}
&= r_t\big(x_t, \xi_t, \hat{\pi}_t(x_t, \xi_t)\big) \\
&+ \sigma_t\bigg(x_t, \xi_t, \int_{\mathscr{Y}} K_t^X\big(x_t, y, \hat{\pi}_t(x_t, \xi_t)\big)\, \xi_t(dy), \\
&x' \mapsto v_{t+1}^*\big(x', \Phi_t(x_t, \xi_t, \hat{\pi}_t(x_t, \xi_t), x')\big)\bigg).
\end{aligned}
$$

This completes the induction step. □

The most essential assumption of Theorem 3 is assumption (ii) of the lower semicontinuity of the transition risk mappings $\sigma_t(\cdot, \cdot, \cdot, \cdot)$. If these mappings are derived from convex or coherent risk measures, as illustrated in Fan and Ruszczyński (2016), their lower semicontinuity with respect to the last argument follows from the corresponding property of the risk measure. In particular, (Ruszczyński and Shapiro 2006a, Cor. 3.1) derives continuity from monotonicity on Banach lattices. The semicontinuity with respect to the third argument, the probability measure, is a more complex issue. The reference Fan and Ruszczyński (2016), Lemmas 4.7 and 4.8, verifies this condition for two popular risk measures: the Average Value at Risk and the mean-semideviation measure. Similar remarks apply to the assumption (iii) about the state-control risk functions. The assumptions (i) and (iv) are the same as in the utility models of Bäuerle and Rieder (2017).

We could have made the sets $\mathscr{U}_t$ depend on ξ_t, but this is hard to justify.

6 Illustration: machine deterioration

6.1 Description of the process

Consider the problem of minimizing costs of using a machine in T periods. The condition of the machine can deteriorate over time, but is not known with certainty. The only information available is the operating cost. The control in any period is to continue using the machine, or to replace it.

At the beginning of period $t = 1, \dots, T$, the condition of the machine is denoted by $Y_t \in \{1, 2\}$, with 1 denoting the “good” and 2 the “bad” state. The controls are denoted by $u_t \in \{0, 1\}$, with 0 meaning “continue”, and 1 meaning “replace”.

The dynamics is Markovian, with the following transition matrices $K^{[u]}, u \in \{0, 1\}$:

$$
K^{[0]} = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}, \quad K^{[1]} = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}. \tag{21}
$$

We can observe the cost incurred during period t, denoted by X_{t+1}. The increment of the time index is due to the fact that the cost becomes known at the end of the period, and provides information for the decision in the next period. The conditional distribution of X_{t+1}, given y_t and u_t, is described by two density functions f_1 and f_2:

$$\begin{aligned}\mathbb{P}\big[X_{t+1} \le C \,\big|\, y_t = i, u_t = 0\big] &= \int_{-\infty}^{C} f_i(x)\,dx, \quad i = 1, 2,\\ \mathbb{P}\big[X_{t+1} \le C \,\big|\, y_t = i, u_t = 1\big] &= \int_{-\infty}^{C} f_1(x)\,dx, \quad i = 1, 2.\end{aligned} \tag{22}$$

Assumption 2 The functions f_1 and f_2 are uniformly bounded and the conditional distribution of X_{t+1} given that the machine is in "good" condition is stochastically smaller than the conditional distribution of x_{t+1} given that the machine is in "bad" condition, i.e.,

$$\int_{-\infty}^{C} f_1(x)\,dx \ge \int_{-\infty}^{C} f_2(x)\,dx, \quad \forall\, C \in \mathbb{R};$$

with a slight abuse of notation, we write it $f_1 \preceq_{\text{st}} f_2$.

Thus the relations (21) and (22) define $\big\{X_t, Y_t\big\}_{t=1,\dots,T+1}$ as a partially observable Markov process controlled by $\big\{u_t\big\}_{t=1,\dots,T}$. Based on observations $(x_1, \dots, x_t)$, the belief state $\xi_t \in [0, 1]$ denotes the conditional probability that $Y_t = 1$. We can update the posterior belief state as follows:

$$\xi_{t+1} = \begin{cases} \Phi(\xi_t, x_{t+1}), & \text{if } u_t = 0;\\ 1 - p, & \text{if } u_t = 1,\end{cases}$$

where Φ is the Bayes operator,

$$\Phi(\xi, x') = \frac{(1-p)\xi f_1(x')}{\xi f_1(x') + (1-\xi) f_2(x')}. \tag{23}$$

We assume that the initial probability $\xi_0 \in [0, 1]$ is known; then $\xi_1(x_1) = \Phi(\xi_0, x_1)$. From (23) we see that $\Phi(0, \cdot\,) = 0$, $\Phi(1, \cdot\,) = 1 - p$, and $\Phi(\,\cdot\,, x')$ is non-decreasing.

6.2 Risk modeling

At the beginning of period t, if we replace the machine ($u_t = 1$), there is an additional fixed replacement cost R. Then the costs incurred are

$$\begin{cases} r_t(x_t, u_t) = R \cdot u_t + x_t, \quad t = 1, \dots, T;\\ r_{T+1}(x_{T+1}) = x_{T+1}.\end{cases} \tag{24}$$

We denote the history of observations by $h_t = (x_1, \dots, x_t)$ and the set of all history-dependent policies by $\Pi := \big\{\pi = (\pi_1, \dots, \pi_T) \,|\, \forall t,\ \pi_t(x_1, \dots, x_t) \in \{0, 1\}\big\}$. We want to evaluate the risk of costs (24) for any $\pi \in \Pi$, and find an optimal policy.

As shown in Theorem 2, construction of Markovian risk measures is equivalent to specifying transition risk mappings $\sigma_t : \mathbb{R} \times \mathscr{P}(\mathbb{R}) \times \mathscr{P}(\mathbb{R}) \times \mathscr{V} \to \mathbb{R}$, where

$\mathscr{V}$ is the space of all bounded and measurable functions from $\mathbb{R}$ to $\mathbb{R}$. For simplicity, we assume that $\sigma_t(\cdot,\cdot,\cdot,\cdot)$ is the same for all t and does not depend on the current state (x_t,ξ_t), that is, $\sigma_t(x,\xi,q,v)=\sigma(q,v)$.

Remark 2 For a probability measure $q\in\mathscr{P}(\mathbb{R})$ that has f as the density function, with slight abuse of notation, we also write $\sigma(f,\cdot)$ instead of $\sigma(q,\cdot)$.

6.3 Value and policy monotonicity

We assume that the transition risk mapping $\sigma:\mathscr{P}(\mathbb{R})\times\mathscr{V}\to\mathbb{R}$ satisfy all assumptions of Theorem 3. Then the optimal value functions v_t^*, $t=1,\dots,T+1$ are memoryless and satisfy the following dynamic programming equations:

$$\begin{aligned} v_t^*(x,\xi) = x + \min\Big(&R+\sigma\big(f_1, x'\mapsto v_{t+1}^*(x',1-p)\big);\\ &\sigma\big(\xi f_1+(1-\xi)f_2, x'\mapsto v_{t+1}^*(x',\Phi(\xi,x'))\big)\Big),\\ &x\in\mathbb{R},\quad \xi\in[0,1],\quad t=1,\dots,T, \end{aligned}\tag{25}$$

with the final stage value $v_{T+1}^*(x,\xi)=x$. Moreover, an optimal Markov policy exists, which is defined by the minimizers in the above dynamic programming equations.

Directly from (25) we see that $v_t^*(x,\xi)=x+w_t^*(\xi)$, $t=1,\dots,T+1$. The dynamic programming Eq. (25) simplify as follows:

$$\begin{aligned} w_t^*(\xi) = \min\Big\{&R+\sigma\big(f_1, x'\mapsto x'+w_{t+1}^*(1-p)\big);\\ &\sigma\big(\xi f_1+(1-\xi)f_2, x'\mapsto x'+w_{t+1}^*(\Phi(\xi,x'))\big)\Big\}, \end{aligned}\tag{26}$$

with the final stage value $w_{T+1}^*(\,\cdot\,)=0$. We can establish monotonicity of $w^*(\cdot)$.

Theorem 4 *If $\frac{f_1}{f_2}$ is non-increasing, then the functions $w_t^*:[0,1]\to\mathbb{R}$, $t=1,\dots,T+1$ are non-increasing.*

Proof Clearly, w_{T+1}^* is non-increasing. Assume by induction that w_{t+1}^* is non-increasing. For any $\xi_1\le\xi_2$ we have:

1. $\xi_1 f_1+(1-\xi_1)f_2\succeq_{\text{st}}\xi_2 f_1+(1-\xi_2)f_2$, because $f_1\preceq_{\text{st}} f_2$.
2. For all x', we have $x'+w_{t+1}^*(\Phi(\xi_1,x'))\ge x'+w_{t+1}^*(\Phi(\xi_2,x'))$, as w_{t+1}^* is non-increasing and $\Phi(\cdot,x')$ is non-decreasing.
3. the mapping $x'\mapsto x'+w_{t+1}^*(\Phi(\xi,x'))$ is non-decreasing for all ξ. To show that, it is sufficient to establish that $x'\mapsto\Phi(\xi,x')$ is non-increasing, and this can be seen from the formula (for $0\le p<1$):

$$\frac{1}{\Phi(\xi,x')}=\frac{1}{1-p}\left(1+\frac{f_2(x')}{f_1(x')}\left(\frac{1}{\xi}-1\right)\right).$$

Thus

$$\begin{aligned}
&\sigma\big(\xi_1 f_1 + (1-\xi_1) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\xi_1, x'))\big)\\
&\quad \geq \sigma\big(\xi_1 f_1 + (1-\xi_1) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\xi_2, x'))\big) && \text{(because of 2.)}\\
&\quad \geq \sigma\big(\xi_2 f_1 + (1-\xi_2) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\xi_2, x'))\big) && \text{(because of 1. and 3.)}
\end{aligned}$$

which completes the induction step. □

The monotonicity assumption on $\frac{f_1}{f_2}$ is in fact a sufficient condition for $f_1 \preceq_{\text{st}} f_2$. From Theorem 4 we obtain the following threshold property of the policy.

Theorem 5 *Under the assumptions of Theorem 4, there exist thresholds $\xi^*_t \in [0,1]$, $t = 1, \dots, T$ such that the policy*

$$u^*_t = \begin{cases} 0 & \text{if } \xi_t > \xi^*_t,\\ 1 & \text{if } \xi_t \leq \xi^*_t, \end{cases}$$

is optimal.

Proof Suppose ξ is such that replacement at time t is optimal:

$$R + \sigma\big(f_1, x' \mapsto x' + w^*_{t+1}(1-p)\big) \leq \sigma\big(\xi f_1 + (1-\xi) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\xi, x'))\big).$$

Then for any $\zeta \leq \xi$, we have $\xi f_1 + (1-\xi) f_2 \preceq_{\text{st}} \zeta f_1 + (1-\zeta) f_2$ and $\Phi(\xi, x') \geq \Phi(\zeta, x')$. Consequently,

$$\begin{aligned}
&R + \sigma\big(f_1, x' \mapsto x' + w^*_{t+1}(1-p)\big)\\
&\quad \leq \sigma\big(\xi f_1 + (1-\xi) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\xi, x'))\big)\\
&\quad \leq \sigma\big(\zeta f_1 + (1-\zeta) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\xi, x'))\big)\\
&\quad \leq \sigma\big(\zeta f_1 + (1-\zeta) f_2, x' \mapsto x' + w^*_{t+1}(\Phi(\zeta, x'))\big),
\end{aligned}$$

and replacement is optimal for ζ as well. □

6.4 Numerical illustration

In this section, we solve the problem in the special case where f_1 and f_2 are density functions $\mathbb{U}(m_1, M_1)$ and $\mathbb{U}(m_2, M_2)$ with $m_1 \leq m_2 \leq M_1 \leq M_2$. Then the Bayes operator is piece-wise constant with respect to x':

$$\Phi(\xi, x') = \begin{cases} 1-p, & \text{if } m_1 \leq x' \leq m_2;\\ \dfrac{(1-p)\xi(M_2-m_2)}{\xi(M_2-m_2)+(1-\xi)(M_1-m_1)} := \hat{\phi}(\xi), & \text{if } m_2 \leq x' \leq M_1;\\ 0, & \text{if } M_1 \leq x' \leq M_2. \end{cases}$$

The conditional distribution of x' given ξ is described by the density function $\xi f_1 + (1-\xi) f_2$, which is also constant in each of the three intervals $[m_1, m_2)$, $[m_2, M_1]$ and $(M_1, M_2]$, with the following probabilities amassed in each of the three intervals:

$$\begin{aligned}
q_1(\xi) &= \frac{\xi(m_2 - m_1)}{M_1 - m_1},\\
q_2(\xi) &= (M_1 - m_2)\left(\frac{\xi}{M_1 - m_1} + \frac{1-\xi}{M_2 - m_2}\right),\\
q_3(\xi) &= \frac{(1-\xi)(M_2 - M_1)}{M_2 - m_2}.
\end{aligned}$$

We use the mean-semideviation transition risk mapping of Example 2, with $p = 1$ and constant $\varkappa$. It is strongly monotonic with respect to stochastic order and lower semi-continuous with respect to (q, v). Then the dynamic programming Eq. (26) for $t = 1, \ldots, T$ become:

$$\begin{aligned}
w_t^*(\xi) = \min\Big\{ & R + E_t^*(1) + \mathbb{E}_{f_1}\big(x' \mapsto x' + w_{t+1}^*(1-p) - E_t^*(1)\big)_+;\\
& E_t^*(\xi) + \mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w_{t+1}^*(\Phi(\xi, x')) - E_t^*(\xi)\big)_+\Big\}, \qquad (27)
\end{aligned}$$

where

$$\begin{aligned}
E_t^*(\xi) :&= \mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w_{t+1}^*(\Phi(\xi, x'))\big)\\
&= q_1(\xi)\left(\frac{m_1 + m_2}{2} + w_{t+1}^*(1-p)\right)\\
&\quad + q_2(\xi)\left(\frac{m_2 + M_1}{2} + w_{t+1}^*(\hat{\phi}(\xi))\right)\\
&\quad + q_3(\xi)\left(\frac{M_1 + M_2}{2} + w_{t+1}^*(0)\right). \qquad (28)
\end{aligned}$$

As $x' \mapsto x' + w_{t+1}^*(\Phi(\xi, x')) - E_t^*(\xi)$ is linear in each of the intervals $[m_1, m_2]$, $[m_2, M_1]$ and $[M_1, M_2]$, we have

$$\begin{aligned}
&\mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w_{t+1}^*(\Phi(\xi, x')) - E_t^*(\xi)\big)_+\\
&\quad = q_1(\xi)\,\theta\big(m_1, m_2, E_t^*(\xi) - w_{t+1}^*(1-p)\big)\\
&\qquad + q_2(\xi)\,\theta\big(m_2, M_1, E_t^*(\xi) - w_{t+1}^*(\hat{\phi}(\xi))\big)\\
&\qquad + q_3(\xi)\,\theta\big(M_1, M_2, E_t^*(\xi) - w_{t+1}^*(0)\big),
\end{aligned} \qquad (29)$$

where, for $a_1 \le a_2$,

$$\theta(a_1, a_2, a_3) := \frac{\int_{a_1}^{a_2} (\,\cdot\, - a_3)_+}{a_2 - a_1} = \begin{cases} \frac{1}{2}(a_1 + a_2), & \text{if } a_3 \le a_1;\\ \frac{1}{2}(a_3 + a_2)\frac{a_2 - a_3}{a_2 - a_1}, & \text{if } a_1 < a_3 < a_2;\\ 0, & \text{if } a_3 \ge a_2. \end{cases}$$

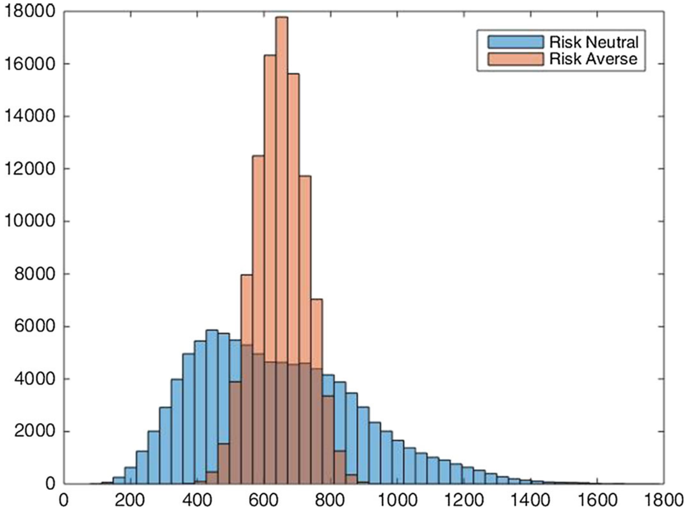

Fig. 1 Empirical distribution of the total cost for the risk-neutral model (blue) and the risk-averse model (orange)

For any t and any ξ, if we know $w^*_{t+1}(1-p)$, $w^*_{t+1}(\hat{\phi}(\xi))$ and $w^*_{t+1}(0)$, then the computation of $w^*_t(\xi)$ can be accomplished in three steps:

1. Compute $E^*_t(1)$ and $E^*_t(\xi)$ by (28).
2. Compute $\mathbb{E}_{f_1}\big(x' \mapsto x' + w^*_{t+1}(1-p) - E^*_t(1)\big)_+$ and $\mathbb{E}_{\xi f_1+(1-\xi)f_2}\big(x' \mapsto x' + w^*_{t+1}(\Phi(\xi, x')) - E^*_t(\xi)\big)_+$ by (29).
3. Compute $w^*_t(\xi)$ using the dynamic programming Eq. (27).

Since $w^*_{T+1} = 0$, all $w^*_t(\xi)$ can be calculated by recursion backward in time.

In Fig. 1, we display the distribution of the total cost obtained by simulating 100,000 runs of the system with two policies: risk-neutral and risk-averse for an example with $m_1 = 0$, $m_2 = 80$, $M_1 = 100$, $M_2 = 500$, $p = 0.2$, $T = 6$, $R = 50$, and $\varkappa = 0.9$.

The application of the risk-averse model increases the threshold values ξ^*_t of the optimal policies and results in a significantly less dispersed distribution of the total cost. For additional details, see Fan (2017).

Acknowledgements Funding was provided by the National Science Foundation, Division of Mathematical Sciences (Grant No. 1312016).

References

Arlotto A, Gans N, Steele JM (2014) Markov decision problems where means bound variances. Oper Res 62(4):864–875

Artzner P, Delbaen F, Eber J-M, Heath D, Ku H (2007) Coherent multiperiod risk adjusted values and Bellman's principle. Ann Oper Res 152:5–22

Aubin J-P, Frankowska H (2009) Set-valued analysis. Birkhäuser, Boston

Bäuerle N, Rieder U (2011) Markov decision processes with applications to finance. Universitext. Springer, Heidelberg

Bäuerle N, Rieder U (2013) More risk-sensitive Markov decision processes. Math Oper Res 39(1):105–120

Bäuerle N, Rieder U (2017) Partially observable risk-sensitive Markov decision processes. Math Oper Res 42:1180–1196

Bertsekas DP, Shreve SE (1978) Stochastic optimal control, volume 139 of mathematics in science and engineering. Academic Press, New York

Çavus Ö, Ruszczyński A (2014a) Computational methods for risk-averse undiscounted transient Markov models. Oper Res 62(2):401–417

Çavus Ö, Ruszczyński A (2014b) Risk-averse control of undiscounted transient Markov models. SIAM J Control Optim 52(6):3935–3966

Chen Z, Li G, Zhao Y (2014) Time-consistent investment policies in Markovian markets: a case of mean-variance analysis. J Econ Dyn Control 40:293–316

Cheridito P, Delbaen F, Kupper M (2006) Dynamic monetary risk measures for bounded discrete-time processes. Electron J Probab 11:57–106

Cheridito P, Kupper M (2011) Composition of time-consistent dynamic monetary risk measures in discrete time. Int J Theor Appl Finance 14(01):137–162

Chu S, Zhang Y (2014) Markov decision processes with iterated coherent risk measures. Int J Control 87(11):2286–2293

Coraluppi SP, Marcus SI (1999) Risk-sensitive and minimax control of discrete-time, finite-state Markov decision processes. Automatica 35(2):301–309

Dai Pra P, Meneghini L, Runggaldier WJ (1998) Explicit solutions for multivariate, discrete-time control problems under uncertainty. Syst Control Lett 34(4):169–176

Denardo EV, Rothblum UG (1979) Optimal stopping, exponential utility, and linear programming. Math Program 16(2):228–244

Di Masi GB, Stettner Ł (1999) Risk-sensitive control of discrete-time Markov processes with infinite horizon. SIAM J Control Optim 38(1):61–78

Fan J (2017) Process-based risk measures and risk-averse control of observable and partially observable discrete-time systems. Ph.D. Dissertation, Rutgers University

Fan J, Ruszczyński A (2016) Process-based risk measures and risk-averse control of discrete-time systems. arXiv:1411.2675

Feinberg EA, Kasyanov PO, Zgurovsky MZ (2016) Partially observable total-cost Markov decision processes with weakly continuous transition probabilities. Math Oper Res 41(2):656–681

Fernández-Gaucherand E, Marcus SI (1997) Risk-sensitive optimal control of hidden Markov models: structural results. IEEE Trans Autom Control 42(10):1418–1422

Filar JA, Kallenberg LCM, Lee H-M (1989) Variance-penalized Markov decision processes. Math Oper Res 14(1):147–161

Föllmer H, Penner I (2006) Convex risk measures and the dynamics of their penalty functions. Stat Decis 24(1/2006):61–96

Hinderer K (1970) Foundations of non-stationary dynamic programming with discrete time parameter. Springer, Berlin

Howard RA, Matheson JE (1971/72) Risk-sensitive Markov decision processes. Manag Sci. 18:356–369

James MR, Baras JS, Elliott RJ (1994) Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems. IEEE Trans Autom Control 39(4):780–792

Jaquette SC (1973) Markov decision processes with a new optimality criterion: discrete time. Ann Statist 1:496–505

Jaśkiewicz A, Matkowski J, Nowak AS (2013) Persistently optimal policies in stochastic dynamic programming with generalized discounting. Math Oper Res 38(1):108–121

Jobert A, Rogers LCG (2008) Valuations and dynamic convex risk measures. Math Finance 18(1):1–22

Klöppel S, Schweizer M (2007) Dynamic indifference valuation via convex risk measures. Math Finance 17(4):599–627

Kuratowski K, Ryll-Nardzewski C (1965) A general theorem on selectors. Bull Acad Polon Sci Ser Sci Math Astron Phys 13(1):397–403

Levitt S, Ben-Israel A (2001) On modeling risk in Markov decision processes. In: Rubinov A, Glover B (eds) Optimization and related topics . Applied Optimization, vol 47. Springer, Boston, MA, pp 27–40

Lin K, Marcus SI (2013) Dynamic programming with non-convex risk-sensitive measures. In: American control conference (ACC), 2013, IEEE, pp 6778–6783

Mannor S, Tsitsiklis JN (2013) Algorithmic aspects of mean-variance optimization in Markov decision processes. Eur J Oper Res 231(3):645–653

Marcus, SI, Fernández-Gaucherand E, Hernández-Hernández D, Coraluppi S, Fard P (1997) Risk sensitive Markov decision processes. In: Byrnes CI, Datta BN, Martin CF, Gilliam DS (eds) Systems and control in the twenty-first century. Systems & Control: Foundations & Applications, vol 22. Birkhäuser, Boston, MA, pp 263–279

Ogryczak W, Ruszczyński A (1999) From stochastic dominance to mean-risk models: semideviations as risk measures. Eur J Oper Res 116(1):33–50

Ogryczak W, Ruszczyński A (2001) On consistency of stochastic dominance and mean-semideviation models. Math Program 89(2):217–232

Pflug ChG, Römisch W (2007) Modeling, measuring and managing risk. World Scientific, Singapore

Riedel F (2004) Dynamic coherent risk measures. Stoch Process Their Appl 112:185–200

Roorda B, Schumacher JM, Engwerda J (2005) Coherent acceptability measures in multiperiod models. Math Finance 15(4):589–612

Runggaldier WJ (1998) Concepts and methods for discrete and continuous time control under uncertainty. Insur Math Econ 22(1):25–39

Ruszczyński A (2010) Risk-averse dynamic programming for Markov decision processes. Math Program 125(2, Ser. B):235–261

Ruszczyński A, Shapiro A (2006a) Optimization of convex risk functions. Math Oper Res 31:433–542

Ruszczyński A, Shapiro A (2006b) Conditional risk mappings. Math Oper Res 31:544–561

Scandolo G (2003) Risk measures in a dynamic setting. Ph.D. thesis, Università degli Studi di Milano

Shen Y, Stannat W, Obermayer K (2013) Risk-sensitive Markov control processes. SIAM J Control Optim 51(5):3652–3672

White DJ (1988) Mean, variance, and probabilistic criteria in finite Markov decision processes: a review. J Optim Theory Appl 56(1):1–29