Mathematical Methods of Operations Research
https://doi.org/10.1007/s00186-023-00834-z

ORIGINAL ARTICLE

Risk filtering and risk-averse control of Markovian systems subject to model uncertainty

Tomasz R. Bielecki[1] · **Igor Cialenco**[1] · **Andrzej Ruszczyński**[2]

Received: 30 April 2023 / Revised: 14 August 2023 / Accepted: 15 August 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
We consider a Markov decision process subject to model uncertainty in a Bayesian framework, where we assume that the state process is observed but its law is unknown to the observer. In addition, while the state process and the controls are observed at time t, the actual cost that may depend on the unknown parameter is not known at time t. The controller optimizes the total cost by using a family of special risk measures, called risk filters, that are appropriately defined to take into account the model uncertainty of the controlled system. These key features lead to non-standard and non-trivial risk-averse control problems, for which we derive the Bellman principle of optimality. We illustrate the general theory on two practical examples: clinical trials and optimal investment.

Keywords Markov decision processes · Model uncertainty · Dynamic measures of risk · Dynamic programming

1 Introduction

We study a risk-averse Markov decision problem (MDP) subject to uncertainty about the underlying dynamics as well as uncertainty about the risk-averse criterion. The uncertainty about the underlying dynamics, that is the lack of perfect knowledge of

✉ Andrzej Ruszczyński
rusz@rutgers.edu

Tomasz R. Bielecki
tbielecki@iit.edu

Igor Cialenco
cialenco@iit.edu

[1] Department of Applied Mathematics, Illinois Institute of Technology, 10 W 32nd Str, Bld RE, Room 220, Chicago, IL 60616, USA

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

Published online: 02 September 2023

the form of the controlled underlying dynamical system, is the type of uncertainty that is referred to as the Knightian uncertainty (after Frank Knight). In the present paper, we additionally face the uncertainty about the control criterion. Accordingly, here, we understand by Knightian uncertainty both the uncertainty about the underlying dynamics as well as uncertainty about the risk-averse criterion.

The literature concerning risk-averse MDPs is rather abundant, and we refer to e.g. Fan and Ruszczyński (2022, 2018) and references therein. Similarly, there is a vast literature on MDPs subject to uncertainty about the model dynamics, and we refer to (Bielecki et al., 2019) for an overview of the classical methodologies on this topic. MDPs subject to model dynamics uncertainty have been studied using both the robust methodology and the Bayesian methodology (see e.g. Wolff et al. (2012), Lin et al. (2021) and references therein). However, to the best of our knowledge, the present study is the first systematic study of *risk-averse* MDPs subject to model uncertainty. An earlier effort to deal with a risk-averse MDP subject to model uncertainty in Lin et al. (2021) focuses on the CVaR criterion that has an equivalent expected value formulation and is only addressing the uncertainty about model dynamics. As already said above, we are not only concerned with uncertainty regarding the underlying dynamics, but also uncertainty about the optimization criterion, which is a novel and important practical feature, as Examples 3.17, 3.18, and 3.19 below show. While frequent in machine learning literature, although mainly concerned with the expected value criterion [cf. Lattimore and Szepesvári (2020); Sutton and Barto (2018)], that has not been addressed in the risk-averse case.

The Knightian uncertainty that we consider is parametric in nature, and our approach to tackling the respective MDP is rooted in the Bayesian methodology. This means that we treat the unknown parameter as a random variable, denoted by Θ below, and a part of the hybrid state that leads to the evaluation of the control policy is the posterior distribution of this random variable [*cf* equation (5.1)], which is updated via an explicit recursion given in Proposition 5.2. It came to us as quite a surprise that accounting for possible uncertainty about the optimization criterion leads to rather intricate conceptual ideas and technical manipulations. In order to avoid measurability and integrability issues that are notorious and intrinsic in MDPs on general state and action spaces, and that would quite likely burden the main takeaways from this study, we decided to work with discrete state, action, and parameter spaces. However, morally, the results should hold true in much more generality, which will be addressed in future works. We chose to use integral notation with respect to the state variables, which is much more pleasing to the eye and lighter than the summation notation. We keep the summation with respect to the time variable though, whenever needed.

The solution to the considered risk-averse MDP hinges on the key and new concepts of dynamic risk filters and recursive dynamic risk filters, as well as the notions of parameter consistency and time consistency for dynamic risk filters. These, in particular, allow to derive a version of the dynamic programming routine suited to the needs of our uncertain risk-averse MDP.

The main contributions of and takeaways from this paper are the following:

1. To the best of our knowledge the present study is the first systematic study of risk-averse MDPs subject to model uncertainty encompassing both the uncertainty about the model dynamics as well as uncertainty about the performance criterion.
2. The concepts of parameter-consistency and time-consistency of dynamic risk filters.
3. Theorem 3.14 that characterizes the dynamic risk filters satisfying conditions stated in Definition 3.3.
4. Explicit recursion for the posterior distribution (belief state) for Θ, given in Proposition 5.2.
5. Theorem 5.6 that provides a solution to our risk-averse control of Markovian systems subject to model uncertainty.

The paper is organized as follows. In Sect. 2 we set the stage and define MDP and the model uncertainty framework. Also here we introduce a series of probability measures and some of their properties used frequently in the sequel. Section 3 is devoted to risk filters, starting with the definition and some fundamental properties of these objects. The key concept of parameter consistency of risk filters is introduced in Sect. 3.2, while the time consistency of risk filters is studied in Sect. 3.3. In this Sect. we provide a characterization of parameter-consistent and time-consistent risk filters; cf. Theorem 3.14. Also here we discuss two important examples of risk filters: the expectation of an additive functional, Example 3.17, and risk-sensitive criteria in the context of clinical trials, Example 3.18. The structure Theorem 3.14 leads to the notation of recursive risk filters, introduced in Sect. 4. Sect. 5.3 is devoted to the risk-averse control problem. In Sect. 5.1 we derive the Bayes kernel for the posterior distribution of the parameter of interest. Then, we derive the dynamic programming backward recursion for the classical additive reward case; Sect. 5.2. Here, as a particular case, we briefly discuss the optimal investment and consumption problem, when the investor faces Knightian uncertainty and unknown risk-aversion parameter; Example 3.19. We conclude with the solution to the optimal control problem for a general recursive risk filer: Theorem 5.6.

Finally, we want to mention that while writing this manuscript we strove to keep a balance between heavy notations and rigor. Nevertheless, some formulas still may appear overwhelming, which is typically the case for MDPs.

2 Markov decision processes with model uncertainty

We consider an observed, controlled random process $X = \{X_t\}_{t=1,\dots,T}$. The corresponding state space is a finite set $\mathcal{X}$. The underlying probability space that we will work with is canonical. It includes the space of paths of X: $\Omega = \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{T \text{ times}} = (\mathcal{X})^T$, endowed with the canonical product σ-field $\mathcal{F} = \underbrace{2^{\mathcal{X}} \otimes \cdots \otimes 2^{\mathcal{X}}}_{T \text{ times}}$. The elements of Ω are $\omega = (\omega_1, \dots, \omega_T)$. We use x_t to denote the canonical projections at time t, so that $X_t(\omega) = x_t = \omega_t$. We let $\{\mathcal{F}_t^X\}_{t=1,\dots,T}$ to denote the canonical filtration generated by the process X, so that $\mathcal{F}_t^X = \underbrace{2^{\mathcal{X}} \otimes \cdots \otimes 2^{\mathcal{X}}}_{t \text{ times}} \otimes \underbrace{\{\Omega, \emptyset\} \otimes \cdots \otimes \{\Omega, \emptyset\}}_{T-t \text{ times}}$. We will make use of the notations $\mathcal{T} = \{1, \dots, T\}$ and $\mathcal{T}_t = \{t, \dots, T\}$.

The control space is given by a finite set $\mathcal{U}$, and the set of admissible controls at step t is given by a set-valued function (or a multifunction) $\mathcal{U}_t : \mathcal{X} \rightrightarrows \mathcal{U}$ with nonempty values. We consider a parametric family of transition kernels $K_\theta : \mathcal{X} \times \mathcal{U} \to \mathcal{P}(\mathcal{X})$, where $\mathcal{P}(\mathcal{X})$ is the space of probability measures on $\mathcal{X}$, and $\theta \in \widehat{\mathbf{\Theta}}$ represents an unknown parameter. Here, $\widehat{\mathbf{\Theta}}$ is a finite set. The unknown true value of the parameter θ is θ^*.

We will consider the Bayesian setting, and therefore, we consider the product space $\widehat{\Omega} = \Omega \times \widehat{\mathbf{\Theta}}$ endowed with product σ-algebra $\widehat{\mathcal{F}} := \mathcal{F} \otimes 2^{\widehat{\mathbf{\Theta}}}$. We denote $\widehat{\omega} = (\omega, \theta)$ and $\widehat{\omega}_t = (\omega_t, \theta)$. In accordance with the Bayesian setting we denote by Θ a random variable on $(\widehat{\Omega}, \widehat{\mathcal{F}})$ with values in $\widehat{\mathbf{\Theta}}$, and with $\Theta(\widehat{\omega}) = \theta$. We also assume that some *prior distribution* ξ_1 of Θ (supported in $\widehat{\mathbf{\Theta}}$) is available.

Remark 2.1 We chose to work with finite sets $\mathcal{X}, \mathcal{U}$, and $\widehat{\mathbf{\Theta}}$ so to avoid dealing with technical and delicate issues of measurability of various functions that we encounter throughout the analysis, as well as the issues of existence and nature of measurable selectors. These technical problems will be tackled in the future.

The process $\{X_t\}_{t\in\mathcal{T}}$ considered as a process on $(\widehat{\Omega}, \widehat{\mathcal{F}})$ is denoted as $\widehat{X} = \{\widehat{X}_t\}_{t\in\mathcal{T}}$, and $\widehat{X}_t(\widehat{\omega}) = \widehat{X}_t(\omega, \theta) = X_t(\omega) = \omega_t$. Accordingly, the canonical filtration generated by the process $\widehat{X}$ is given as $\{\mathcal{F}_t^{\widehat{X}} = \mathcal{F}_t^X \otimes \{\emptyset, \widehat{\mathbf{\Theta}}\},\ t \in \mathcal{T}\}$.

At time t, the history of observed states is $h_t = (x_1, x_2, \ldots, x_t)$, while all the information available for making a decision is $g_t = (x_1, u_1, x_2, u_2, \ldots, x_t)$. We use $\mathcal{H}_t := \mathcal{X}^t = \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{t \text{ times}}$ to denote the spaces of possible state histories h_t. We make distinction of g_t and h_t because we should make the decision of u_t based on g_t as the past controls $u_1, \ldots, u_{t-1}$ are also taken into consideration when estimating the conditional distribution of θ. We write H_t for $(X_1, \ldots, X_t)$ and $\widehat{H}_t$ for $(\widehat{X}_1, \ldots, \widehat{X}_t)$.

A *history-dependent admissible policy* $\pi = (\pi_1, \ldots, \pi_T)$ is a sequence of functions $\pi_t(g_t)$ such that $\pi_t(g_t) \in \mathcal{U}_t(x_t)$ for all possible g_t. One can easily prove that for such an admissible policy π, each π_t reduces to a function of $h_t = (x_1, x_2, \ldots, x_t)$,[1] as $u_s = \pi_s(x_1, \ldots, x_s)$ for all $s = 1, \ldots, t-1$. Therefore the set of admissible policies is

$$\Pi = \left\{ \pi = (\pi_1, \ldots, \pi_T) \ :\ \pi_t(x_1, \ldots, x_t) \in \mathcal{U}_t(x_t),\ t \in \mathcal{T} \right\}.$$

Any policy $\pi \in \Pi$ defines the control process, also denoted by $\pi = \{\pi_t\}_{t\in\mathcal{T}}$, with $\pi_t = \pi_t(X_1, \ldots, X_t)$. We make a distinction between $u_t = \pi_t(x_1, \ldots, x_t)$ and $\pi_t = \pi_t(X_1, \ldots, X_t)$.

As said in the Introduction, even though we work with discrete spaces $\mathcal{X}$ and $\widehat{\mathbf{\Theta}}$, we are using the more convenient integral notation, rather than the summation notation.

[1] We are still using π_s to denote the decision rule; it will not lead to any misunderstanding.

For a fixed initial state x_1, every policy $\pi \in \Pi$, and every $\theta \in \widehat{\mathbf{\Theta}}$, a probability measure P_θ^π on $(\Omega, \mathcal{F})$ is uniquely defined by:

$$\begin{aligned} P_\theta^\pi(A_1 \times A_2 \times \ldots \times A_{T-1} \times A_T) \\ = \int_{A_1} \int_{A_2} \cdots \int_{A_{T-1}} K_\theta(A_T | x_{T-1}, \pi_{T-1}(x_1, \ldots, x_{T-1})) \\ \times K_\theta(dx_{T-1} | x_{T-2}, \pi_{T-2}(x_1, \ldots, x_{T-2})) \times \cdots \\ \times K_\theta(dx_2 | x_1, \pi_1(x_1)) \delta_{x_1}(dy), \qquad A_t \subset \mathcal{X},\ t \in \mathcal{T}, \end{aligned} \tag{2.1}$$

where, as usual, δ_x denotes the Dirac measure concentrated at x. In particular,

$$P_\theta^\pi(A) = P_\theta^\pi(X \in A) = P_\theta^\pi(\{\omega \in \Omega : X(\omega) \in A\}), \quad A \subset \Omega.$$

The true but unknown measure under the policy π is $P_{\theta^*}^\pi$. This measure gives the true law of the canonical process X subject to the control strategy π.

Given the prior distribution ξ_1, a probability measure P^π on $(\widehat{\Omega}, \widehat{\mathcal{F}})$ is defined as well:

$$P^\pi(A \times D) = \int_D P_\theta^\pi(A)\, \xi_1(d\theta), \quad A \subset \Omega,\ D \subset \widehat{\mathbf{\Theta}}. \tag{2.2}$$

In particular,

$$P^\pi(A \times D) = P^\pi(\{\widehat{\omega} \in \widehat{\Omega} : \widehat{X}(\widehat{\omega}) \in A,\ \Theta(\widehat{\omega}) \in D\}).$$

Clearly, ξ_1 is the marginal of P^π, that is $\xi_1(D) = P^\pi(\Omega \times D)$. To simplify the ensuing study, we assume that for any $t \in \mathcal{T}$ and $h_t \in \mathcal{H}_t$ we have $P^\pi(\widehat{H}_t = h_t) > 0$. This assumption is of course an assumption about the kernels $K_\theta, \theta \in \widehat{\mathbf{\Theta}}$.

Furthermore, for each $t = 1, \ldots, T-1$ and for each history $h_t \in \mathcal{H}_t$, we define the set of *tail control strategies*

$$\begin{aligned} \Pi^{t,h_t} = \{\pi^{t,h_t} : \pi_t^{t,h_t} = \pi_t(h_t),\ \pi_s^{t,h_t}(x_{t+1}, \ldots, x_s) \\ = \pi_s(h_t, x_{t+1}, \ldots, x_s),\ s \in \mathcal{T}_{t+1},\ \pi \in \Pi\}. \end{aligned}$$

In addition, for $t = 1, \ldots, T-1$, and for each $\theta \in \widehat{\mathbf{\Theta}}$, $h_t \in \mathcal{H}_t$, and $\pi^{t,h_t} \in \Pi^{t,h_t}$ we construct a probability measure $P_{\theta,t+1,T}^{\pi^{t,h_t}}$ on $\mathcal{X}^{T-t}$ in analogy to (2.1). Specifically, we put

$$\begin{aligned} P_{\theta,t+1,T}^{\pi^{t,h_t}}(A_{t+1} \times \cdots \times A_T) \\ = \int_{A_{t+1}} \int_{A_{t+2}} \cdots \int_{A_{T-1}} K_\theta(A_T | x_{T-1}, \pi_{T-1}(h_t, x_{t+1}, \ldots, x_{T-1})) \\ \cdot K_\theta(dx_{T-1} | x_{T-2}, \pi_{T-2}(h_t, x_{t+1}, \ldots, x_{T-2})) \cdots K_\theta(dx_{t+2} | x_{t+1}, \pi_{t+1}(h_t, x_{t+1})) \\ \cdot K_\theta(dx_{t+1} | x_t, \pi_t(h_t)), \qquad A_s \subset \mathcal{X},\ s \in \mathcal{T}_{t+1}. \end{aligned} \tag{2.3}$$

We proceed with three technical results that are rather straightforward consequences of the above set-up.

Lemma 2.2 *For any $h_t \in \mathcal{H}_t$, and $A_s \subset \mathcal{X}$, $s \in \mathcal{T}_{t+1}$, $\pi \in \Pi$, and the corresponding $\pi^{t,h_t} \in \Pi^{t,h_t}$ we have that*

$$P^{\pi^{t,h_t}}_{\theta,t+1,T}(A_{t+1} \times \ldots \times A_T) = P^{\pi}_{\theta}(X_{t+1} \in A_{t+1}, \ldots, X_T \in A_T | H_t = h_t). \quad (2.4)$$

Proof First, note that[2]

$$\begin{aligned} &P^{\pi}_{\theta}(X_1 = x_1, \ldots, X_t = x_t) \\ &= K_\theta(x_2|x_1, \pi_1(x_1))\, K_\theta(x_3|x_2, \pi_2(x_1, x_2)) \cdots \\ &\quad K_\theta(x_t|x_{t-1}, \pi_{t-1}(x_1, \ldots x_t)). \end{aligned}$$

On the other hand,

$$\begin{aligned} &{}^{\pi}_{\theta}(X_{t+1} \in A_{t+1}, \ldots, X_T \in A_T, H_t = h_t) \\ &\quad = P^{\pi}_{\theta}(X_1 = x_1, \ldots, X_t = x_t, X_{t+1} \in A_{t+1}, \ldots, X_T \in A_T) \\ &\quad = \int_{\{x_1\}} \ldots \int_{\{x_t\}} P^{\pi^{t,\bar{h}_t}}_{\theta,t+1,T}(A_{t+1} \times \cdots \times A_T) \\ &\qquad K_\theta(d\bar{x}_t|\bar{x}_{t-1}, \pi_{t-1}(\bar{h}_{t-1})) \cdots K_\theta(d\bar{x}_2|\bar{x}_1, \pi_1(\bar{x}_1))\delta_{x_1}(y) \\ &\quad = P^{\pi^{t,h_t}}_{\theta,t+1,T}(A_{t+1} \times \cdots \times A_T) K_\theta(x_t|x_{t-1}, \pi_{t-1}(h_{t-1})) \cdots K_\theta(x_2|x_1, \pi_1(x_1)). \end{aligned}$$

Combining the above we immediately have (2.4). □

For future reference we denote by $P^{\pi^{t,h_t}}_{\theta,t+1}$ a measure on $(\mathcal{X}, 2^{\mathcal{X}})$ defined as

$$P^{\pi^{t,h_t}}_{\theta,t+1}(B) = P^{\pi^{t,h_t}}_{\theta,t+1,T}(B \times \mathcal{X}^{T-t-1}). \quad (2.5)$$

Thus, we have that, for $t \leq T-1$,

$$\begin{aligned} P^{\pi^{t,h_t}}_{\theta,t+1}(B) &= \int_B K_\theta(dx_{t+1} \mid x_t, \pi_t(h_t)) = K_\theta(B \mid x_t, \pi_t(h_t)) \\ &= P^{\pi}_{\theta}(X_{t+1} \in B | H_t = h_t). \end{aligned} \quad (2.6)$$

Next, we construct a probability measure $P^{\pi^{t,h_t}}_{t+1,T}$ on $\mathcal{X}^{T-t} \times \widehat{\mathbf{\Theta}}$ as

$$P^{\pi^{t,h_t}}_{t+1,T}(A \times D) = \int_D P^{\pi^{t,h_t}}_{\theta,t+1,T}(A)\, \xi^{\pi,h_t}_t(d\theta), \quad A \in \underbrace{2^{\mathcal{X}} \otimes \cdots \otimes 2^{\mathcal{X}}}_{T-t \text{ times}},\ D \in 2^{\widehat{\mathbf{\Theta}}}, \quad (2.7)$$

[2] To further simplify the notation we write $K_\theta(x|\ldots)$ in place of $K_\theta(\{x\}|\ldots)$. In a similar way, for a probability measure Q on $\mathcal{Q}$, and $y \in \mathcal{Q}$, we may write $Q(y)$ instead of $Q(\{y\})$.

where $\xi_t^{\pi,h_t} \in \mathcal{P}(\widehat{\mathbf{\Theta}})$, is given as

$$\xi_t^{\pi,h_t}(D) = P^\pi(\Theta \in D \mid \widehat{H}_t = h_t), \text{ for } t = 2, \dots, T, \quad \text{and} \quad \xi_1^{\pi,h_1}(D) = \xi_1(D). \tag{2.8}$$

With some abuse of terminology, we define a conditional measure

$$P_{t+1,T}^{\pi^t,h_t}(A \mid \Theta = \theta) := \frac{P_{t+1,T}^{\pi^t,h_t}(A \times \{\theta\})}{P_{t+1,T}^{\pi^t,h_t}(\mathcal{X}^{T-t} \times \{\theta\})}, \quad \theta \in \mathbf{\Theta},\ A \in 2^{\mathcal{X}} \otimes \cdots \otimes 2^{\mathcal{X}}. \tag{2.9}$$

Then, clearly,

$$P_{t+1,T}^{\pi^t,h_t}(A \mid \Theta = \theta) = P_{\theta,t+1,T}^{\pi^t,h_t}(A), \quad \theta \in \mathbf{\Theta},\ A \in 2^{\mathcal{X}} \otimes \cdots \otimes 2^{\mathcal{X}}. \tag{2.10}$$

Occasionally, in what follows we will use a simplified notation $P_{t+1,T\mid\theta}^{\pi^t,h_t}(A)$ for $P_{t+1,T}^{\pi^t,h_t}(A \mid \Theta = \theta)$

Lemma 2.3 *Let* $A \in \underbrace{2^{\mathcal{X}} \otimes \cdots \otimes 2^{\mathcal{X}}}_{T-t \text{ times}}$, $D \in 2^{\widehat{\mathbf{\Theta}}}$, $h_t \in \mathcal{H}_t$, *and* $\pi \in \Pi$. *Then*

$$P_{t+1,T}^{\pi^t,h_t}(A \times D) = P^\pi\big((\widehat{X}_{t+1}, \widehat{X}_{t+2}, \dots, \widehat{X}_{T-1}, \widehat{X}_T) \in A, \widehat{\mathbf{\Theta}} \in D \,\big|\, \widehat{H}_t = h_t\big). \tag{2.11}$$

Proof First, in view of (2.8) and (2.2) we note that

$$\xi_t^{\pi,h_t}(D) = \frac{\int_D P_\theta^\pi(H_t = h_t)\, \xi_1(d\theta)}{P^\pi(\widehat{H}_t = h_t)}.$$

Thus,

$$P_{t+1,T}^{\pi^t,h_t}(A \times D) = \frac{\int_D P_{\theta,t+1,T}^{\pi^t,h_t}(A) P_\theta^\pi(H_t = h_t)\, \xi_1(d\theta)}{P^\pi(\widehat{H}_t = h_t)}. \tag{2.12}$$

On the other hand, using (2.2) and Lemma 2.2, we have

$$\begin{aligned}
&P^{\pi}((\widehat{X}_{t+1}, \widehat{X}_{t+2}, \ldots, \widehat{X}_{T-1}, \widehat{X}_T) \in A, \Theta \in D | \widehat{H}_t = h_t) \\
&= \frac{\int_D P^{\pi}_{\theta}((X_{t+1}, \ldots, X_T) \in A, H_t = h_t)\, \xi_1(d\theta)}{P^{\pi}(\widehat{H}_t = h_t)} \\
&= \frac{\int_D P^{\pi}_{\theta}((X_{t+1}, \ldots, X_T) \in A \mid H_t = h_t) P^{\pi}_{\theta}(H_t = h_t)\, \xi_1(d\theta)}{P^{\pi}(\widehat{H}_t = h_t)} \\
&= \frac{\int_D P^{\pi^{t,h_t}}_{\theta,t+1}(A) P^{\pi}_{\theta}(H_t = h_t)\, \xi_1(d\theta)}{P^{\pi}(\widehat{H}_t = h_t)}.
\end{aligned}$$

This, combined with (2.12) concludes the proof of part (i). □

Remark 2.4 Formally, taking $T = t + 1$ in (2.10) we obtain

$$P^{\pi^{t,h_t}}_{t+1,t+1|\theta}(A) = P^{\pi^{t,h_t}}_{\theta,t+1,t+1}(A) = P^{\pi^{t,h_t}}_{\theta,t+1}(A), \tag{2.13}$$

where in the last equality we used (2.5).

Lemma 2.5 *Let $t \in \{1, \ldots, T-1\}$, and let F be a function on $\mathcal{X}^{T-t} \times \widehat{\mathbf{\Theta}}$. Then, for each $h_t \in \mathcal{H}_t$ we have*

$$\begin{aligned}
&\mathbb{E}^{\pi}[F(\widehat{X}_{t+1}, \widehat{X}_{t+2}, \ldots, \widehat{X}_{T-1}, \widehat{X}_T, \Theta) \mid \widehat{H}_t = h_t] = \\
&\int_{\widehat{\mathbf{\Theta}}} \int_{\mathcal{X}^{T-t}} F(x_{t+1}, \ldots, x_T, \theta) P^{\pi^{T-1,h_{T-1}}}_{\theta,T}(dx_T) \cdots P^{\pi^{t,h_t}}_{\theta,t+1}(dx_{t+1}) \xi^{\pi,h_t}_t(d\theta),
\end{aligned} \tag{2.14}$$

where $\mathbb{E}^{\pi}$ denotes the expectation with respect to probability P^{π}.

Proof In view of Lemma 2.3, we have

$$P^{\pi}(dx_{t+1}, \ldots, dx_T; d\theta \mid \widehat{H}_t = h_t) = P^{\pi^{t,h_t}}_{t+1,T}(dx_{t+1}, \ldots, dx_T; d\theta).$$

Consequently, by (2.7), we continue

$$P^{\pi^{t,h_t}}_{t+1,T}(dx_{t+1}, \ldots, dx_T; d\theta) = P^{\pi^{t,h_t}}_{\theta,t+1,T}(dx_{t+1}, \ldots, dx_T) \xi^{\pi^{t,h_t}}_t(d\theta).$$

This combined with (2.3) and (2.6) yields the identity (2.14). □

For future reference we denote by $P^{\pi^{t,h_t}}_{t+1}$ the measure on $\mathcal{X} \times \widehat{\mathbf{\Theta}}$ defined as

$$P^{\pi^{t,h_t}}_{t+1}(B \times D) = P^{\pi^{t,h_t}}_{t+1,T}(B \times \mathcal{X}^{T-t-1} \times D) = P^{\pi}(\widehat{X}_{t+1} \in B, \Theta \in D \mid \widehat{H}_t = h_t), \tag{2.15}$$

for $t = 1, \ldots, T-1$, with the convention, employed throughout, that $B \times \mathcal{X}^0 \times D = B \times D$. The second equality in (2.15) follows from (2.11).

For $t = T$ and $h_T \in \mathcal{H}_T$ we construct a measure $P_{T+1,T}^{\pi^{T,h_T}}$ on $\widehat{\mathbf{\Theta}}$ as

$$P_{T+1,T}^{\pi^{T,h_T}}(D) = P^{\pi}(\Theta \in D \mid \widehat{H}_T = h_T) = \xi_T^{\pi,h_T}(D).$$

Given that a strategy π is used, then at each time $t \in \mathcal{T}$ a random cost $Z_{\theta^*,t}^{\pi}$ is incurred, with

$$Z_{\theta,t}^{\pi} = c_t(X_t, \pi_t, \theta),$$

where $c_t : \mathcal{X} \times \mathcal{U} \times \widehat{\mathbf{\Theta}} \to \mathbb{R}_+$.

Remark 2.6 It is important to note that even though X_t and π_t are observed at time t, the actual cost $c_t(X_t, \pi_t, \theta^*)$ is not known (or observed) at time t as θ^* is not known. The dependence of both the transition kernel and the accrued costs on the unknown parameter is an important practical situation, leading to non-standard and non-trivial risk-averse Markov decision problems.

To proceed, for each $t \in \mathcal{T}$ and each history $h_t \in \mathcal{H}_t$ we denote

$$Z_{\theta,t,t}^{\pi,h_t} = c_t(x_t, \pi_t(h_t), \theta), \tag{2.16}$$

and for each $s = t+1, \ldots, T$ we put

$$Z_{\theta,t,s}^{\pi,h_t,x_{t+1},\ldots,x_s} = c_s(x_s, \pi_s^{t,h_t}(x_{t+1}, \ldots, x_s), \theta). \tag{2.17}$$

Note that, for a fixed strategy π and a fixed $h_t \in \mathcal{H}_t$, we have that $c_t(x_t, \pi_t(h_t), \cdot)$ is a function on $\widehat{\mathbf{\Theta}}$, and $c_s(\cdot, \pi_s^{t,h_t}(\cdot, \ldots, \cdot), \theta)$ is a function on $\mathcal{X}^{s-t} \times \widehat{\mathbf{\Theta}}$.

3 Risk filters for MDPs with model uncertainty

An underlying feature of our approach is a desire to assess the riskiness of the uncertain costs induced by any policy π in a time-consistent way. This desire is fulfilled via the concept of a time-consistent dynamic risk filter, also satisfying additional properties of normalization, monotonicity, translation invariance, support, and parameter consistency.

3.1 Dynamic risk filters

For $t = 1, \ldots, T-1$ and $s = t, \ldots, T$, we denote by $\mathcal{Z}_t^{\mathcal{X}}$ and $\mathcal{Z}_{t,s}$ the spaces of real valued functions on $\mathcal{X}^t$ and $\mathcal{X}^{s-t} \times \widehat{\mathbf{\Theta}}$, respectively, where $\mathcal{X}^0 \times \widehat{\mathbf{\Theta}} := \widehat{\mathbf{\Theta}}$, so that $\mathcal{Z}_{t,t}$ is the space of real valued functions on $\widehat{\mathbf{\Theta}}$. For $Z_{t,s}, W_{t,s} \in \mathcal{Z}_{t,s}$, the comparison between these functions is understood point-wise; $Z_{t,s} \leq W_{t,s}$ means that $Z_{t,s}(x_{t+1}, \ldots, x_s, \theta) \leq W_{t,s}(x_{t+1}, \ldots, x_s, \theta)$ for all $(x_{t+1}, \ldots, x_s, \theta) \in \mathcal{X}^{s-t} \times \widehat{\mathbf{\Theta}}$.

For any policy $\pi \in \Pi$, our objective is to evaluate at each time $t \in \mathcal{T}$, the riskiness of the sequence of costs $Z^{\pi,h_t}_{\Theta,t}, Z^{\pi,h_t,X_{t+1}}_{\Theta,t,t+1}, \ldots, Z^{\pi,h_t,X_{t+1},\ldots,X_T}_{\Theta,t,T}$, given history h_t, in such a way that the evaluation is $\mathcal{F}^X_t$-measurable. We denote by

$$\mathcal{Z}^{t,T} = \mathcal{Z}_{t,t} \times \mathcal{Z}_{t,t+1} \times \cdots \times \mathcal{Z}_{t,T}$$

the space of *conditional cost functions*[3] in periods $t, \ldots, T$.

For $t = 1, \ldots, T$ and $s = t, \ldots, T$, we also use $\mathcal{P}_{t,s}$ to denote the space of probability measures on the space of paths starting at time t and ending at time s, and on realizations of the parameter θ, that is on the space $\mathcal{X}^{s-t+1} \times \widehat{\boldsymbol{\Theta}}$. Additionally, we understand $\mathcal{P}_{T+1,T}$ as $\mathcal{P}(\widehat{\boldsymbol{\Theta}})$, because no future paths are possible. Note, in particular, that $P^{\pi^t,h_t}_{t+1,T} \in \mathcal{P}_{t+1,T}$, for $t \in \mathcal{T}$, and $P^{\pi^t,h_t}_{t+1,t+1} \in \mathcal{P}_{t+1,t+1}$.

Observe that at time $t = 1, \ldots, T-1$ we know the history h_t, and, for any policy $\pi \in \Pi$ (in principle), we can evaluate the distribution of $(X_{t+1}, \ldots, X_T, \Theta)$ under the measure $P^{\pi^t,h_t}_{t+1,T}$.

We proceed with stating three key definitions.

Definition 3.1 For a fixed $t \in \mathcal{T}$, a mapping $\rho_t : \mathcal{Z}^{t,T} \times \mathcal{P}_{t+1,T} \to \mathbb{R}$, is called a *conditional risk filter*.

It should be stressed that the concept of a conditional risk filter is substantially different than that of a conditional measure of risk, discussed in the risk theory literature, such as, *inter alia*, Scandolo (2003); Frittelli and Scandolo (2006); Cheridito et al. (2006); Ruszczyński and Shapiro (2006); Artzner et al. (2007); Ch et al. (2007); Shapiro et al. (2021), because it is real-valued and admits the probability measure as its second argument, which is pertinent to our setting. It is closely related to the concept of a *risk form* recently introduced in Dentcheva and Ruszczyński (2020). Note that, in particular, for any $(Z_{t,t}, \ldots, Z_{t,T}) \in \mathcal{Z}^{t,T}$ and $\pi \in \Pi$ we have

$$\rho_t(Z^{\pi,h_t}_{\theta,t,t}, \ldots, Z^{\pi,h_t,x_{t+1},\ldots,x_T}_{\theta,t,T}; P^{\pi^t,h_t}_{t+1,T}) = R(h_t), \text{ for all } h_t \in \mathcal{X}^t,$$

for some function $R : \mathcal{X}^t \to \mathbb{R}$. This explains the term *conditional* in the conditional risk filter. Also, it needs to be stressed that conditional risk filters should not be confused with the linear or non-linear filters that are studied in the filtering theory (see e.g. Krishnamurthy (2016)).

Definition 3.2 Let $t \in \mathcal{T}$. A conditional risk filter ρ_t

i is *normalized* if $\rho_t(0, 0, \ldots, 0; P_{t+1,T}) = 0$ for all $P_{t+1,T} \in \mathcal{P}_{t+1,T}$;

ii is *monotonic* if $\rho_t(Z_{t,t}, \ldots, Z_{t,T}; P_{t+1,T}) \leq \rho_t(W_{t,t}, \ldots, W_{t,T}; P_{t+1,T})$ for all $P_{t+1,T} \in \mathcal{P}_{t+1,T}$, and all $(Z_{t,t}, \ldots, Z_{t,T})$ and $(W_{t,t}, \ldots, W_{t,T})$ in $\mathcal{Z}^{t,T}$, such that $Z_{t,s} \leq W_{t,s}$ for all $s \in \mathcal{T}_t$;

[3] The term *conditional* refers to the fact that at any time t we consider cost functions that depend on a history h_t.

iii is *translation invariant* if for all $(Z_{t,t}, \ldots, Z_{t,T}) \in \mathcal{Z}^{t,T}$, all $V \in \mathbb{R}$, and all $P_{t+1,T} \in \mathcal{P}_{t+1,T}$,

$$\rho_t(V + Z_{t,t}, Z_{t,t+1}, \ldots, Z_{t,T}; P_{t+1,T}) = V + \rho_t(Z_{t,t}, Z_{t,t+1}, \ldots, Z_{t,T}; P_{t+1,T});$$

iv has the *support property*, if

$$\rho_t\big(Z_{t,t}, \ldots, Z_{t,T}; P_{t+1,T}\big) = \rho_t\big(Z_{t,t}\mathbb{1}_{\operatorname{supp}_t(P_{t+1,T})}, \ldots, Z_{t,T}\mathbb{1}_{\operatorname{supp}_T(P_{t+1,T})}; P_{t+1,T}\big),$$

for all $(Z_{t,t}, \ldots, Z_{t,T}) \in \mathcal{Z}^{t,T}$, and all $P_{t+1,T} \in \mathcal{P}_{t+1,T}$, and where $\operatorname{supp}_s(P_{t+1,T})$ denotes the projection of $\operatorname{supp}(P_{t+1,T})$ on $\mathcal{X}^{s-t} \times \widehat{\mathbf{\Theta}}$, for $s \geq t$.

Remark 3.3 Let $s = t, \ldots, T$ and let $\{Z^y_{s,T}, y \in \mathcal{Y}\}$ be a family of functions parameterized by y, for some non-empty set $\mathcal{Y}$. Then, by the normalization property, for any $A \subset \mathcal{Y}$, $y \in \mathcal{Y}$, and $P \in \mathcal{P}_{t+1,T}$, we have that

$$\mathbb{1}_A(y)\rho_t(Z^y_{t,T}, \ldots, Z^y_{T,T}; P) = \rho_t(\mathbb{1}_A(y)Z^y_{t,T}, \ldots, \mathbb{1}_A(y)Z^y_{T,T}; P).$$

Definition 3.4 A *dynamic risk filter* $\rho = \{\rho_t\}_{t\in\mathcal{T}}$ is a sequence of conditional risk filters $\rho_t : \mathcal{Z}^{t,T} \times \mathcal{P}_{t+1,T} \to \mathbb{R}$. We say that it is normalized, monotonic, translation invariant, or has the support property, if all $\rho_t, t \in \mathcal{T}$, satisfy the respective conditions of Definition 3.2.

3.2 Parameter consistency

Let $t = 1, \ldots, T-1$ and $s = t, \ldots, T$. For any probability measure $P_{t,s} \in \mathcal{P}_{t,s}$, we denote by $P_{t,s|\Theta}(\cdot,\cdot)$, the stochastic kernel from $\widehat{\mathbf{\Theta}}$ to $\mathcal{X}^{s-t+1}$ defined as[4]

$$P_{t,s|\Theta}(\theta, A) = \frac{P_{t,s}(A \times \{\theta\})}{P_{t,s}(\mathcal{X}^{s-t+1} \times \{\theta\})}, \quad A \subset \mathcal{X}^{s-t+1},\ \theta \in \widehat{\mathbf{\Theta}}. \tag{3.1}$$

The corresponding marginal on $\widehat{\mathbf{\Theta}}$ is denoted by $P_{t,s,\Theta}$, so that

$$P_{t,s,\Theta}(D) = P_{t,s}(\mathcal{X}^{s-t+1} \times D), \quad D \subset \widehat{\mathbf{\Theta}}. \tag{3.2}$$

Clearly, the measure $P_{t,s}$ admits the disintegration

$$P_{t,s}(A \times B) = \int_B P_{t,s|\theta}(A) P_{t,s,\Theta}(d\theta) =: P_{t,s,\Theta} \circledast P_{t,s|\Theta}(A \times B),$$

where we use a simplified notation

$$P_{t,s|\theta}(A) := P_{t,s|\Theta}(\theta, A). \tag{3.3}$$

[4] Note that (2.9) is an example of the above.

We note that for any stochastic kernel $\kappa_{s,t}(\cdot,\cdot)$ from $\widehat{\mathbf{\Theta}}$ to $\mathcal{X}^{s-t}$ and for any probability measure μ on $2^{\widehat{\mathbf{\Theta}}}$ one can construct a unique probability measure on the product space $\mathcal{X}^{s-t+1} \times \mathbf{\Theta}$ as

$$m_{t,s}(A \times B) = \int_B \kappa_{t,s}(\theta, A)\, \mu(d\theta) =: \mu \circledast \kappa_{t,s}(A \times B).$$

In particular, with $\mu = \delta_\theta$ and $\kappa_{t,s} = P_{t,s|\Theta}$, with $P_{t,s} \in \mathcal{P}_{t,s}$, we get

$$m_{t,s}(A \times B) = \delta_\theta \circledast P_{t,s|\Theta}(A \times B) = P_{t,s|\theta}(A)\mathbb{1}_B(\theta) = P_{t,s|\theta}(A)\delta_\theta(B). \quad (3.4)$$

Remark 3.5 In our convention, $P_{T+1,T}(\cdot)$ is a measure on $\widehat{\mathbf{\Theta}}$. This means that, formally, $P_{T+1,T,\mathbf{\Theta}}$
$= P_{T+1,T}$ and $P_{T+1,T|\mathbf{\Theta}} \equiv 1$, in which case (formally)

$$\delta_\theta \circledast P_{T+1,T|\Theta} = \delta_\theta.$$

Example 3.6 Fix $t \in \mathcal{T}$, $h_t \in \mathcal{H}_t$ and $\pi \in \Pi$. Take $P_{t+1,T} = P^{\pi^{t,h_t}}_{t+1,T} \in \mathcal{P}_{t+1,T}$ and $P_{t+1,t+1} = P^{\pi^{t,h_t}}_{t+1,t+1} \in \mathcal{P}_{t+1,t+1}$. Then,

$$P_{t+1,T|\theta} = P^{\pi^{t,h_t}}_{\theta,t+1,T}, \quad P_{t+1,t+1|\theta} = P^{\pi^{t,h_t}}_{\theta,t+1}, \quad P_{t+1,T,\Theta} = \xi^{\pi^{t,h_t}}_t. \quad (3.5)$$

The first equality above comes from (2.10). The second one comes from (2.13). The third one is just (2.7) with $A = \mathcal{X}^{T-t}$. Note that (3.4) and (3.5) imply that

$$\begin{aligned} \delta_\theta \circledast P^{\pi^{t,h_t}}_{t+1,T|\Theta}(A \times B) &= \delta_\theta \otimes P^{\pi^{t,h_t}}_{\theta,t+1,T}(A \times B), \\ \delta_\theta \circledast P^{\pi^{t,h_t}}_{t+1,t+1|\Theta}(A \times B) &= \delta_\theta \otimes P^{\pi^{t,h_t}}_{\theta,t+1,t+1}(A \times B). \end{aligned} \quad (3.6)$$

We introduce the following key concept.

Definition 3.7 A conditional risk filter $\rho_t : \mathcal{Z}^{t,T} \times \mathcal{P}_{t+1,T} \to \mathbb{R}$ is *parameter consistent*, if for all $(Z_{t,t}, \ldots, Z_{t,T})$, $(W_{t,t}, \ldots, W_{t,T}) \in \mathcal{Z}^{t,T}$, and all $P_{t+1,T}, Q_{t+1,T} \in \mathcal{P}_{t+1,T}$ the relations

$$P_{t+1,T,\Theta} = Q_{t+1,T,\Theta}$$

and

$$\rho_t\big(Z_{t,t}, \ldots, Z_{t,T}; \delta_\theta \circledast P_{t+1,T|\Theta}\big) \le \rho_t\big(W_{t,t}, \ldots, W_{t,T}; \delta_\theta \circledast Q_{t+1,T|\Theta}\big), \quad \text{for all } \theta \in \widehat{\mathbf{\Theta}}, \quad (3.7)$$

imply that

$$\rho_t\big(Z_{t,t}, \ldots, Z_{t,T}; P_{t+1,T}\big) \le \rho_t\big(W_{t,t}, \ldots, W_{t,T}; Q_{t+1,T}\big). \quad (3.8)$$

In words, if the marginal distributions of P and Q on $\widehat{\boldsymbol{\Theta}}$ are the same, and the conditional risk of $Z_{t:T} := (Z_{t,t}, \dots, Z_{t,T})$ under P is not greater than that of $W_{t:T} := (W_{t,t}, \dots, W_{t,T})$ under Q for every value of θ, then the risk of $Z_{t:T}$ under P should be not greater than that of $W_{t:T}$ under Q.

Remark 3.8 Note that parameter consistency at $t = T$ follows from the support property, translation invariance, monotonicity, and normalization of ρ_T. Indeed, first observe that according to Remark 3.5 the equality $P_{T+1,T,\Theta} = Q_{T+1,T,\Theta} = 1$ implies that $P_{T+1,T} = Q_{T+1,T}$. Thus, for any $\theta \in \widehat{\boldsymbol{\Theta}}$

$$\begin{aligned}\rho_T(Z_{T,T}, \delta_\theta) \le \rho_T(W_{T,T}, \delta_\theta) \quad &\Leftrightarrow \quad \rho_T(Z_{T,T}(\theta), \delta_\theta) \le \rho_T(W_{T,T}(\theta), \delta_\theta) \quad \Leftrightarrow \\ Z_{T,T}(\theta) + \rho_T(0, \delta_\theta) \le W_{T,T}(\theta) + \rho_T(0, \delta_\theta) \quad &\Leftrightarrow \quad Z_{T,T}(\theta) \le W_{T,T}(\theta).\end{aligned}$$

By monotonicity, we have that

$$\rho_T(Z_{T,T}, P_{T+1,T}) \le \rho_T(W_{T,T}, P_{T+1,T}) = \rho_T(W_{T,T}, Q_{T+1,T}).$$

This remark is used in Proposition 3.10 and also in Theorem 3.14.

We have the following risk decomposition formula.

Theorem 3.9 *Take $t = 1, \dots, T$. If a conditional risk filter $\rho_t : \mathcal{Z}^{t,T} \times \mathcal{P}_{t+1,T} \to \mathbb{R}$ is parameter consistent, then there exists a mapping $\widehat{\rho}_t : \mathcal{Z}_{t,t} \times \mathcal{P}(\widehat{\boldsymbol{\Theta}}) \to \mathbb{R}$ such that for all $Z_{t:T}$ and $P_{t+1,T}$,*

$$\begin{aligned}&\rho_t\big(Z_{t,t}, \dots, Z_{t,T}; P_{t+1,T}\big) \\ &\quad = \widehat{\rho}_t\Big(\{\rho_t\big(Z_{t,t}, \dots, Z_{t,T}; \delta_\theta \circledast P_{t+1,T|\Theta}\big),\ \theta \in \widehat{\boldsymbol{\Theta}}\}; P_{t+1,T,\Theta}\Big). \qquad (3.9)\end{aligned}$$

Proof Suppose two sequences $Z_{t:T}$ and $W_{t:T}$ in $\mathcal{Z}^{t,T}$, and two measures $P_{t+1,T}$ and $Q_{t+1,T}$ in $\mathcal{P}_{t+1,T}$ are such that $P_{t+1,T,\widehat{\boldsymbol{\Theta}}} = Q_{t+1,T,\widehat{\boldsymbol{\Theta}}}$ and

$$\rho_t\big(Z_{t,t}, \dots, Z_{t,T}; \delta_\theta \circledast P_{t+1,T|\Theta}\big) = \rho_t\big(W_{t,t}, \dots, W_{t,T}; \delta_\theta \circledast Q_{t+1,T|\Theta}\big), \quad \forall\, \theta \in \widehat{\boldsymbol{\Theta}}.$$

Then it follows from Definition 3.7 that

$$\rho_t\big(Z_{t,t}, \dots, Z_{t,T}; P_{t+1,T}\big) = \rho_t\big(W_{t,t}, \dots, W_{t,T}; Q_{t+1,T}\big).$$

This means that formula (3.9) is true. □

Thus, parameter consistency allows us to disintegrate the risk filtering task into two stages. First, we evaluate the risk in a fully observed system, with the parameter θ fixed, and then we integrate the results by using the operator $\widehat{\rho}_t$, which we call the *marginal risk filter*.

Proposition 3.10 *Take $t = 1, \dots, T$. If the conditional risk filter ρ_t is parameter consistent, normalized, monotonic, and has the translation invariant property, or the support property, then the mapping $\widehat{\rho}_t(\cdot; \cdot)$ has the corresponding properties as well (in the sense indicated in the proof below).*

Proof Indeed, consider any measure $\Lambda \in \mathcal{P}(\widehat{\mathbf{\Theta}})$. Then for any $P_{t+1,T} \in \mathcal{P}_{t+1,T}$ such that $P_{t+1,T,\Theta} = \Lambda$, we will use the formula (3.9) to analyze the implied properties of $\widehat{\rho}_t$.

1) Suppose ρ_t is normalized. Then (the symbol $\mathbf{0}$ below denotes a function on $\widehat{\mathbf{\Theta}}$ that is identically equal to zero)

$$\begin{aligned}\widehat{\rho}_t(\mathbf{0};\Lambda) &= \widehat{\rho}_t\Big(\{\rho_t\big(0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\widehat{\mathbf{\Theta}}}\big),\ \theta\in\widehat{\mathbf{\Theta}}\};\ P_{t+1,T,\Theta}\Big)\\ &= \rho_t\big(0,\ldots,0;\ P_{t+1,T}\big) = 0.\end{aligned}$$

Thus, $\widehat{\rho}$ is normalized.

2) Suppose ρ_t is normalized and translation invariant and has the support property. Then for any $V \in \mathbb{R}$, we have

$$\rho_t\big(V,0,\ldots,0;\ P_{t+1,T}\big) = V + \rho_t\big(0,0,\ldots,0;\ P_{t+1,T}\big) = V.$$

Therefore, for any $U \in \mathcal{Z}^{t,t}$ and any $a \in \mathbb{R}$, by the support, translation invariance, and normalization properties of ρ_t, we have for any $\theta \in \mathbf{\Theta}$,

$$\begin{aligned}\rho_t\big(U+a,0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big) &= \rho_t\big(U(\theta)+a,0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big)\\ &= U(\theta)+a. \qquad (3.10)\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{\rho}_t(U+a;\Lambda) &= \widehat{\rho}_t\Big(\{\rho_t\big(U+a,0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big),\ \theta\in\widehat{\mathbf{\Theta}}\};\ P_{t+1,T,\Theta}\Big)\\ &= \rho_t\big(U+a,0,\ldots,0;\ P_{t+1,T}\big) = a + \rho_t\big(U,0,\ldots,0;\ P_{t+1,T}\big)\\ &= a + \widehat{\rho}_t\Big(\{\rho_t\big(U,0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big),\ \theta\in\widehat{\mathbf{\Theta}}\};\ P_{t+1,T,\Theta}\Big)\\ &= a + \widehat{\rho}_t(U;\Lambda).\end{aligned}$$

Hence, $\widehat{\rho}$ is translation invariant.

Similarly, for any $U \in \mathcal{Z}^{t,t}$, noting that $\mathrm{supp}_{t,t}(P_{t+1,T}) = \mathrm{supp}(\Lambda)$, we deduce

$$\begin{aligned}\widehat{\rho}_t(U,\Lambda) &= \rho_t\big(U,0,\ldots,0;\ P_{t+1,T}\big) = \rho_t\big(\mathbb{1}_{\ \mathrm{supp}_{t,t}(P_{t+1,T})}U,0,\ldots,0;\ P_{t+1,T}\big)\\ &= \rho_t\big(\mathbb{1}_{\ \mathrm{supp}(\Lambda)}U,0,\ldots,0;\ P_{t+1,T}\big) = \widehat{\rho}_t(\mathbb{1}_{\ \mathrm{supp}(\Lambda)}U,\Lambda).\end{aligned}$$

Thus, $\widehat{\rho}_t$ also has the support property.

3) Suppose ρ_t is normalized, translation invariant, and monotonic. Then, for all $U, W \in \mathcal{Z}^{t,t}$ such that $U \le W$, employing (3.10) with $a = 0$, we have

$$\begin{aligned}\widehat{\rho}_t(U;\Lambda) &= \widehat{\rho}_t\Big(\{\rho_t\big(U,0,\dots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big),\ \theta\in\widehat{\mathbf{\Theta}}\};P_{t+1,T,\Theta}\Big)\\ &= \rho_t\big(U,0,\dots,0;P_{t+1,T}\big)\\ &\le \rho_t\big(W,0,\dots,0;P_{t+1,T}\big)\\ &= \widehat{\rho}_t\Big(\{\rho_t\big(W,0,\dots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big),\ \theta\in\widehat{\mathbf{\Theta}}\};P_{t+1,T,\Theta}\Big)\\ &= \widehat{\rho}_t(W;\Lambda),\end{aligned}$$

and this proves the monotonicity of $\widehat{\rho}_t$.
4) If ρ is monotonic, normalized, parameter consistent, translation invariant, and with support property, then $\widehat{\rho}_t$ is also normalized, monotonic, normalized, translation invariant, and has the support property, and in view of Remark 3.8, $\widehat{\rho}_t$ is also parameter consistent. □

3.3 Time consistency

We now consider the notion of time consistency of risk filters.

Definition 3.11 Let $t \in \{1,\dots,T-1\}$. For any positive measure μ_{t+1} on $\mathcal{X}^{T-t}\times\widehat{\mathbf{\Theta}}$ and for any $x\in\mathcal{X}$, we denote by $\mu_{t+1}(\cdot\,\|x)$ the measure on $\mathcal{X}^{T-t-1}\times\widehat{\mathbf{\Theta}}$ given as

$$\mu_{t+1}(A\times B\|x) = \frac{\mu_{t+1}(\{x\}\times A\times B)}{\mu_{t+1}(\{x\}\times\mathcal{X}^{T-t-1}\times\widehat{\mathbf{\Theta}})}. \tag{3.11}$$

Clearly, $\mu_{t+1}(\cdot\,\|x)$ is a probability measure.

In particular, taking $\mu_{t+1} = \delta_\theta \circledast P_{t+1,T|\Theta}$,

$$\delta_\theta \circledast P_{t+1,T|\Theta}(A\times B\,\|x) = \frac{\delta_\theta \circledast P_{t+1,T|\Theta}(\{x\}\times A\times B)}{\delta_\theta \circledast P_{t+1,T|\Theta}(\{x\}\times\mathcal{X}^{T-t-1}\times\widehat{\mathbf{\Theta}})}. \tag{3.12}$$

It follows from (3.4) and (3.12) that

$$\begin{aligned}\delta_\theta \circledast P_{t+1,T|\Theta}(A\times B\,\|\,x) &= \frac{P_{t+1,T|\theta}(\{x\}\times A)\delta_\theta(B)}{P_{t+1,T|\theta}(\{x\}\times\mathcal{X}^{T-t-1})} = \frac{P_{t+1,T|\theta}(\{x\}\times A)}{P_{t+1,t+1|\theta}(\{x\})}\delta_\theta(B)\\ &=: \widetilde{P}_{t+1,T|\theta}(A\,\|\,x)\delta_\theta(B),\end{aligned} \tag{3.13}$$

which, in view of (3.6) gives

$$\delta_\theta \circledast P^{\pi^{t,h_t}}_{t+1,T|\Theta}(A\times B\|x) = \frac{P^{\pi^{t,h_t}}_{\theta,t+1,T}(\{x\}\times A)}{P^{\pi^{t,h_t}}_{\theta,t+1,t+1}(\{x\})}\delta_\theta(B) =: \widetilde{P}^{\pi^{t,h_t}}_{\theta,t+1,T}(A\,\|x)\delta_\theta(B). \tag{3.14}$$

The next definition is a version of the dynamic conditional time consistency used in Fan and Ruszczyński (2018), adapted to the set-up of the present paper. It is significantly different than the standard time consistency concept, discussed, *inter alia*, in Cheridito et al. (2006); Artzner et al. (2007); Cheridito and Kupper (2011); Ruszczyński (2010), by the explicit use of the conditional distributions, and by formulating it for each fixed value of θ.

Definition 3.12 A dynamic risk filter $\rho = \{\rho_t\}_{t=1,\dots,T}$ is *time consistent* if for any $t = 1, \dots,$
$T-1$, for any $P_{t+1,T}, Q_{t+1,T} \in \mathcal{P}_{t+1,T}$, such that $P_{t+1,t+1|\Theta} = Q_{t+1,t+1|\Theta}$, and for any functions[5] $Z_{t,s}(\cdot_{s-t}, \cdot_1), W_{t,s}(\cdot_{s-t}, \cdot_1) \in \mathcal{Z}_{t,s}$, $s = t+1, \dots, T$, the inequalities

$$\begin{aligned}&\rho_{t+1}\big(Z_{t,t+1}(x_{t+1}, \cdot_1), \dots, Z_{t,T}(x_{t+1}, \cdot_{T-t-1}, \cdot_1); \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x_{t+1})\big)\\ &\quad\le \rho_{t+1}\big(W_{t,t+1}(x_{t+1}, \cdot_1), \dots, W_{t,T}(x_{t+1}, \cdot_{T-t-1}, \cdot_1); \delta_\theta \circledast Q_{t+1,T|\Theta}(\cdot\,\|x_{t+1})\big),\\ &\qquad\qquad \forall\,\theta \in \widehat{\boldsymbol{\Theta}}, \quad \forall\, x_{t+1} \in \mathcal{X},\end{aligned} \tag{3.15}$$

imply that for any function $f_t \in \mathcal{Z}_{t,t}$

$$\begin{aligned}&\rho_t\big(f_t(\cdot_1), Z_{t,t+1}(\cdot_1, \cdot_1), \dots, Z_{t,T}(\cdot_{T-t}, \cdot_1)); \delta_\theta \circledast P_{t+1,T|\Theta}\big)\big)\\ &\quad\le \rho_t\big(f_t(\cdot_1), W_{t,t+1}(\cdot_1, \cdot_1), \dots, W_{t,T}(\cdot_{T-t}, \cdot_1)); \delta_\theta \circledast Q_{t+1,T|\Theta}\big)\big), \quad \forall\,\theta \in \widehat{\boldsymbol{\Theta}}.\end{aligned} \tag{3.16}$$

In words, if the risk of the system starting at time $t+1$ from any possible next state x_{t+1} and any possible parameter value θ is lower for Z than for W (if the one-step conditional measures of P and Q coincide), and if Z and W are equal at time t, then the risk at time t is also lower for Z than for W, no matter what the value of θ is.

Lemma 3.13 *Suppose a dynamic risk filter $\{\rho_t\}_{t=1,\dots,T}$ is normalized, translation invariant, has the support property, and is time consistent. Let $\theta \in \widehat{\boldsymbol{\Theta}}$ be fixed. Then the function on $\mathcal{Z}_{t,t+1} \times \mathcal{P}(\mathcal{X}^{T-t} \times \widehat{\boldsymbol{\Theta}})$ given as*

$$\rho_t\big(0, w, 0, \dots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}\big) \tag{3.17}$$

depends only on the probability measure $P_{t+1,t+1|\theta}$ and on the function $w(\cdot, \theta)$.

Proof For any $P_{t+1,T}$ and $x \in \mathcal{X}$, the support property of ρ_{t+1} implies that

$$\begin{aligned}&\rho_{t+1}\big(w(x, \cdot), 0, \dots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\|x)\big)\\ &\quad= \rho_{t+1}\big(w(x, \theta), 0, \dots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\|x)\big).\end{aligned}$$

Then, by the translation invariance and the normalization properties of ρ_{t+1} we obtain

$$\rho_{t+1}\big(w(x, \cdot), 0, \dots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\|x)\big) = w(x, \theta),$$

[5] The notation $\cdot_k$ is a place-holder for k variables.

which does not depend on $P_{t+2,T|\Theta}$. Hence, for any $Q_{t+1,T} \in \mathcal{P}_{t+1,T}$ we have

$$\begin{aligned}&\rho_{t+1}\big(w(x,\cdot),0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x)\big)\\&\quad=\rho_{t+1}\big(w(x,\cdot),0,\ldots,0;\delta_\theta \circledast Q_{t+1,T|\Theta}(\cdot\,\|x)\big).\end{aligned}$$

If in addition, $P_{t+1,t+1|\Theta} = Q_{t+1,t+1|\Theta}$, then, by the time consistency,

$$\rho_t\big(0,w,0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big) = \rho_t\big(0,w,0,\ldots,0;\delta_\theta \circledast Q_{t+1,T|\Theta}\big).$$

which proves that only the conditional measure $P_{t+1,t+1|\Theta}$ matters in this calculation. The fact that the knowledge of $w(\cdot,\theta)$ is sufficient, follows from the support property. This concludes the proof. □

In accordance with the above lemma we define the functions

$$\sigma_t : Z_1^{\mathcal{X}} \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}, \quad t = 1,\ldots,T-1,$$

as

$$\sigma_t(v;P_{t+1,t+1|\theta}) = \rho_t\big(0,w,0,\ldots,0;\delta_\theta \circledast P_{t+1,T|\Theta}\big), \tag{3.18}$$

where $w(\cdot,\theta) \equiv v(\cdot)$ and can be arbitrary otherwise. We refer to these functions as *transition risk mappings*.

Note that if ρ_t is normalized, monotonic, translation invariant and has support property, then so is σ_t.

Theorem 3.14 *A dynamic risk filter $\rho = \{\rho_t\}_{t=1,\ldots,T}$ is normalized, monotonic, translation invariant, has the support property, is parameter consistent, and time consistent, if and only if the following conditions are satisfied:*

1) *Marginal risk mappings $\widehat{\rho}_t : \mathcal{Z}_{t,t} \times \mathcal{P}(\widehat{\boldsymbol{\Theta}}) \to \mathbb{R}$, $t \in \mathcal{T}$, exist, which are normalized, monotonic, translation invariant, and have the support property;*
2) *Transition risk mappings given in (3.18) are such that*
   - (i) *For all $t = 1,\ldots,T-1$, $\sigma_t(\cdot;\cdot)$ is normalized, monotonic, translation invariant, and has the support property;*
   - (ii) *For any $P_{t+1,T} \in \mathcal{P}_{t+1,T}$, $t = 1,\ldots,T-1$, and for any functions $Z_{t,s} \in \mathcal{Z}_{t,s}$, $s \in \mathcal{T}_t$, we have that*[6]

$$\begin{aligned}&\rho_t(Z_{t,t},Z_{t,t+1},\ldots,Z_{t,T};P_{t+1,T})\\&\quad=\widehat{\rho}_t\Big(\Big\{Z_{t,t}(\theta)+\sigma_t\big(\rho_{t+1}\big(Z_{t,t+1}(\diamond,\cdot_1),\ldots,Z_{t,T}(\diamond,\cdot_{T-t-1},\cdot_1);\\&\delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|\diamond)\big);P_{t+1,t+1|\theta}\big),\theta\in\widehat{\boldsymbol{\Theta}}\Big\};P_{t+1,T,\Theta}\Big).\end{aligned} \tag{3.19}$$

[6] Recall Definition 3.11. The notation $\sigma_t\big(\rho_{t+1}\big(Z_{t,t+1}(\diamond,\theta),\ldots,Z_{t,T}(\diamond,\cdot_{T-t-1},\theta);\delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|\diamond)\big);P_{t+1,t+1|\theta}\big)$, where we use $\diamond$ as place holder, means that $P_{t+1,t+1|\theta}$ acts on $w(x,\theta) = \rho_{t+1}\big(Z_{t,t+1}(x,\theta),\ldots,f_{t,T}(x,\cdot_{T-t-1},\theta);\delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x)\big)$ as a function of x.

For any function $Z_{T,T} \in \mathcal{Z}_{T,T}$ *and any* $P_{T+1,T} \in \mathcal{P}(\mathbf{\Theta})$

$$\rho_T(Z_{T,T}; P_{T+1,T}) = \widehat{\rho}_T\Big(Z_{T,T}, P_{T+1,T}\Big). \tag{3.20}$$

Proof We fix $P_{t+1,T} \in \mathcal{P}_{t+1,T}$ and $Z_{t,s} \in \mathcal{Z}_{t,s}$, $s = t, \ldots, T$. Since t is fixed, to alleviate the notations we simply write Z_s, instead of $Z_{t,s}$ in this proof.

Since the risk filter is parameter consistent, Theorem 3.9 yields the existence of mappings $\widehat{\rho}_t$ such that

$$\begin{aligned}&\rho_t(Z_t, Z_{t+1}, \ldots, Z_T; P_{t+1,T})\\ &\quad= \widehat{\rho}_t\Big(\Big\{\rho_t\big(Z_t, Z_{t+1}, \ldots, Z_T; \delta_\theta \circledast P_{t+1,T|\Theta}\big), \theta \in \widehat{\mathbf{\Theta}}\Big\}; P_{t+1,T,\Theta}\Big).\end{aligned}$$

It follows from Proposition 3.10 that $\widehat{\rho}$ is normalized, monotonic, translation invariant, has the support property.

Next, we derive an equivalent expression for the first argument of $\widehat{\rho}_t$ that will prove (3.19). Define the function

$$w(x,\theta) = \rho_{t+1}\big(Z_{t+1}(x,\cdot_1), \ldots, Z_T(x, \cdot_{T-t-1}, \cdot_1); \delta_\theta \circledast P_{t+1,T|\widehat{\mathbf{\Theta}}}(\cdot\,\|x)\big), \quad x \in \mathcal{X},$$
$$\theta \in \widehat{\mathbf{\Theta}}.$$

Then, for any fixed $x \in \mathcal{X}$ and $\theta \in \mathbf{\Theta}$, we use the support, translation invariance, and the normalization properties in the chain of equations below:

$$\begin{aligned}&\rho_{t+1}\big(w(x,\cdot), 0, \ldots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x)\big)\\ &\quad= \rho_{t+1}\big(w(x,\theta), 0, \ldots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x)\big)\\ &\quad= w(x,\theta) + \rho_{t+1}\big(0, 0, \ldots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x)\big) = w(x,\theta)\\ &\quad= \rho_{t+1}\big(Z_{t+1}(x,\cdot_1), \ldots, Z_T(x,\cdot_{T-t-1},\cdot_1); \delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\,\|x)\big).\end{aligned}$$

In view of the assumed time consistency of ρ, the above implies that for every $Z_{t,t} \in \mathcal{Z}_{t,t}$,

$$\rho_t\big(Z_t, Z_{t+1}, \ldots, Z_T; \delta_\theta \circledast P_{t+1,T|\Theta}\big) = \rho_t\big(Z_t, w, 0, \ldots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}\big) =: I_1.$$

Thus, by using the translation invariance and the support properties again, we conclude that, for all $\theta \in \widehat{\mathbf{\Theta}}$,

$$\begin{aligned}I_1 &= \rho_t\big(Z_t(\theta), w, 0, \ldots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}\big)\\ &= Z_t(\theta) + \rho_t\big(0, w, 0, \ldots, 0; \delta_\theta \circledast P_{t+1,T|\Theta}\big).\end{aligned}$$

Using (3.18), we get

$$I_1 = Z_t(\theta) + \sigma_t\big(w(\cdot,\theta), P_{t+1,t+1|\theta}\big).$$

Finally, from here, note that by support property

$$w(x,\theta) = \rho_{t+1}\big(Z_{t+1}(x,\theta),\dots,Z_T(x,\cdot_{T-t-1},\theta);\delta_\theta \circledast P_{t+1,T|\Theta}(\cdot\|x)\big),\quad x\in\mathcal{X},\quad \theta\in\widehat{\mathbf{\Theta}}, \tag{3.21}$$

we obtain the representation (3.19). The representation (3.20) follows from the definition of ρ_T and the form of $\widehat{\rho}_T$.

Next, we prove the converse statement by backward induction in time. For $t = T$, the conditional risk filter (3.20) has all the postulated properties, with the exception of the time consistency, because $\widehat{\rho}_T$ does (see Proposition 3.10).

Suppose the conditional risk filters ρ_s, $s = t+1,\dots,T$ are normalized, monotonic, translation invariant, have the support property, are parameter consistent, and time consistent. We will verify these properties for ρ_t given by formula (3.19). The translation invariance follows from the translation invariance of $\widehat{\rho}_t$. The normalization and the monotonicity follow immediately from the normalization and the monotonicity of σ_t, ρ_{t+1} and $\widehat{\rho}_t$.

We now verify the support property. For every θ, and x define $\mu_{\theta,x}(\cdot) = \delta_\theta \circledast P_{t+1,T|\theta}(\cdot\|x)$, $A(\theta) = \operatorname{supp}(P_{t+1,t+1|\theta}) \subset \mathcal{X}$, $B = \operatorname{supp}(P_{t+1,T,\Theta}) \subset \mathbf{\Theta}$. Then, by (3.19), (3.21), the support property of $\hat{\rho}_t$ and σ_t, and by Remark 3.3 applied to σ_t, we deduce that

$$\begin{aligned}
&\rho_t(Z_t, Z_{t+1},\dots,Z_T; P_{t+1,T}) = \hat{\rho}_t\Big(\Big\{Z_t(\theta) + \sigma_t(w(\diamond,\theta); P_{t+1,t+1,|\theta}); \theta\in\mathbf{\Theta}\Big\}; P_{t+1,T,\Theta}\Big)\\
&= \hat{\rho}_t\Big(\Big\{\mathbb{1}_B(\theta)Z_t(\theta) + \mathbb{1}_B(\theta)\sigma_t(w(\diamond,\theta); P_{t+1,t+1,|\theta}); \theta\in\mathbf{\Theta}\Big\}; P_{t+1,T,\Theta}\Big)\\
&= \hat{\rho}_t\Big(\Big\{\mathbb{1}_B(\theta)Z_t(\theta) + \sigma_t(\mathbb{1}_B(\theta)\mathbb{1}_{A(\theta)}(\diamond)w(\diamond,\theta); P_{t+1,t+1,|\theta}); \theta\in\mathbf{\Theta}\Big\}; P_{t+1,T,\Theta}\Big).
\end{aligned} \tag{3.22}$$

By the assumed support property of ρ_{t+1}, and in view of Remark 3.3 applied to ρ_{t+1}, we obtain, using (3.21) again,

$$\begin{aligned}
\mathbb{1}_{A(\theta)}(x)\mathbb{1}_B(\theta)w(x,\theta) = \rho_{t+1}\big(&\mathbb{1}_{A(\theta)}(x)\mathbb{1}_B(\theta)Z_{t+1}(x,\theta),\\
&\mathbb{1}_{A(\theta)}(x)\mathbb{1}_B(\theta)\mathbb{1}_{\operatorname{supp}_{t+2}(\mu_{\theta,x})}Z_{t+2}(x,\cdot,\theta),\dots,\\
&\mathbb{1}_{A(\theta)}(x)\mathbb{1}_B(\theta)\mathbb{1}_{\operatorname{supp}_T(\mu_{\theta,x})}Z_T(x,\cdot_{T-t-1},\theta);\mu_{\theta,x}\big),
\end{aligned}$$

for every $x\in\mathcal{X}$ and $\theta\in\mathbf{\Theta}$. From here and (3.22), combined with the normalization property of ρ_{t+1}, and the fact that $\mathbb{1}_{A(\theta)}(x)\mathbb{1}_B(\theta)\mathbb{1}_{\operatorname{supp}_s(\mu_{\theta,x})} \le 1_{\operatorname{supp}_s(P_{t+1,T})}$, $s = t,\dots,T$, we obtain the support property of ρ_t.

Next we prove the parameter consistency. Assume that (3.7) is satisfied for a fixed $\bar{\theta}\in\mathbf{\Theta}$, and denote by $\bar{P}_{t+1,T} = \delta_{\bar{\theta}} \circledast P_{t+1,T|\Theta}$ and $\bar{Q}_{t+1,T} = \delta_{\bar{\theta}} \circledast Q_{t+1,T|\Theta}$. We note that[7]

$$\begin{aligned}
\bar{P}_{t+1,t+1|\theta} &= P_{t+1,t+1|\bar{\theta}}\cdot\mathbb{1}_{\bar{\theta}}(\theta), \quad \bar{P}_{t+1,T,\Theta} = \delta_{\bar{\theta}}, \quad \bar{P}_{t+1,T|\theta}(\cdot\|x)\\
&= P_{t+1,T|\bar{\theta}}(\cdot\|x)\mathbb{1}_{\bar{\theta}}(\theta).
\end{aligned}$$

[7] We use the convention that $\frac{0}{0} = 0$ when considering $\bar{P}_{t+1,t+1|\theta}$ and $\bar{P}^x_{t+1,T|\theta}(\cdot\|x)$.

Using this, and in view of (3.19), we can write (3.7) as follows (with measures $\bar{P}_{t+1,T}$ and $\bar{Q}_{t+1,T}$ in place of $P_{t+1,T}$ and $Q_{t+1,T}$):

$$\begin{aligned}
&\widehat{\rho}_t\Big(\Big\{Z_t(\theta)+\sigma_t\big(\rho_{t+1}\big(Z_{t+1}(\diamond,\theta),\dots,Z_T(\diamond,\cdot_{T-t-1},\theta);\\
&\delta_{\bar{\theta}}(\cdot)P_{t+1,T|\bar{\theta}}(\cdot\,\|\diamond)\mathbb{1}_{\bar{\theta}}(\theta);\,P_{t+1,t+1|\bar{\theta}}\mathbb{1}_{\bar{\theta}}(\theta)\big),\theta\in\widehat{\boldsymbol{\Theta}}\Big\};\delta_{\bar{\theta}}\Big)\\
&\quad\le\widehat{\rho}_t\Big(\Big\{W_t(\theta)+\sigma_t\big(\rho_{t+1}\big(W_{t+1}(\diamond,\theta),\dots,W_T(\diamond,\cdot_{T-t-1},\theta);\\
&\delta_{\bar{\theta}}(\cdot)Q_{t+1,T|\bar{\theta}}(\cdot\,\|\diamond)\mathbb{1}_{\bar{\theta}}(\theta);\,Q_{t+1,t+1|\bar{\theta}}\mathbb{1}_{\bar{\theta}}(\theta)\big),\theta\in\widehat{\boldsymbol{\Theta}}\Big\};\delta_{\bar{\theta}}\Big).
\end{aligned}$$

By the support property of $\hat{\rho}_t$ and σ_t, and by the normalization and monotonicity of $\hat{\rho}_t$, we obtain that

$$\begin{aligned}
&Z_t(\bar{\theta})+\sigma_t\big(\rho_{t+1}\big(Z_{t+1}(\diamond,\bar{\theta}),\dots,Z_T(\diamond,\cdot_{T-t-1},\bar{\theta});\delta_{\bar{\theta}}(\cdot)P_{t+1,T|\bar{\theta}}(\cdot\,\|\diamond);\,P_{t+1,t+1|\bar{\theta}}\big)\\
&\quad\le W_t(\bar{\theta})+\sigma_t\big(\rho_{t+1}\big(W_{t+1}(\diamond,\bar{\theta}),\dots,W_T(\diamond,\cdot_{T-t-1},\bar{\theta});\delta_{\bar{\theta}}(\cdot)Q_{t+1,T|\bar{\theta}}(\cdot\,\|\diamond);\,Q_{t+1,t+1|\bar{\theta}}\big),
\end{aligned}$$

for any $\bar{\theta}\in\boldsymbol{\Theta}$. From here, applying $\hat{\rho}_t$ to both sides, since we assumed that $P_{t+1,T,\widehat{\boldsymbol{\Theta}}}=Q_{t+1,T,\widehat{\boldsymbol{\Theta}}}$, employing monotonicity of $\hat{\rho}_t$, we obtain (3.8), and the parameter consistency of ρ_t is proved.

Finally, let us verify the time consistency at time t. If the inequalities (3.15) are satisfied, then it follows from the monotonicity of σ_t with respect to its first argument that for all $\theta\in\widehat{\boldsymbol{\Theta}}$

$$\begin{aligned}
&G_1(\theta):=\sigma_t\big(\rho_{t+1}\big(Z_{t+1}(\diamond,\theta),\dots,Z_T(\diamond,\cdot_{T-t-1},\theta);\delta_\theta\circledast P_{t+1,T|\Theta}(\cdot\,\|\diamond)\big);\,P_{t+1,t+1|\theta}\big)\\
&\quad\le\sigma_t\big(\rho_{t+1}\big(W_{t+1}(\diamond,\theta),\dots,W_T(\diamond,\cdot_{T-t-1},\theta);\delta_\theta\circledast Q_{t+1,T|\Theta}(\cdot\,\|\diamond)\big);\,P_{t+1,t+1|\theta}\big)=:G_2(\theta).
\end{aligned}$$

Then, from the monotonicity of $\widehat{\rho}_t$ we get, for any function $f_t\in\mathcal{Z}_{t,t}$,

$$\hat{\rho}_t(f_t+G_1;\delta_\theta)\le\hat{\rho}_t(f_t+G_2;\delta_\theta),\quad\theta\in\boldsymbol{\Theta}.$$

From here, using the support property of $\widehat{\rho}_t$, σ_t and ρ_{t+1}, along (3.19), we obtain (3.16), and thus time consistency at t is verified.

By induction, all properties hold true for $t=1,\dots,T$. □

Remark 3.15 Note that even though we have proved the existence of the conditional risk mappings $\widehat{\rho}_t$, the specific form of these mappings depends on the given dynamic risk filter.

Example 3.16 We consider a very special conditional risk filter, given as the expectation of an additive functional under the measure $P_{t+1,T}$. Specifically, we let

$$\begin{aligned}
\rho_t(Z_{t,t},\dots,Z_{t,T};P_{t+1,T})&=\int_{\mathcal{X}^{T-t}\times\widehat{\boldsymbol{\Theta}}}\sum_{k=t}^{T}Z_{t,k}(x_{t+1},\dots,x_k,\theta)P_{t+1,T}(dx_{t+1},\cdots,dx_T,d\theta)\\
&=\mathbb{E}_{P_{t+1,T}}\sum_{k=t}^{T}Z_{t,k}.
\end{aligned}$$

Clearly, this ρ_t is normalized, monotonic, translation invariant, and has the support property (cf. Definition 3.2).

Next, note that for this ρ_t the inequality (3.7) becomes (cf. (3.4))

$$\int_{\mathcal{X}^{T-t}} \sum_{k=t}^{T} Z_{t,k}(x_{t+1:k}, \theta) P_{t+1,T|\theta}(dx_{t+1}, \cdots, dx_T)$$
$$\leq \int_{\mathcal{X}^{T-t}} \sum_{k=t}^{T} Z_{t,k}(x_{t+1:k}, \theta) Q_{t+1,T|\theta}(dx_{t+1}, \cdots, dx_T),$$

for any $\theta \in \mathbf{\Theta}$. Assuming that $P_{t+1,T,\Theta} = Q_{t+1,T,\Theta}$, multiplying the last inequality by $P_{t+1,T,\Theta}(\theta)$, and summing up with respect to $\theta \in \mathbf{\Theta}$, the inequality (3.8) follows, and hence the parameter consistency is true.

The time consistency follows by similar arguments. Indeed, (3.15) becomes (cf. (3.13))

$$\int_{\mathcal{X}^{T-t-1}} \sum_{k=t+1}^{T} Z_{t,k}(x_{t+1}, x_{t+2:k}, \theta)\ \widetilde{P}_{t+1,T|\theta}(dx_{t+2}, \cdots, dx_T \| x_{t+1})$$
$$\leq \int_{\mathcal{X}^{T-t-1}} \sum_{k=t+1}^{T} Z_{t,k}(x_{t+1}, x_{t+2:k}, \theta)\ \widetilde{Q}_{t+1,T|\theta}(dx_{t+2}, \cdots, dx_T \| x_{t+1}),$$

for any $x_{t+1} \in \mathcal{X}$ and $\theta \in \mathbf{\Theta}$. Assuming that $P_{t+1,t+1|\Theta} = Q_{t+1,t+1|\Theta}$, multiplying both parts by $P_{t+1,t+1|\theta}(x_{t+1})$, and noting that (cf. (3.13))

$$\widetilde{P}_{t+1,T|\theta}(\cdot \| x_{t+1}) P_{t+1,t+1|\theta}(x_{t+1}) = P_{t+1,T|\theta}(x_{t+1}, \cdot),$$

for any function $f_t \in \mathcal{Z}_{t,t}$ we have (cf. (3.13))

$$f_t(\theta) + \int_{\mathcal{X}^{T-t-1}} \sum_{k=t+1}^{T} Z_{t,k}(x_{t+1}, x_{t+2:k}, \theta)\ P_{t+1,T|\theta,}(\{x_{t+1}\}, dx_{t+2}, \cdots, dx_T)$$
$$\leq f_t(\theta) + \int_{\mathcal{X}^{T-t-1}} \sum_{k=t+1}^{T} Z_{t,k}(x_{t+1}, x_{t+2:k}, \theta)\ Q_{t+1,T|\theta}(\{x_{t+1}\}, dx_{t+2}, \cdots, dx_T).$$

After summing up with respect to x_{t+1} we obtain (3.16), and thus the time consistency is proved.

We complete this example by observing that in the this case we have that $\widehat{\rho}_t(f, P') = \mathbb{E}_{P'}(f)$, for $f \in Z_{t,t}$ and $P' \in \mathcal{P}(\mathbf{\Theta})$, and that $\sigma_t(v; P'') = \mathbb{E}_{P''}(v)$, for $v \in \mathcal{Z}_1^{\mathcal{X}}$ and $P'' \in \mathcal{P}(\mathcal{X})$. □

Example 3.17 Let us cast Example 3.16 in the setup of Sect. 2. For this, we fix a history $h_t = (x_1, \ldots, x_t) \in \mathcal{H}_t$ and $\pi \in \Pi$, and we take

$$\begin{aligned} Z_{t,t}(\theta) &:= Z^{\pi,h_t}_{\theta,t,t} = c_t(x_t, \pi_t(h_t), \theta), \\ Z_{t,s}(x_{t+1}, \ldots, x_s, \theta) &:= Z^{\pi,h_t,x_{t+1},\ldots,x_s}_{\theta,t,s} \\ &= c_s(x_s, \pi^{t,h_t}_s(x_{t+1}, \ldots, x_s), \theta), \ s = t+1, \ldots, T, \end{aligned}$$

and

$$P_{t+1,T} = P^{\pi^{t,h_t}}_{t+1,T}.$$

The conditional risk filter of Example 3.16 becomes the conditional expectation (cf. Lemma 2.3)

$$\begin{aligned} &\rho_t\Big(c_t(x_t, \pi_t(h_t), \cdot), c_{t+1}(\cdot, \pi_{t+1}(h_t, \cdot), \cdot), \cdots, c_T(\cdot, \pi_T(h_t, \cdot, \ldots, \cdot), \cdot), P^{\pi^{t,h_t}}_{t+1,T}\Big) \\ &\quad = \mathbb{E}^{\pi}\left[c_t(x_t, \pi_t(h_t), \Theta) + \sum_{s=t+1}^{T} c_s(\widehat{X}_s, \pi^{t,h_t}_s(\widehat{X}_{t+1}, \ldots, \widehat{X}_s), \Theta) \mid \widehat{H}_t = h_t\right], \end{aligned} \tag{3.23}$$

for $t = 1, \ldots, T$, where we use the standard convention that an empty sum is zero (i.e. $\sum_{s=T+1}^{T} \cdots = 0$ in our case).

In view of (2.7) we also have

$$\begin{aligned} &\rho_t\Big(c_t(x_t, \pi_t(h_t), \cdot), c_{t+1}(\cdot, \pi_{t+1}(h_t, \cdot), \cdot), \cdots, c_T(\cdot, \pi_T(h_t, \cdot, \ldots, \cdot), \cdot), P^{\pi^{t,h_t}}_{t+1,T}\Big) \\ &\quad = \widehat{\rho}_t\Big(\Big\{c_t(x_t, \pi_t(h_t), \theta) + \sigma_t\Big(\rho_{t+1}\Big(c_{t+1}(\diamond, \pi_{t+1}(h_t, \diamond), \cdot), c_{t+2}(\cdot, \pi_{t+2}(h_t, \diamond, \cdot), \cdot), \\ &\qquad \cdots, c_T(\cdot, \pi_T(h_t, \diamond, \cdot, \ldots, \cdot), \cdot); \delta_\theta \circledast P^{\pi^{t,h_t}}_{t+1,T|\Theta}(\cdot \,\|\diamond)\Big); P^{\pi^{t,h_t}}_{t+1,t+1|\theta}\Big), \theta \in \widehat{\mathbf{\Theta}}\Big\}; \xi^{\pi,h_t}_t\Big) \\ &\quad = \widehat{\rho}_t\Big(\Big\{c_t(x_t, \pi_t(h_t), \theta) + \sigma_t\Big(\rho_{t+1}\Big(c_{t+1}(\diamond, \pi_{t+1}(h_t, \diamond), \cdot), c_{t+2}(\cdot, \pi_{t+2}(h_t, \diamond, \cdot), \cdot), \\ &\qquad \cdots, c_T(\cdot, \pi_T(h_t, \diamond, \cdot, \ldots, \cdot), \cdot); P^{\pi^{t,h_t}}_{\theta,t+1,T}(\{\diamond\} \times \cdot)\delta_\theta(\cdot)\Big); P^{\pi^{t,h_t}}_{\theta,t+1}\Big), \theta \in \widehat{\mathbf{\Theta}}\Big\}; \xi^{\pi,h_t}_t\Big), \end{aligned} \tag{3.24}$$

where in the last equality we used (3.5) and (3.14), where

$$\widehat{\rho}_t\Big(\big\{f(\theta), \theta \in \widehat{\mathbf{\Theta}}\big\}; \xi^{\pi,h_t}_t\Big) = \widehat{\rho}_t\big(f; \xi^{\pi,h_t}_t\big) = \int_{\widehat{\mathbf{\Theta}}} f(\theta)\, \xi^{\pi,h_t}_t(d\theta) = \mathbb{E}_{\xi^{\pi,h_t}_t}(f), \tag{3.25}$$

and where, for a function v on $\mathcal{X}$, we have (cf. (3.5), (3.18), and (3.6))

$$\sigma_t\left(v, P^{\pi^{t,h_t}}_{\theta,t+1}\right) = \int_{\mathcal{X}} v(x) P^{\pi^{t,h_t}}_{\theta,t+1} d(x) = \mathbb{E}_{P^{\pi^{t,h_t}}_{\theta,t+1}}(v). \tag{3.26}$$

Example 3.18 In the previous example we proceeded from ρ to σ (via $\widehat{\rho}$). Here, we will do the opposite.

In clinical trials, the potency of a drug is characterized by an unknown parameter θ. The purpose of the trials is to estimate θ and to determine the optimal dose. Let us assume for simplicity that θ *is* the optimal dose. If a dose u_1 is administered to a patient, a response X_2 is observed (the subscript 2 indicates that X_2 is not known when u_1 is determined). X_2 is a Bernoulli random variable, with $X_2 = 1$ representing toxic response, and $X_2 = 0$ nontoxic. The probability of toxic response is a function of θ and u_1, that is, $P[X_2 = 1] = \Psi(\theta, u_1)$. The "cost" is $c(\theta, u_1)$; it depends on both the applied and best doses. The cost is not observed; we only know whether the patient was toxic or not. In the second stage, the dose u_2 is administered to the next patient, the patient's response X_3 observed, and cost $c(\theta, u_2)$ incurred. The process continues for T stages, with u_T being the final dose recommendation, whose cost is equal to $c(\theta, u_T)$. For example, the cost may have the form $c(\theta, u) = |u - \theta|$ to penalize for the over- and under-dosage. It is never observed.

The problem can be cast to our setting. The state space $\mathcal{X}$ is $\{0, 1\}$, while the unknown parameter space $\widehat{\boldsymbol{\Theta}}$ is an interval of the real line or a finite subset of the real line. Given the set-up adopted in this paper, we assume that $\widehat{\boldsymbol{\Theta}}$ is a finite subset of the real line. The transition kernel does not depend on X at all; the distribution of the next X_{t+1} depends on θ and u:

$$K_\theta(0|x, u) = 1 - \Psi(\theta, u), \quad K_\theta(1|x, u) = \Psi(\theta, u).$$

Thus, we have

$$\begin{aligned} P^{\pi^{t,h_t}}_{\theta,t+1}(y) &= K_\theta(y|x_t, \pi_t(h_t)) \\ &= \mathbb{1}_{\{y=0\}}\big(1 - \Psi(\theta, \pi_t(h_t))\big) + \mathbb{1}_{\{y=1\}}\Psi(\theta, \pi_t(h_t)), \quad y \in \{0, 1\}. \end{aligned}$$

There is considerable freedom to choose the form of σ_t in a way consistent with our set-up. For example, one can choose σ_t by generalizing the entropic risk measure to allow for dependence on the probability measure:

$$\sigma_t(v; P) = \frac{1}{\varkappa} \ln \int_{\mathcal{X}} e^{\varkappa v(y)} P(dy),$$

for a function v on $\mathcal{X}$, $P \in \mathcal{P}(\mathcal{X})$, and a constant $\varkappa > 0$. Consequently, for $t = 1, \ldots, T-1$, using (2.5) and (2.10) we obtain

$$\begin{aligned} \sigma_t(w(\cdot, \theta); P^{\pi^{t,h_t}}_{\theta,t+1}) &= \frac{1}{\varkappa} \ln \int_{\mathcal{X}} e^{\varkappa w(y,\theta)} P^{\pi^{t,h_t}}_{\theta,t+1}(dy) \\ &= \frac{1}{\varkappa} \ln \Big(\big(1 - \Psi(\theta, \pi_t(h_t))\big) e^{\varkappa w(0,\theta)} + \Psi(\theta, \pi_t(h_t)) e^{\varkappa w(1,\theta)} \Big), \end{aligned}$$

with $\sigma_T = 0$.

Now, for a function f on $\widehat{\mathbf{\Theta}}$ and a measure $\xi \in \mathcal{P}(\widehat{\mathbf{\Theta}})$, let

$$\widehat{\rho}_t\Big(\big\{f(\theta), \theta \in \widehat{\mathbf{\Theta}}\big\}; \xi\Big) = \widehat{\rho}_t(f; \xi) = \frac{1}{\varkappa} \ln \int_{\widehat{\mathbf{\Theta}}} e^{\varkappa f(\theta)} \xi(d\theta), \qquad t \in \mathcal{T}.$$

Given the above, we obtain for $t = T$

$$\widehat{\rho}_T\Big(\big\{f(\theta), \theta \in \widehat{\mathbf{\Theta}}\big\}; \xi_T^{\pi,h_T}\Big) = \frac{1}{\varkappa} \ln \int_{\widehat{\mathbf{\Theta}}} e^{\varkappa f(\theta)} \xi_T^{\pi,h_T}(d\theta), \tag{3.27}$$

and for $t = 1, \ldots, T-1$,

$$\begin{aligned}
&\widehat{\rho}_t\Big(\big\{f(\theta) + \sigma_t(w(\cdot,\theta); P^{\pi,h_t}_{\theta,t+1}), \theta \in \widehat{\mathbf{\Theta}}\big\}; \xi_t^{\pi,h_t}\Big) \\
&= \frac{1}{\varkappa} \ln \int_{\widehat{\mathbf{\Theta}}} \int_{\mathcal{X}} e^{\varkappa(f(\theta)+w(y,\theta))} P^{\pi,h_t}_{\theta,t+1}(dy) \xi_t^{\pi,h_t}(d\theta) \\
&= \frac{1}{\varkappa} \ln \mathbb{E}^{\pi}[e^{\varkappa(f(\Theta)+w(X_{t+1},\Theta))} | \widehat{H}_t = h_t],
\end{aligned}$$

where the last equality follows from Lemma 2.5.

We will now derive a generic formula for ρ_t, resulting from (3.19) and σ_t and $\widehat{\rho}_t$ as above, in case of the generic cost functions, as in (2.16) and (2.17). Let us fix an admissible strategy π. For $t = T$ we have

$$\begin{aligned}
&\rho_T(c_T(x_T, \pi_T(h_T), \cdot), P^{\pi,h_T}_{T+1,T}) = \widehat{\rho}_T(\{c_T(x_T, \pi_T(h_T), \theta), \theta \in \widehat{\mathbf{\Theta}}\}; P^{\pi,h_T}_{T+1,T,\Theta}) \\
&= \widehat{\rho}_T(c_T(x_T, \pi_T(h_T), \cdot); P^{\pi,h_T}_{T+1,T,\Theta}) = \widehat{\rho}_T(c_T(x_T, \pi_T(h_T), \cdot); \xi_T^{\pi,h_T}) \\
&= \frac{1}{\varkappa} \ln \int_{\widehat{\mathbf{\Theta}}} e^{\varkappa c_T(x_T, \pi_T(h_T), \theta)} \xi_T^{\pi,h_T}(d\theta) = \frac{1}{\varkappa} \ln \mathbb{E}^{\pi}(e^{\varkappa c_T(x_T, \pi_T(h_T), \Theta)} | \widehat{H}_T = h_T).
\end{aligned}$$

Now, note that

$$\rho_T(c_T(x_T, \pi_T(h_T), \cdot), \delta_\theta) = c_T(x_T, \pi_T(h_T), \theta),$$

and thus

$$\begin{aligned}
&\sigma_{T-1}\big(\rho_T\big(c_T(\diamond, \pi_T(h_{T-1}, \diamond), \theta); \delta_\theta\big); P^{\pi^{T-1,h_{T-1}}}_{T,T|\theta}\big) \\
&= \sigma_{T-1}\big(c_T(\diamond, \pi_T(h_{T-1}, \diamond), \theta); P^{\pi^{T-1,h_{T-1}}}_{T,T|\theta}\big) \\
&= \frac{1}{\varkappa} \ln \int_{\mathcal{X}} e^{\varkappa c_T(x_T, \pi_T(h_{T-1}, x_T), \theta)} P^{\pi^{T-1,h_{T-1}}}_{\theta,T}(dx_T).
\end{aligned}$$

So, for $t = T-1$, we have

$$\begin{aligned}
&\rho_{T-1}(c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \cdot), c_T(\cdot, \pi_T(h_{T-1}, \cdot), \cdot), P_{T,T}^{\pi\, T-1, h_{T-1}}) \\
&= \widehat{\rho}_{T-1}\Big(\Big\{c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \theta) + \sigma_{T-1}\big(\rho_T\big(c_T(\diamond, \pi_T(h_{T-1}, \diamond), \cdot); \delta_\theta\big); P_{T,T|\theta}^{\pi\, T-1, h_{T-1}}\big)\big), \\
&\quad \theta \in \widehat{\Theta}\Big\}; P_{T,T,\Theta}^{\pi\, T-1, h_{T-1}}\Big) \\
&= \widehat{\rho}_{T-1}\Big(\Big\{c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \theta) + \sigma_{T-1}\big(\rho_T\big(c_T(\diamond, \pi_T(h_{T-1}, \diamond), \cdot); \delta_\theta\big); \\
&\quad P_{T,T|\theta}^{\pi\, T-1, h_{T-1}}\big)\big), \theta \in \widehat{\Theta}\Big\}; \xi_{T-1}^{\pi, h_{T-1}}\Big) \\
&= \frac{1}{\varkappa} \ln \int_{\widehat{\Theta}} \int_{\mathcal{X}} e^{\varkappa(c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \theta) + c_T(x_T, \pi_T(h_{T-1}, x_T), \theta))} P_{\theta,T}^{\pi\, T-1, h_{T-1}}(dx_T) \xi_{T-1}^{\pi, h_{T-1}}(d\theta) \\
&= \frac{1}{\varkappa} \ln \mathbb{E}^\pi\big(e^{\varkappa(c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \Theta) + c_T(X_T, \pi_T(h_{T-1}, X_T), \Theta))} | \widehat{H}_{T-1} = h_{T-1}\big),
\end{aligned}$$

where we used (2.5) and (2.10) for the second to the last equality, and where the last equality follows from Lemma 2.5.

Proceeding in the analogous way for $t = T-2, \ldots, 1$ we finally obtain

$$\begin{aligned}
&\rho_{1,T}(c_1(x_1, \pi_1(h_1), \cdot), \ldots, c_T(\cdot, \pi_T(h_1, \cdot, \ldots, \cdot), \cdot), P_{2,T}^{\pi\, 1, h_1}) \\
&\qquad = \frac{1}{\varkappa} \ln \mathbb{E}^\pi\left(e^{\varkappa \sum_{k=1}^T c_k(X_k, \pi_k(h_k)), \Theta)} \,\middle|\, \widehat{H}_1 = h_1\right), \qquad (3.28)
\end{aligned}$$

which gives us the risk-sensitive criterion with the entropic utility (cf. Bäuerle and Rieder (Feb 2014); Davis and Lleo (2014)).

Example 3.19 A prominent example that can be cast in our framework is the classical optimal investment and consumption problem subject to model uncertainty. Namely, consider an investor with initial capital $\bar{x}_1$, who can invest in d assets, with $\bar{X}_t$ denoting the portfolio value at time t, which of course is observed by the investor. The investor rebalances the portfolio at each time t, following a self-financing trading strategy (policy) $\bar{\pi}$, that may satisfy additional trading constrains, such as short selling constraints, turnover constraints, etc. The investor is also allowed to consume at each time t part of the wealth, say z_t, that does not exceed $\bar{X}_t$. We postulate that the investor maximizes the expected utility of consumptions and terminal wealth using the utility functions V^β and U^γ, respectively, where $\beta, \gamma \in \mathbb{R}$ stand for risk aversion parameters. We refer the reader to [Bäuerle and Rieder (2017),Section 4] for detailed formulation of this problem in the MDP framework.

Additionally, we assume that the investor faces the Knightian uncertainty about the model of the underlying assets, described in terms of a (finite) parametric set $\mathbf{\Lambda} \subset \mathbb{R}^m$; see (Bielecki et al., 2019) for an overview of MDPs under Knightian uncertainty. Let $\{P_\lambda\}_{\lambda \in \Lambda}$ be a parameterized family of probability measures, and assume that the wealth process $\bar{X}_t$, follows the dynamics

$$X_{t+1} = G(X_t, \bar{\pi}_t, Z_{t+1}), \quad t = 1, \ldots, T-1,$$

where Z_{t+1} is the random disturbance (e.g. the log-returns of the underlying investment instruments), and G a deterministic function.

Moreover, we suppose that the investor is also uncertain about her risk aversion parameters $(\beta, \gamma) \in \mathbf{\Gamma} \subset \mathbb{R}^2$. We emphasize that this additional feature of an unknown risk aversion parameter is practically important. Generally speaking, it is difficult to determine the investor's risk aversion parameter, which is well documented in the behavioral finance literature. This becomes especially relevant in the context of fast-growing robo-advising industry that typically deals with unsophisticated investors, and which establishes investor's risk preferences without human intervention. At each time t, the investor reports through process Y_t her subjective degree of happiness about the performance of her investment. For example, one can take Y_t to be a Bernoulli random variable with $Y_t = 1$ corresponding to happy and $Y_t = 0$ meaning unhappy about her investment, and then we follow a similar setup to the clinical trials Example 3.6 and incorporate the uncertainty about (β, γ) into the original MDP formulation.

Namely, we consider the observed state process $X_t = (\bar{X}_t, Y_t)$, and we take $\theta = (\beta, \gamma, \lambda) \in \mathbf{\Theta} = \mathbf{\Gamma} \times \mathbf{\Lambda}$ representing the model uncertainty in this model. We assume that choosing the 'optimal' risk aversion parameters $\check{\pi}_t = (v_t^1, v_t^2)$ is part of the policy, and that the robo-advisor/investor is making this choice at each time step. Overall, the policy at time t becomes $\pi_t = (\bar{\pi}_t, \check{\pi}_t)$. With this at hand we define the transition kernels

$$K_\theta(\bar{x}_{t+1}, y_{t+1} \mid x_t, y_t, \pi_t(h_t)) = P_\theta(G(\bar{x}_t, \bar{\pi}_t(h_t), Z_{t+1}) = \bar{x}_{t+1})\Psi(\beta, \gamma, \check{\pi}_t(h_t); y_{t+1}),$$

for some function Ψ.

Consequently, we define the cost functionals

$$\begin{aligned} c_t &= V^\beta(z_t(\bar{x}_t, \bar{\pi}_t)) + F(y_t, \check{\pi}_t, \beta, \gamma), \quad t = 1, \dots, T-1, \\ c_T &= V^\beta(z_T(\bar{x}_T, \bar{\pi}_T)) + U^\gamma(\bar{x}_T), \end{aligned}$$

where F is a penalty for 'deviating' from the true risk-aversion parameters. Using the expectation as the risk functional (cf. Example 3.16), we obtain a generalization of the classical expected utility criteria in the context of optimal investment.

4 Recursive risk filters

Let us fix $t \in \{1, \dots, T-1\}$. Since t is fixed, we will again simply write Z_s instead of $Z_{t,s}$, for $s \in \mathcal{T}_t$. We introduce two families of functions:

$$\begin{aligned} v_t^\pi(h_t) &= \rho_t(Z_t, Z_{t+1}, \dots, Z_T; P_{t+1,T}^{\pi^t, h_t}) \\ \widetilde{v}_{t+1}^{\pi,\theta}((h_t, x_{t+1})) &:= \rho_{t+1}\big(Z_{t+1}(x_{t+1}, \cdot_1), \dots, Z_T(x_{t+1}, \cdot_{T-t-1}, \cdot_1); \delta_\theta \circledast P_{t+1,T|\Theta}^{\pi^t, h_t}(\cdot \,\|\, x_{t+1})\big) \\ &= \rho_{t+1}\big(Z_{t+1}(x_{t+1}, \cdot_1), \dots, Z_T(x_{t+1}, \cdot_{T-t-1}, \cdot_1); P_{\theta,t+1,T}^{\pi^t, h_t}(\{x_{t+1}\} \times \cdot)\delta_\theta(\cdot)\big), \end{aligned} \tag{4.1}$$

where for the last equality we used (3.14).

The quantity $v_t^\pi(h_t)$ evaluates the policy π at the time t and with the history h_t in the original problem.

Recall that (cf. (3.5)) $P^{\pi^{t,h_t}}_{t+1,T,\Theta} = \xi_t^{\pi,h_t}$. Thus, the key equation (3.19) can be written more compactly as follows:

$$\begin{aligned}
v_t^\pi(h_t) &= \widehat{\rho}_t\Big(\Big\{Z_t(\theta) + \sigma_t\big(\rho_{t+1}\big(Z_{t+1}(\diamond,\cdot_1),\ldots,Z_T(\diamond,\cdot_{T-t-1},\cdot_1);\\
&\qquad\qquad \delta_\theta \circledast P^{\pi^{t,h_t}}_{t+1,T|\Theta}(\cdot\,\|\diamond)\big);\, P^{\pi^{t,h_t}}_{t+1,t+1|\theta}\big),\, \theta\in\widehat{\mathbf{\Theta}}\Big\};\, P^{\pi^{t,h_t}}_{t+1,T,\Theta}\Big)\\
&= \widehat{\rho}_t\Big(\Big\{Z_t(\theta) + \sigma_t\big(Z_{t+1}(\diamond,\cdot_1),\ldots,Z_T(\diamond,\cdot_{T-t-1},\cdot_1);\\
&\qquad\qquad \delta_\theta \circledast P^{\pi^{t,h_t}}_{t+1,T|\Theta}(\cdot\,\|\diamond)\big);\, P^{\pi^{t,h_t}}_{t+1,t+1|\theta}\big),\, \theta\in\widehat{\mathbf{\Theta}}\Big\};\, \xi_t^{\pi,h_t}\Big)\\
&= \widehat{\rho}_t\Big(\Big\{Z_t(\theta) + \sigma_t\big(\widetilde{v}^{\pi,\theta}_{t+1}((h_t,\diamond));\, P^{\pi^{t,h_t}}_{t+1,t+1|\theta}\big),\, \theta\in\widehat{\mathbf{\Theta}}\Big\};\, \xi_t^{\pi,h_t}\Big),\\
&= \widehat{\rho}_t\Big(\Big\{Z_t(\theta) + \sigma_t\big(\widetilde{v}^{\pi,\theta}_{t+1}((h_t,\diamond));\, P^{\pi^{t,h_t}}_{\theta,t+1}\big),\, \theta\in\widehat{\mathbf{\Theta}}\Big\};\, \xi_t^{\pi,h_t}\Big),
\end{aligned}\tag{4.2}$$

with σ_t given in (3.18), and where we used (2.13) in the last equality.

Note that in equation (4.2) we have $\widetilde{v}^{\pi,\theta}_{t+1}$ on the right hand side. Thus, this equation does not provide a convenient recursion for the quantities v_t^π. By convenient we mean recursion in terms of v_t^π and v_{t+1}^π, rather than in terms of v_t^π and $\widetilde{v}_{t+1}^\pi$. Such convenient recursion will allow us to successfully tackle the risk-averse control problem of Sect. 5. This leads us to the following concept.

Definition 4.1 A dynamic risk filter ρ is called *recursive* if it satisfies the properties stated in Theorem 3.14 and

$$v_t^\pi(h_t) = \widehat{\rho}_t\Big(\Big\{Z_{t,t}(\theta) + \sigma_t\big(v^\pi_{t+1}((h_t,\diamond));\, P^{\pi^{t,h_t}}_{\theta,t+1}\big),\, \theta\in\widehat{\mathbf{\Theta}}\Big\};\, \xi_t^{\pi,h_t}\Big),$$

for $t = T-1,\ldots,1$, with

$$v_T^\pi(h_T) = \widehat{\rho}_t\Big(\Big\{Z_{T,T}(\theta),\, \theta\in\widehat{\mathbf{\Theta}}\Big\};\, \xi_T^{\pi,h_T}\Big).$$

Remark 4.2 In what follows, we will give examples of recursive dynamic risk-filters. In problems where the corresponding dynamic risk filter ρ is not recursive, one will need to tackle the risk averse control problem of Sect. 5 by exploiting a recursion of the pair of functions $(v_t^\pi, \widetilde{v}_t^{\pi,\theta})$, with use of equation (4.2) in particular. This will be done in a follow-up work.

4.1 Examples of recursive dynamic risk filters

The common feature of the dynamic risk filters of Example 3.17 and Example 3.18 is that in both cases we have

$$\widehat{\rho}_t(f;\xi) = U^{-1}\left(\int_{\widehat{\mathbf{\Theta}}} U(f(\theta))\,\xi(d\theta)\right),\tag{4.3}$$

for a function f on $\widehat{\mathbf{\Theta}}$ and $\xi \in \mathcal{P}(\widehat{\mathbf{\Theta}})$, and

$$\sigma_t(\nu; P) = U^{-1}\left(\int_{\mathcal{X}} U(\nu(y))\ P(dy)\right), \tag{4.4}$$

for a function ν on $\mathcal{X}$ and $P \in \mathcal{P}(\mathcal{X})$, where U is an invertible utility function. In Example 3.17: $U(a) = a$, and in Example 3.18: $U(a) = e^{\varkappa a}$. Note that in both cases it holds that

$$\widehat{\rho}_t\left(f(\diamond) + \sigma_t(w(\cdot, \diamond); P); \xi\right) = U^{-1}\left(\int_{\widehat{\mathbf{\Theta}}}\int_{\mathcal{X}} U(f(\theta) + w(\theta, y))\ P(dy)\xi(d\theta)\right). \tag{4.5}$$

Also note, that in both examples the function U is such that $\widehat{\rho}_t$ and σ_t satisfy properties stated in Theorem 3.14.

Given a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$, a utility function $U : \mathbb{R} \to \mathbb{R}$, and a real-valued random variable Y, the quantity CE satisfying

$$U(CE) = E_{\widetilde{P}}[U(Y)],$$

is called the certainty equivalent for Y relative to U and $\widetilde{P}$. If U is invertible, then

$$CE = U^{-1}\left(E_P[U(Y)]\right).$$

Therefore, we dub dynamic risk filters $\rho = \left\{\rho_t\right\}_{t=1,\ldots,T}$ given as in (3.19) and (3.20), with $\widehat{\rho}_t$ and σ_t satisfying (4.3), (4.4) and (4.5), the certainty equivalent dynamic risk filters.

We will show here that if $\rho = \left\{\rho_t\right\}_{t=1,\ldots,T}$ is a certainty equivalent dynamic risk filter, and if the function U is such that $\widehat{\rho}_t$ and σ_t satisfy the properties stated in Theorem 3.14, then ρ is a recursive dynamic risk filter. We will do this for the only case of interest to us, that is for case where, for $h_t = (x_1, \ldots, x_t) \in \mathcal{H}_t$ and $\pi \in \Pi$,

$$\begin{aligned} Z_{t,t}(\theta) &:= Z^{\pi,h_t}_{\theta,t,t} = c_t(x_t, \pi_t(h_t), \theta), \\ Z_{t,s}(x_{t+1}, \ldots, x_s, \theta) &= Z^{\pi,h_t,x_{t+1},\ldots,x_s}_{\theta,t,s} = c_s(x_s, \pi^{t,h_t}_s(x_{t+1}, \ldots, x_s), \theta),\ s = t+1, \ldots, T. \end{aligned}$$

Thus, we get

$$\begin{aligned} v^{\pi}_t(h_t) &= \rho_t\Big(c_t(x_t, \pi_t(h_t), \cdot), c_{t+1}(\cdot, \pi_{t+1}(h_t, \cdot), \cdot), \\ &\qquad c_{t+2}(\cdot, \pi_{t+2}(h_t, \cdot, \cdot), \cdot), \cdots, c_T(\cdot, \pi_T(h_t, \cdot, \ldots, \cdot), \cdot), P^{\pi^{t,h_t}}_{t+1,T}\Big) \\ &= U^{-1}\left(\mathbb{E}^{\pi}\left[U\left(c_t(x_t, \pi_t(h_t), \Theta) + \sum_{s=t+1}^{T} c_s(\widehat{X}_s, \pi_s(\widehat{X}_{t+1}, \ldots, \widehat{X}_s), \Theta)\right)\ \middle|\ \widehat{H}_t = h_t\right]\right), \end{aligned} \tag{4.6}$$

for $t = 1, \ldots, T$, where again we use the standard convention that an empty sum is zero (i.e. $\sum_{s=T+1}^{T} \cdots = 0$ in our case). Note that, in particular,

$$\begin{aligned} v_T^\pi(h_T) &= U^{-1}\left(\mathbb{E}^\pi\left[U\left(c_T(x_T, \pi_T(h_T), \Theta)\right) \mid \widehat{H}_T = h_T\right]\right) \\ &= U^{-1}\left(\int_{\widehat{\mathbf{\Theta}}} U(c_T(x_T, \pi_T(h_T), \theta))\xi_T^{\pi,h_T}(d\theta)\right) \\ &= \widehat{\rho}_T(c_T(x_T, \pi_T(h_T), \cdot); \xi_T^{\pi,h_T}). \end{aligned} \tag{4.7}$$

Now, using (4.3)–(4.5), (4.6) and Lemma 2.5, combined with the tower property of the conditional expectations, we obtain

$$\begin{aligned} v_t^\pi(\widehat{H}_t) &= U^{-1}\Bigg(\mathbb{E}^\pi\Bigg[U\Bigg(c_t(\widehat{X}_t, \pi_t(\widehat{H}_t), \Theta) \\ &\quad + U^{-1}\left(\mathbb{E}^\pi\left[U\Big(\sum_{s=t+1}^{T} c_s(\widehat{X}_s, \pi_s(\widehat{X}_{t+1}, \ldots, \widehat{X}_s), \Theta)\Big)\middle|\widehat{H}_{t+1}\right]\right)\Bigg)\middle|\widehat{H}_t\Bigg]\Bigg), \\ &= U^{-1}\left(\mathbb{E}^\pi[U(c_t(\widehat{X}_t, \pi_t(\widehat{H}_t), \Theta) + v_{t+1}^\pi(\widehat{H}_{t+1}))\mid \widehat{H}_t]\right), \end{aligned}$$

and so, using (4.5) and Lemma 2.5 again, we obtain

$$\begin{aligned} v_t^\pi(h_t) &= U^{-1}\left(\mathbb{E}^\pi[U\left(c_t(x_t, \pi_t(h_t), \Theta) + v_{t+1}^\pi(h_t, \widehat{X}_{t+1})\right) \mid \widehat{H}_t = h_t]\right) \\ &= U^{-1}\left(\int_{\mathbf{\Theta}}\int_{\mathcal{X}} U\left(c_t(x_t, \pi_t(h_t), \theta) + v_{t+1}^\pi(h_t, x_{t+1})\right) P_{\theta,t+1}^{\pi^t,h_t}(dx_{t+1})\, \xi_t^{\pi^t,h_t}(d\theta)\right) \\ &= \widehat{\rho}_t\left(\left\{c_t(x_t, \pi_t(h_t), \theta) + \sigma_t\left(v_{t+1}^\pi((h_t, \cdot)); P_{\theta,t+1}^{\pi^t,h_t}\right), \theta \in \widehat{\mathbf{\Theta}}\right\}; \xi_t^{\pi,h_t}\right), \end{aligned} \tag{4.8}$$

for $t = T-1, \ldots, 1$. In view of (4.8) and (4.7) the dynamic risk filter ρ is recursive. In particular, the dynamic risk filters of Example 3.17 and Example 3.18 are recursive.

5 Risk-averse control problem

Let v_1^π be as in (4.1). The control problem is to find

$$\min_{\pi\in\Pi} v_1^\pi(h_1), \tag{5.1}$$

as well as the optimal policy, say π^*, for which $v_1^{\pi^*}(h_1) = \min_{\pi\in\Pi} v_1^\pi(h_1)$. Note that given our set-up, an optimal policy does exist because the set Π is finite. However, we are interested in seeking an optimal policy in the class of quasi-Markov policies.

Definition 5.1 A policy $\pi \in \Pi$ is *quasi-Markov (QMP)* if

$$\pi_t(h_t) = \phi_t(x_t, \xi_t^{\pi,h_t})$$

for some function $\phi_t : \mathcal{X} \times \mathcal{P}(\widehat{\mathbf{\Theta}}) \to \mathcal{U}$, $t = 1, \ldots, T$.

Please see Remark 5.3 with regard to Definition 5.1.

5.1 The Bayes Kernel

At each time t and for every policy π and history h_t, the measure P_{t+1}^{π,h_t} (cf. (2.15)) describes the conditional joint distribution of the pair $(\widehat{X}_{t+1}, \Theta)$ in $\mathcal{X} \times \widehat{\mathbf{\Theta}}$.

This measure admits two natural disintegrations. One of them is already obtained from (2.7), repeated here:

$$\begin{aligned} P_{t+1}^{\pi,h_t}(B \times D) &= P_{t+1,T}^{\pi,h_t}(B \times \mathcal{X}^{T-t-1} \times D) = \int_D P_{\theta,t+1,T}^{\pi,h_t}(B \times \mathcal{X}^{T-t-1})\, \xi_t^{\pi,h_t}(d\theta) \\ &= P^{\pi}[\widehat{X}_{t+1} \in B, \Theta \in D \mid \widehat{H}_t = h_t], \end{aligned}$$

where $\xi_t^{\pi,h_t} \in \mathcal{P}(\mathbf{\Theta})$, is given as (cf. (2.8) and (2.11)) $\xi_t^{\pi,h_t}(D) = P^{\pi}[\Theta \in D \mid \widehat{H}_t = h_t] = P_{t+1,\Theta}^{\pi,h_t}(D)$. One can also disintegrate P_{t+1}^{π,h_t} into its marginal on $\mathcal{X}$, say[8] $P_{t+1,X}^{\pi,h_t}$, and the corresponding stochastic kernel, say $P_{t+1|X}^{\pi,h_t}$ from $\mathcal{X}$ to $\widehat{\mathbf{\Theta}}$. That is, for any $B \times D \subset \mathcal{X} \times \mathbf{\Theta}$,

$$\begin{aligned} P_{t+1}^{\pi,h_t}(B \times D) &= (P_{t+1,X}^{\pi,h_t} \circledast P_{t+1|X}^{\pi,h_t})(B \times D), \\ &= \int_B P_{t+1|x}^{\pi,h_t}(D) P_{t+1,X}^{\pi,h_t}(dx) \\ &= P^{\pi}[\widehat{X}_{t+1} \in B, \Theta \in D \mid \widehat{H}_t = h_t], \end{aligned} \tag{5.2}$$

where we used the simplified notation $P_{t+1|x}^{\pi,h_t}(D)$ for $P_{t+1|X}^{\pi,h_t}(x, D)$.

The kernel $P_{t+1|x}^{\pi,h_t}$ is the Bayes kernel which describes the dynamics of the belief states, that is the posterior distributions of Θ, as documented in the next result.

Proposition 5.2 *For $t = 1, \ldots, T-1$, $h_t \in H_t$, $x_{t+1} \in \mathcal{X}$ and $D \subset \widehat{\mathbf{\Theta}}$, we have*

$$\begin{aligned} \xi_{t+1}^{\pi,(h_t,x_{t+1})}(D) &= P_{t+1\,|\,x_{t+1}}^{\pi,h_t}(D) \\ &= \frac{\int_D K_\theta(x_{t+1}|x_t, \pi_t(h_t))\, \xi_t^{\pi,h_t}(d\theta)}{P_{t+1}^{\pi,h_t}[\{x_{t+1}\} \times \widehat{\mathbf{\Theta}}]}, \end{aligned} \tag{5.3}$$

where

$$\xi_1^{\pi,x_1}(\theta) = \xi_1(\theta). \tag{5.4}$$

Proof First, note that

$$P_{t+1,X}^{\pi,h_t}(B) = P^{\pi}[\widehat{X}_{t+1} \in B \mid \widehat{H}_t = h_t]. \tag{5.5}$$

Take $B = \{x_{t+1}\}$. Then, using (5.2) and (5.5), we obtain

$$P^{\pi}[\widehat{X}_{t+1} = x_{t+1}, \Theta \in D \mid \widehat{H}_t = h_t] = P_{t+1|x_{t+1}}^{\pi,h_t}(D) P^{\pi}[\widehat{X}_{t+1} = x_{t+1} \mid \widehat{H}_t = h_t],$$

[8] For simplicity of notations, we write $P_{t+1,X}^{\pi,h_t}$ instead of more coherent notation $P_{t+1,X_{t+1}}^{\pi,h_t}$. Similar remark applies to the kernel $P_{t+1|X}^{\pi,h_t}$.

and thus

$$P^{t,h_t}_{t+1|x_{t+1}}(D) = \frac{P^\pi[\widehat{X}_{t+1} = x_{t+1}, \Theta \in D \mid \widehat{H}_t = h_t]}{P^\pi[\widehat{X}_{t+1} = x_{t+1} \mid \widehat{H}_t = h_t]}$$
$$= P^\pi[\Theta \in D \mid \widehat{H}_{t+1} = (h_t, x_{t+1})] = \xi^{\pi,(h_t,x_{t+1})}_{t+1}(D),$$

which proves the first equality in (5.3). The second one follows from the following chain of equalities,

$$\xi^{\pi,(h_t,x_{t+1})}_{t+1}(\theta) = P^\pi[\Theta = \theta \mid \widehat{H}_{t+1} = (h_t, x_{t+1})]$$
$$= P^\pi[\Theta = \theta \mid \widehat{H}_t = h_t] \frac{P^\pi[\widehat{X}_{t+1} = x_{t+1} \mid \widehat{\mathbf{\Theta}} = \theta, \widehat{H}_t = h_t]}{P^\pi[\widehat{X}_{t+1} = x_{t+1} \mid \widehat{H}_t = h_t]}$$
$$= \xi^{\pi,h_t}_t(\theta) \frac{K_\theta(x_{t+1}|x_t, \pi_t(h_t))}{P^{\pi^{t,h_t}}_{t+1}[\{x_{t+1}\} \times \widehat{\mathbf{\Theta}}]},$$

where in the last equality we used (2.1) and (2.2) to deduce that $P^\pi[\widehat{X}_{t+1} = x_{t+1} \mid \widehat{\mathbf{\Theta}} = \theta, \widehat{H}_t = h_t] = K_\theta(x_{t+1}|x_t, \pi_t(h_t))$, and we used (2.15) to deduce that $P^\pi[\widehat{X}_{t+1} = x_{t+1} \mid \widehat{H}_t = h_t] = P^{\pi^{t,h_t}}_{t+1}[\{x_{t+1}\} \times \widehat{\mathbf{\Theta}}]$. □

Remark 5.3 Note that (5.3) represents learning about the unknown parameter θ^* in the sense of updating the posterior distributions of Θ. Also, in view of (5.3) one might surmise that, by extending the canonical space by products of $\mathcal{P}(\widehat{\mathbf{\Theta}})$, the process (X_t, ξ^{ϕ,H_t}_t) would be Markov for a Markov strategy, say $\psi_t(X_t, \xi^{\psi,H_t}_t)$, and therefore one might seek an optimal strategy in the class of Markov strategies. This however is not as straightforward as it might appear. One of the reasons being that even though (5.3) is a deterministic recursion, it is not exactly of Markovian type. This is why we work here with quasi-Markov policies. The issue of Markov strategies will be investigated in a follow-up paper.

5.2 The optimal control problem corresponding to Example 3.16

In this section we will study the optimal control problem corresponding to the Example 3.16 classical additive reward case, that will serve as the base for the general case. In what follows, we denote by (x, ξ) an element of the set $\mathcal{X} \times \mathcal{P}(\widehat{\mathbf{\Theta}})$.

Recall (4.6). Accordingly, we have for $t = T$

$$v^\pi_T(h_T) = \rho_{T,T}\Big(c_T(x_T, \pi_T(h_T), \cdot), P^{\pi^{T,h_T}}_{T+1,T}\Big)$$
$$= \int_{\widehat{\mathbf{\Theta}}} c_T(x_T, \pi_T(h_T), \theta))\ P^{\pi^{T,h_T}}_{T+1,T}(d\theta)$$
$$= \int_{\widehat{\mathbf{\Theta}}} c_T(x_T, \pi_T(h_T), \theta)\ \xi^{\pi,h_T}_T(d\theta)$$
$$= \mathbb{E}^\pi\Big(c_T(x_T, \pi_T(h_T), \Theta) \mid \widehat{H}_T = h_T\Big). \tag{5.6}$$

Thus, observing that ξ_T^{π,h_T}, does not depend on π_T, letting $x_T = x$ and $\xi_T^{\pi,h_T} = \xi$, we compute the candidate-optimal quasi-Markov control ϕ_T as

$$\phi_T(x,\xi) = \arg\min_{u\in\mathcal{U}} \int_{\widehat{\Theta}} c_T(x,u,\theta)\,\xi(d\theta). \tag{5.7}$$

We define the Bellman function at time $t = T$:

$$V_T(x,\xi) = \min_{u\in\mathcal{U}} \int_{\widehat{\Theta}} c_T(x,u,\theta)\,\xi(d\theta) = \int_{\widehat{\Theta}} c_T(x,\phi_T(x,\xi),\theta)\,\xi(d\theta). \tag{5.8}$$

Now, we proceed to time $t = T-1$. Noting that $\xi_{T-1}^{\pi,h_{T-1}}$, does not depend on π_{T-1}, letting $x_{T-1} = x$ and $\xi_{T-1}^{\pi,h_{T-1}} = \xi$, we compute the candidate-optimal quasi-Markov control ϕ_{T-1} as

$$\begin{aligned}\phi_{T-1}(x,\xi) = \arg\min_{u\in\mathcal{U}} \int_{\widehat{\Theta}} \Big(& c_{T-1}(x,u,\theta) \\ &+ \int_{\mathcal{X}} V_T(x_T, \widetilde{\xi}_T^{u,x_T,\xi}) K_\theta(dx_T|x,u)\Big)\,\xi(d\theta),\end{aligned} \tag{5.9}$$

where (cf. (5.3))

$$\widetilde{\xi}_T^{u,x_T,\xi}(\theta) = \xi(\theta)\frac{K_\theta(x_T|x,u)}{\int_{\widehat{\Theta}} K_\theta(x_T|x,u)\,\xi(d\theta)}. \tag{5.10}$$

The corresponding Bellman function is

$$\begin{aligned}V_{T-1}(x,\xi) &= \min_{u\in\mathcal{U}} \int_{\widehat{\Theta}} \Big(c_{T-1}(x,u,\theta) + \int_{\mathcal{X}} V_T(x_T, \widetilde{\xi}_T^{u,x_T,\xi}) K_\theta(dx_T|x,u)\Big)\,\xi(d\theta) \\ &= \int_{\widehat{\Theta}} \Big(c_{T-1}(x,\phi_{T-1}(x,\xi),\theta) \\ &\quad + \int_{\mathcal{X}} V_T(x_T, \widetilde{\xi}_T^{\phi_{T-1}(x,\xi),x_T,\xi}) K_\theta(dx_T|x,\phi_{T-1}(x,\xi))\Big)\,\xi(d\theta).\end{aligned}$$

Following this pattern, we arrive at the dynamic programming (DP) backward recursion:

$$V_t(x,\xi) = \min_{u\in\mathcal{U}} \int_{\widehat{\Theta}} \Big(c_t(x,u,\theta) + \int_{\mathcal{X}} V_{t+1}(x_{t+1}, \widetilde{\xi}_{t+1}^{u,x_{t+1},\xi}) K_\theta(dx_{t+1}|x,u)\Big)\,\xi(d\theta), \quad t\in\mathcal{T}, \tag{5.11}$$

where (cf. (5.3))

$$\widetilde{\xi}_{t+1}^{u,x_{t+1},\xi}(\theta) = \xi(\theta)\frac{K_\theta(x_{t+1}|x,u)}{\int_{\widehat{\Theta}} K_\theta(x_{t+1}|x,u)\,\xi(d\theta)}, \tag{5.12}$$

and

$$V_{T+1} \equiv 0. \tag{5.13}$$

Accordingly, for $t = 1, \ldots, T$ we define the candidate-optimal quasi-Markov control ϕ_t as

$$\phi_t(x, \xi) = \arg\min_{u \in \mathcal{U}} \int_{\widehat{\Theta}} \Big(c_t(x, u, \theta) + \int_{\mathcal{X}} V_{t+1}(x_{t+1}, \widetilde{\xi}_{t+1}^{u, x_{t+1}, \xi}) K_\theta(dx_{t+1} | x, u) \Big) \xi(d\theta). \tag{5.14}$$

Recall that ξ_1 is a given prior distribution for Θ. Also, recall that $h_1 = x_1$.
Next, define a policy π^* as follows,

$$\begin{aligned} \pi_1^*(h_1) &= \phi_1(x_1, \xi_1) \\ \pi_t^*(h_t) &= \phi_t(x_t, \widehat{\xi}_t^{\pi^*, h_t}), \quad t = 2, \ldots, T, \end{aligned} \tag{5.15}$$

where

$$\widehat{\xi}_2^{\pi^*, h_2} = \widetilde{\xi}_2^{\pi_1^*(h_1), x_2, \xi_1}, \quad \widehat{\xi}_3^{\pi^*, h_3} = \widetilde{\xi}_3^{\pi_2^*(h_2), x_3, \widehat{\xi}_2^{\pi^*, h_2}}, \ldots \tag{5.16}$$

The next result is the optimality verification theorem.

Theorem 5.4 *We have,*

$$\min_{\pi \in \Pi} v_1^\pi(h_1) = v_1^{\pi^*}(h_1) = V_1(x_1, \xi_1).$$

Proof Let $\pi \in \Pi$. For $t = T$ we have

$$v_T^\pi(h_T) \geq V_T(x_T, \xi_T^{\pi, h_T}) = \int_{\widehat{\Theta}} c_T(x_T, \phi_T(x_T, \xi_T^{\pi, h_T}), \theta) \, \xi_T^{\pi, h_T}(d\theta).$$

For $t = T - 1$, using the above, the recursion in (4.8), and (5.3), we have

$$\begin{aligned} & v_{T-1}^\pi(h_{T-1}) \\ &= \int_{\widehat{\Theta}} \Big(c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \theta) + \int_{\mathcal{X}} v_T^\pi(h_T) K_\theta\big(dx_T | x_{T-1}, \pi_{T-1}(h_{T-1})\big) \Big) \xi_{T-1}^{\pi, h_{T-1}}(d\theta) \\ &\geq \int_{\widehat{\Theta}} \Big(c_{T-1}(x_{T-1}, \pi_{T-1}(h_{T-1}), \theta) \\ &\quad + \int_{\mathcal{X}} V_T(x_T, \xi_T^{\pi, h_T}) K_\theta\big(dx_T | x_{T-1}, \pi_{T-1}(h_{T-1})\big) \Big) \xi_{T-1}^{\pi, h_{T-1}}(d\theta) \\ &\geq \int_{\widehat{\Theta}} \Big(c_{T-1}(x_{T-1}, \phi_{T-1}(x_{T-1}, \xi_{T-1}^{\pi, h_{T-1}}), \theta) \\ &\quad + \int_{\mathcal{X}} V_T(x_T, \widetilde{\xi}_T^{\phi_{T-1}(x_{T-1}, \xi_{T-1}^{\pi, h_{T-1}}), x_T, \xi_{T-1}^{\pi, h_{T-1}}}) K_\theta\big(dx_T | x_{T-1}, \phi_{T-1}(x_{T-1}, \xi_{T-1}^{\pi, h_{T-1}})\big) \Big) \xi_{T-1}^{\pi, h_{T-1}}(d\theta) \\ &= V_{T-1}(x_{T-1}, \xi_{T-1}^{\pi, h_{T-1}}). \end{aligned}$$

Likewise, for $t = 1, \dots, T-2$, we have

$$v_t^{\pi}(h_t) \geq V_t(x_t, \xi_t^{\pi,h_t}) = \int_{\widehat{\Theta}} \Big(c_t(x_t, \phi_t(x_t, \xi_t^{\pi,h_t}), \theta) + \int_{\mathcal{X}} V_{t+1}(x_{t+1}, \phi_{t+1}(x_{t+1}, \widetilde{\xi}_{t+1}^{\phi_t(x_t,\xi_t^{\pi,h_t}),x_{t+1},\xi_t^{\pi,h_t}})) K_\theta\big(dx_{t+1}|x_t, \phi_t(x_t, \xi_t^{\pi,h_t})\big)\Big) \xi_t^{\pi,h_t}(d\theta).$$

Now, if π and ξ_t^{π,h_t}, $t \in \mathcal{T}$, above are replaced with π^* and $\widehat{\xi}_t^{\pi^*,h_t}$, $t \in \mathcal{T}$, respectively, then the inequalities above become equalities, proving that π^* is an optimal strategy. □

Recalling (3.25) and (3.26), we note that the key DP recursion (5.9) can be written as

$$V_t(x,\xi) = \min_{u \in \mathcal{U}} \widehat{\rho}_t\Big(\Big\{c_t(x,u,\theta) + \sigma_t(V_{t+1}(\cdot, \widetilde{\xi}_{t+1}^{u,\cdot,\xi}); K_\theta(x,u)), \theta \in \widehat{\mathbf{\Theta}}\Big\}; \xi\Big),$$

subject to (5.12) and (5.13).

We remark that the solution to the optimal control problem associated with Example 3.19 is also given by Theorem 5.4. Detailed model specification and analysis of Example 3.19 is beyond the scope of this work and will be addressed in future work.

5.3 The optimal control problem corresponding to Example 3.18

We will present the solution to the optimal control problem for the clinical trials example with the risk-sensitive criterion. Namely, for $t \in \mathcal{T}$, we consider

$$\begin{aligned} v_t^{\pi}(h_t) &= \rho_t\Big(c_t(x_t, \pi_t(h_t), \cdot), c_{t+1}(\cdot, \pi_{t+1}(h_t, \cdot), \cdot), \\ &\qquad c_{t+2}(\cdot, \pi_{t+2}(h_t, \cdot, \cdot), \cdot), \cdots, c_T(\cdot, \pi_T(h_t, \cdot, \dots, \cdot), \cdot), P_{t+1,T}^{\pi^{t,h_t}}\Big) \\ &= \frac{1}{\varkappa} \ln \mathbb{E}^{\pi}\Big(\exp\Big(\varkappa\Big(c_t(x_t, \pi_t(h_t), \Theta) + \sum_{k=t+1}^{T} c_k(\widehat{X}_k, \pi_k(h_t, \widehat{X}_{t+1}, \dots, \widehat{X}_k), \Theta)\Big)\Big) \Big| \widehat{H}_t = h_t\Big). \end{aligned} \tag{5.17}$$

For $t = T$ we have

$$\begin{aligned} v_T^{\pi}(h_T) &= \frac{1}{\varkappa} \ln\Big(\mathbb{E}^{\pi}\Big(\exp\left(\varkappa\, c_T(x_T, \pi_T(h_T), \Theta)\right) \big| \widehat{H}_T = h_T\Big)\Big) \\ &= \frac{1}{\varkappa} \ln\Big(\int_{\widehat{\Theta}} e^{\varkappa\, c_T(x_T, \pi_T(h_T), \theta)} \xi_T^{\pi,h_T}(d\theta)\Big). \end{aligned} \tag{5.18}$$

As above, we denote by (x,ξ) an element of the set $\mathcal{X} \times \mathcal{P}(\widehat{\mathbf{\Theta}})$. Thus, observing that ξ_T^{π,h_T} does not depend on π_T, and letting $x_T = x$ and $\xi_T = \xi$, we compute the candidate optimal quasi-Markov control φ_T as

$$\varphi_T(x,\xi) = \arg\min_{u \in \mathcal{U}} \int_{\widehat{\Theta}} e^{\varkappa\, c_T(x,u,\theta)}\, \xi(d\theta). \tag{5.19}$$

We define the Bellman function at time $t = T$ as

$$V_T(x,\xi) = \min_{u\in\mathcal{U}} \frac{1}{\varkappa} \ln\left(\int_{\widehat{\Theta}} e^{\varkappa c_T(x,u,\theta)}\, \xi(d\theta)\right) = \frac{1}{\varkappa} \ln\left(\int_{\widehat{\Theta}} e^{\varkappa c_T(x,\varphi_T(x,\xi),\theta)}\, \xi(d\theta)\right). \tag{5.20}$$

Now, we proceed to time $t = T-1$. Given $x_{T-1} = x$ and $\xi_{T-1} = \xi$ we compute the candidate optimal quasi-Markov control φ_{T-1} as

$$\varphi_{T-1}(x,\xi) = \arg\min_{u\in\mathcal{U}} \int_{\widehat{\Theta}} \int_{\mathcal{X}} e^{\varkappa c_{T-1}(x,u,\theta)} V_T(x_T, \widetilde{\xi}_T^{u,x_T,\xi}) K_\theta(dx_T|x,u)\, \xi(d\theta), \tag{5.21}$$

where $\widetilde{\xi}_T^{u,x_T,\xi}$ is given by (5.10). The corresponding Bellman function is

$$V_{T-1}(x,\xi) = \min_{u\in\mathcal{U}} \frac{1}{\varkappa} \ln\left(\int_{\widehat{\Theta}} \int_{\mathcal{X}} e^{\varkappa c_{T-1}(x,u,\theta)} V_T(x_T, \widetilde{\xi}_T^{u,x_T,\xi})\; K_\theta(dx_T|x,u)\; \xi(d\theta)\right)$$
$$= \frac{1}{\varkappa} \ln\left(\int_{\widehat{\Theta}} \int_{\mathcal{X}} e^{\varkappa c_{T-1}(x,\varphi_{T-1}(x,\xi),\theta)} V_T(x_T, \widetilde{\xi}_T^{\varphi_{T-1}(x,\xi),x_T,\xi})\; K_\theta(dx_T|x,\varphi_{T-1}(x,\xi))\; \xi(d\theta)\right). \tag{5.22}$$

Following this pattern, we arrive at the DP backward recursion:

$$V_t(x,\xi) = \min_{u\in\mathcal{U}} \frac{1}{\varkappa} \ln\left(\int_{\widehat{\Theta}} \int_{\mathcal{X}} e^{\varkappa c_t(x,u,\theta)} V_{t+1}(x_{t+1}, \widetilde{\xi}_{t+1}^{u,x_{t+1},\xi})\; K_\theta(dx_{t+1}|x,u)\; \xi(d\theta)\right), \quad t\in\mathcal{T}, \tag{5.23}$$

whereas in the previous example $\widetilde{\xi}_{t+1}^{u,x_{t+1},\xi}(\{\theta\})$ is given by (5.12), and $V_{T+1} \equiv 1$.

Accordingly, for $t \in \mathcal{T}$, we define the candidate-optimal quasi-Markov control φ_t as

$$\varphi_t(x,\xi) = \arg\min_{u} \int_{\widehat{\Theta}} \int_{\mathcal{X}} e^{\varkappa c_t(x,u,\theta)} V_{t+1}(x_{t+1}, \widetilde{\xi}_{t+1}^{u,x_{t+1},\xi})\; K_\theta(dx_{t+1}|x,u)\; \xi(d\theta), \tag{5.24}$$

with ξ_1 being the given prior distribution for Θ, and $h_1 = x_1$.

The policy π^* is defined by analogy to (5.15). The following verification theorem can proved in a way analogous to the proof of Theorem 5.4, so we skip its proof.

Theorem 5.5 *The following holds true*

$$\min_{\pi\in\Pi} v_1^\pi(h_1) = v_1^{\pi^*}(h_1) = V_1(x_1,\xi_1).$$

We emphasize that the key DP recursion (5.23) may be written as

$$V_t(x,\xi) = \min_{u\in\mathcal{U}} \widehat{\rho}_t\Big(\big\{c_t(x,u,\theta) + \sigma_t\big(V_{t+1}(\cdot, \widetilde{\xi}_{t+1}^{u,\cdot,\xi});\, K_\theta(x,u)\big), \theta\in\widehat{\Theta}\big\};\, \xi\Big),$$

where for a function f on $\widehat{\mathbf{\Theta}}$, $\xi \in \mathcal{P}(\widehat{\mathbf{\Theta}})$, and a function h on $\mathcal{X}$, we have

$$\widehat{\rho}_t\big(\{f(\theta), \theta \in \widehat{\mathbf{\Theta}}\}; \xi\big) = \frac{1}{\varkappa} \ln\left(\int_{\widehat{\mathbf{\Theta}}} e^{\varkappa f(\theta)} \xi(d\theta)\right),$$

and where

$$\sigma_t\big(h; K_\theta(x,u)\big) = \frac{1}{\varkappa} \ln\left(\int_{\mathcal{X}} e^{\varkappa h(x_{t+1})} K_\theta(dx_{t+1}|x,u)\right).$$

5.4 Solution of the optimal control problem for general recursive risk filters

Let ρ be a recursive risk filter, and let

$$v_t^\pi(h_t) = \rho_t\Big(c_t(x_t, \pi_t(h_t), \cdot), c_{t+1}(\cdot, \pi_{t+1}(h_t, \cdot), \cdot), c_{t+2}(\cdot, \pi_{t+2}(h_t, \cdot, \cdot), \cdot), \cdots\cdots,$$
$$c_T(\cdot, \pi_T(h_t, \cdot, \ldots, \cdot), \cdot), P_{t+1,T}^{\pi, h_t}\Big), \quad t \in \mathcal{T}. \tag{5.25}$$

Consider the general problem (5.1) with $v_t^\pi(h_t)$ as in (5.25).

Using reasoning analogous to the one employed in Sects. 5.2 and 5.3 one can prove the following result, proof of which we omit here.

Theorem 5.6 *There exist operators $\widehat{\rho}_t$ and σ_t, $t \in \mathcal{T}$, and a function V^* such that for the functions v_t^* defined recursively as*

$$v_{T+1}^*(x) = V^*(x), \ x \in \mathcal{X},$$
$$v_t^*(x,\xi) = \min_{u \in \mathcal{U}} \widehat{\rho}_t\Big(\Big\{c_t(x,u,\theta) + \sigma_t(v_{t+1}^*(\cdot, \widetilde{\xi}_{t+1}^{u,\cdot,\xi}); K_\theta(x,u)), \theta \in \widehat{\mathbf{\Theta}}\Big\}; \xi\Big),$$
$$t = T-1, \ldots, 1, \quad x \in \mathcal{X}, \ \xi \in \mathcal{P}(\widehat{\mathbf{\Theta}}),$$

subject to

$$\widetilde{\xi}_{t+1}^{u,x',\xi}(\theta) = \xi(\theta) \frac{K_\theta(x'|x,u)}{\int_{\widehat{\mathbf{\Theta}}} K_\theta(x'|x,u)\, \xi(d\theta)}, \quad t \in \mathcal{T}, \ x, x' \in \mathcal{X}, \ \xi \in \mathcal{P}(\widehat{\mathbf{\Theta}}),$$

we have that

$$\min_{\pi \in \Pi} v_1^\pi(h_1) = v_1^*(x_1, \xi_1).$$

Moreover, the policy π^* defined as in (5.15) and (5.16), with the ϕ_t's given as

$$\phi_t(x,\xi) = \arg\min_{u \in \mathcal{U}} \widehat{\rho}_t\Big(\Big\{c_t(x,u,\theta) + \sigma_t(v_{t+1}^*(\cdot, \widetilde{\xi}_{t+1}^{u,\cdot,\xi}); K_\theta(x,u)), \theta \in \widehat{\mathbf{\Theta}}\Big\}; \xi\Big),$$
$$t = 1, \ldots, T-1, \ x \in \mathcal{X}, \ \xi \in \mathcal{P}(\widehat{\mathbf{\Theta}}),$$

is an optimal policy, that is

$$\min_{\pi \in \Pi} v_1^\pi(h_1) = v_1^{\pi^*}(h_1). \tag{5.26}$$

The form of the operators $\widehat{\rho}_t$ and σ_t, $t \in \mathcal{T}$, depends on the form of ρ, and it can be explicitly written in terms of ρ.

Acknowledgements The research of Andrzej Ruszczyński has benefited from partial support from National Science Foundation Award DMS-1907522 and by the Office of Naval Research Award N00014-21-1-2161. Tomasz R. Bielecki and Igor Cialenco acknowledge the support from the National Science Foundation Grant No. DMS-1907568.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

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

Bielecki TR, Chen T, Cialenco I, Cousin A, Jeanblanc M (2019) Adaptive robust control under model uncertainty. SIAM J Control Optim 57(2):925–946

Bäuerle N, Rieder U (2014) More risk-sensitive Markov decision processes. Math Oper Res 39(1):105–120

Bäuerle N, Rieder U (2017) Zero-sum risk-sensitive stochastic games. Stoch Process Appl 127(2):622–642

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 Theoret Appl Finan 14(01):137–162

Davis MHA, Lleo S (2014) Risk-sensitive investment management. Advanced series on statistical science & applied probability. Vol. 19. World Scientific

Dentcheva D, Ruszczyński A (2020) Risk forms: representation, disintegration, and application to partially observable two-stage systems. Math Program 181(2):297–317

Fan J, Ruszczyński A (2018) Risk measurement and risk-averse control of partially observable discrete-time Markov systems. Math Methods Oper Res 88(2):161–184

Fan J, Ruszczyński A (2022) Process-based risk measures and risk-averse control of discrete-time systems. Math Program, 191(1, Ser. B):113–140

Frittelli M, Scandolo G (2006) Risk measures and capital requirements for processes. Math. Finance 16(4):589–612

Krishnamurthy V (2016) Partially observed Markov decision processes: from filtering to controlled sensing. Cambridge University Press, Cambridge

Lin Y, Ren Y, Zhou E (2021) A Bayesian risk approach to MDPs with parameter uncertainty. Preprint arXiv:2106.02558,

Lattimore T, Szepesvári C (2020) Bandit algorithms. Cambridge University Press, Cambridge

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

Ruszczyński A, Shapiro A (2006) Conditional risk mappings. Math Oper Res 31(3):544–561

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

Sutton RS, Barto AG (2018) Reinforcement learning: an introduction. MIT press, Cambridge

Scandolo G (2003) Risk measures in a dynamic setting. PhD thesis, Universitia degli Studi di Milano, Italy,

Shapiro A, Dentcheva D, Ruszczyński A (2021) Lectures on stochastic programming: modeling and theory (3rd ed). SIAM, Philadelphia

Wolff EM, Topcu U, Murray RM (2012) Robust control of uncertain Markov decision processes with temporal logic specifications. In: 2012 IEEE 51st IEEE Conference on decision and control (CDC), pp. 3372–3379

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.