PROCESS-BASED RISK MEASURES AND RISK-AVERSE CONTROL OF OBSERVABLE AND PARTIALLY OBSERVABLE DISCRETE-TIME SYSTEMS

by

JINGNAN FAN

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Operations Research

Written under the direction of

Andrzej Ruszczyński

And approved by

New Brunswick, New Jersey

JANUARY, 2018

ABSTRACT OF THE DISSERTATION

Process-based Risk Measures and Risk-Averse Control of Observable and Partially Observable Discrete-Time Systems

By Jingnan Fan

Dissertation Director:
Andrzej Ruszczyński

In this thesis, we develop theoretical foundations of the theory of dynamic risk measures for controlled stochastic processes, and we apply our theory to Markov decision processes (MDP) and partially observable Markov decision processes (POMDP).

We consider a new class of dynamic risk measures for controlled discrete-time stochastic processes, which we call process-based. By introducing a new concept of stochastic conditional time consistency, we derive the structure of process-based risk measures enjoying this property. It is shown that such risk measures can be equivalently represented by a collection of static law-invariant risk measures on the space of functions of the state of the base process.

The results are first specialized to Markov decision problems (MDP), in which we use process-based dynamic risk measures to evaluate control policies. We derive the refined structure of risk measures for this kind of problems, along with the associated dynamic programming equations.

We then specialize our theory to partially observable Markov decision problems (POMDP). Compared to MDP, in POMDP we can only observe part of the state, and we need to infer the rest of the state conditional on our observations. We derive that the stochastically conditionally time-consistent dynamic risk measures can be represented by a sequence of law-invariant risk measures on the space of function of the observable part of the state. The corresponding dynamic programming equations are also derived.

Finally, as an application to our theory on POMDP, we study a model for machine deterioration problem.

Acknowledgements

I would like to begin by thanking Professor Andrzej Ruszczyński who introduced me to the universe of dynamic risk measures. His passion for research and genuine curiosity are highly contagious. He was always there to help and encourage me when I needed him. He not only taught me how to be a better researcher, but also taught me how to be a better person. It is without doubt that I have benefited immensely from his vast mathematical knowledge and remarkable personality, and this thesis would not have existed without him.

I am very grateful to Professor Eckstein, Professor Boros, Professor Katehakis and Professor Collado for having agreed to join the committee of examiners. Thanks also go out to all my colleagues who made my graduate experience at Rutgers so much fun.

Finally, I would like to thank my parents and my husband for their support; this thesis is dedicated to them.

Dedication

To my family.

Table of Contents

Chapter 1

Introduction and Preliminaries

1.1 Introduction

Quantifying risk in a dynamic multistage setup is a research subject that is intruiging and of immense practical importance. In the extant literature, three basic approaches to introduce risk aversion for discrete-time processes have been employed: utility functions (see, e.g., [31, 32, 19, 7, 33]), mean–variance models (see. e.g., [63, 23, 42, 1]), and entropic (exponential) models (see, e.g., [29, 43, 10, 17, 20, 40, 7]).

In recent years, as a multistage extension to the classical theory of risk measures, the theory of dynamic risk measures emerges as a more general tool for quantifying risk compared to the approaches mentioned above, see [58, 48, 53, 24, 15, 55, 4, 47, 35, 34, 16] and the references therein. The basic setting is the following: we have a probability space $(\Omega, \mathcal{F}, P)$, a filtration $\{\mathcal{F}_t\}_{t=1,\dots,T}$ with a trivial $\mathcal{F}_1$, and we define appropriate spaces $\mathcal{Z}_t$ of $\mathcal{F}_t$-measurable random variables, $t = 1, \dots, T$. For each $t = 1, \dots, T$, a mapping $\rho_{t,T} : \mathcal{Z}_T \to \mathcal{Z}_t$ is called a *conditional risk measure*. The central role in the theory is played by the concept of *time consistency*, which regulates relations between the mappings $\rho_{t,T}$ and $\rho_{s,T}$ for different s and t. One definition employed in the literature is the following: *for all $Z, W \in \mathcal{Z}_T$, if $\rho_{t,T}(Z) \le \rho_{t,T}(W)$ then $\rho_{s,T}(Z) \le \rho_{s,T}(W)$ for all $s < t$.* This can be used to derive recursive relations $\rho_{t,T}(Z) = \rho_t\big(\rho_{t+1,T}(Z)\big)$, with simpler *one-step conditional risk mappings* $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \dots, T-1$. Much effort has been devoted to derive dual representations of the conditional risk mapping and to study their evolution in various settings.

When applied to processes described by controlled kernels, in particular, to Markov processes, the theory of dynamic measures of risk encounters difficulties. The spaces $\mathcal{Z}_t$ are different for different t, and thus each one-step mapping ρ_t has different domain and range spaces. With $\mathcal{Z}_t$ containing all $\mathcal{F}_t$ measurable random variables, arbitrary dependence of ρ_t on the past is allowed. Moreover, no satisfactory theory of law-invariant dynamic risk measures exists, which would be suitable for Markov control problems (the restrictive definitions of law invariance employed in [37] and [59] lead to conclusions of limited practical usefulness, while the translation of the approach of [62] to the Markov case appears to be difficult). These difficulties are compounded in the case of controlled processes, when a control policy changes the probability measure on the space of paths of the process. Risk measurement of the entire family of processes defined by control policies is needed.

Motivated by these issues, a specific class of dynamic risk measures was introduced in [57], which is well-suited for Markov problems. It was postulated in [57] that the one-step conditional risk mappings ρ_t have a special form, which allows for their representation in terms of static risk measures on the space of functions defined on the state space of the Markov process. This restriction allowed for the development of dynamic programming equations and corresponding solution methods, which generalizes the well-known results for expected value problems. The ideas were successfully extended in [13, 12, 41, 61]. However, the construction of the Markov risk measures appeared somewhat arbitrary.

Aiming at building more solid theoretical foundations for the Markov risk measures introduced in [57], in this thesis, we introduce and analyze a general class of risk measures, which we call *process-based*. We consider a controlled process $\{X_t\}_{t=1,\dots,T}$ taking values in a *state space* $\mathcal{X}$, whose conditional distributions are described by controlled history-dependent transition kernels

$$Q_t : \mathcal{X}^t \times \mathcal{U} \to \mathcal{P}(\mathcal{X}), \quad t = 1, \dots, T-1,$$

where

$$\mathcal{X}^t = \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{t \text{ times}},$$

and $\mathcal{U}$ is a certain *control space*. Any history-dependent (measurable) control $u_t = \pi_t(x_1, \dots, x_t)$ is allowed. In this setting, we are only interested in measuring risk of stochastic processes of the form $Z_t = c_t(X_t, u_t)$, $t = 1, \dots, T$, where $c_t : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ can be any bounded measurable function. This restriction of the class of stochastic processes for which risk needs to be measured is one of the two cornerstones of our approach. The other cornerstone is our new concept of *stochastic conditional time consistency*. It is more restrictive than the usual time consistency, because it involves conditional distributions and uses stochastic dominance rather than the pointwise order.

These two foundations allow for the development of a theory of dynamic risk measures for controlled processes. We demonstrate that dynamic risk measures for history-dependent controlled processes can be fully described by a sequence of static law-invariant risk measures on a space $\mathcal{V}$ of measurable functions on the state space $\mathcal{X}$. This proves that our theory generalizes the utility and entropic models. On the other hand, the mean–variance models do not satisfy, in general, the monotonicity and time consistency conditions, except the version of [14].

In the special case of controlled Markov processes, we derive the structure postulated in [57], thus providing its solid theoretical foundations. We also derive dynamic programming equations in a much more general setting than that of [57]. For multistage stochastic programming problems with decision-dependent probabilities, we derive from our axioms the form of risk measures on a scenario tree and we also derive the associated dynamic programming equations.

In the second part of the thesis, we develop risk theory for *partially observable* controlled Markov processes. In the expected-value case, this classical topic is covered in many monographs (see, e.g., [28, 9, 6] and the references therein). The standard approach is to consider the belief state space, involving the space of probability distributions of the unobserved part of the state. The recent article [21] provides the state-of-the-art setting. The risk-averse case has been dealt, so far, with the use of the entropic risk measure [30, 22]. A more general partially-observable utility model was recently analyzed in [8].

In a partially-observable model, with the state process $\{X_t, Y_t\}_{t=1,\dots,T}$, where X_t is observable at time t, while $\{Y_t\}$ is unobservable, the concepts of conditional and dynamic risk measures are insufficient. The reason is that the cost process, in general, is adapted to the full filtration $\{\mathcal{F}_t^{X,Y}\}$ defined by the full state process $\{X_t, Y_t\}$, while the dynamic risk evaluation has to be available at each time t, and thus must be adapted to the sub-filtration $\{\mathcal{F}_t^X\}$ defined by the observable part of the state process. To deal with this difficulty, we introduce the concept of a *risk filter*. We postulate a new property of stochastic conditional time consistency of such a filter. Our main result is that the risk filters can be equivalently modeled by special forms of transition risk mappings: static risk measures on the space of functions defined on the observable part of the state only. We also derive dynamic programming equations for risk-averse partially observable Markov models. In these equations, the state space comprises belief states and observable states, as in the expected value model, but the conditional expectation is replaced by a transition risk mapping.

1.2 Outline of the Dissertation

The thesis is organized as follows.

- We briefly review the fundamental theory of static risk measures in the rest of this chapter.

- In Chapter 2 and 3 we consider general discrete-time stochastic processes and introduce a new class of risk measures that are process-based. A new property for process-based dynamic risk measures, which we call stochastic conditional time consistency, is also introduced (Definitions 2.3.1, 3.2.1). Respectively for uncontrolled processes (Chapter 2) and controlled processes (Chapter 3), we characterize the structure of dynamic risk measures enjoying this property (Theorems 2.4.5, 3.2.3).

- In Chapter 4 we specialize the concepts and results to controlled Markov processes. We introduce the concept of Markov risk measures and we derive its structure (Theorem 4.1.3). In Section 4.2, we prove the dynamic programming equations in this case. In Section 4.3, we consider multistage stochastic programming problems in which the exogenous (data) process is history-dependent, while the physical state evolves according to a controlled state equation.

- In Chapter 5 we extend our theory to partially observable Markov process (POMDP). The concept of risk filters is introduced (Section 5.2) and their structures are derived under stochastic conditional time consistency (Theorem 5.3.8). In Section 5.4, we prove the dynamic programming equations for risk-averse partially observable models.

- In Chapter 6, we apply our theory for POMDP to a machine replacement problem.

1.3 Static Risk Measures

The modern theory of static risk measures has been established since late 1990's [2, 3, 25, 26, 27], while a list of axioms are imposed to define the concept of coherent risk measures. We would like to point out that our convention throughout the thesis might be different compared to some literature: we adopt the convention that smaller values of random variables are preferred, while the values can be interpreted as "costs". Let us begin with presenting the rigorous definition for risk measures.

Definition 1.3.1. *Assume that $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathcal{Z} = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ is the space of all p-integrable random variables with $p \in [1, +\infty]$. A function $\rho : \mathcal{Z} \to \mathbb{R}$ is called a* ***convex risk measure*** *if it satisfies the following axioms:*

- ***Monotonicity****. If $V, Z \in \mathcal{Z}$ and $V \leq Z$, then $\rho(V) \leq \rho(Z)$.*

- ***Convexity****. $\rho(\alpha Z + (1-\alpha)V) \leq \alpha\rho(Z) + (1-\alpha)\rho(V), \quad \forall Z, V \in \mathcal{Z}, \quad \alpha \in [0, 1]$.*

- ***Translation Equivariance****. If $c \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\rho(c + Z) = c + \rho(Z)$.*

Furthermore, a risk measure is called a ***coherent risk measure*** *if additionally it satisfies*

- ***Positive Homogeneity****. If $\gamma \geq 0$ and $Z \in \mathcal{Z}$, then $\rho(\gamma Z) = \gamma\rho(Z)$.*

A convex risk measure satisfies the following property.

Proposition 1.3.2. *If ρ is a convex risk measure, then for all $Z \in \mathcal{Z}$, $\rho(Z)$ is the certainty equivalent to Z, namely*

$$\rho(\rho(Z)) = \rho(Z).$$

Let us discuss each of the properties in Definition 1.3.1 and explain their modeling motivations. The theory of risk measure originates from the need of quantifying risk in financial instruments, and here it is illuminating to think of Z as a random variable representing the loss of a portfolio in different scenarios, where a positive number stands for a loss and a negative number stands for a gain. The word "risk" in our discussion that follows refers to a comprehensive summary of "unacceptability" a portfolio incurs, with a higher number leading to a "less acceptable" portfolio. This includes the level of uncertainty of the return of the portfolio, but also includes the average performance of the portfolio. In contrast, sometimes in the literature, the term "risk" only corresponds to the level of uncertainty.

The monotonicity axiom is straightforward and expresses the fact that if a portfolio loses less than another portfolio in all circumstances, it is a better portfolio. The convexity axiom corresponds to the benefit of diversification: by forming a convex combination of two portfolios, the risk of the newly created hybrid portfolio is never larger than the convex combination of the risks of the original two portfolios, and sometimes the risk is lower. The translation equivariance axiom stipulates that all sure losses contribute directly to the certainty equivalent. Finally, the positive homogeneity means that we cannot achieve lower risk by simply combining identical portfolios.

Let us illustrate the definition of risk measures by several examples.

Example 1.3.3 (Ogryczak and Ruszczyński [44, 45])**.** *The* ***mean-semideviation risk measure*** *is defined as*

$$\rho(Z) = \mathbb{E}[Z] + \kappa \|(Z - \mathbb{E}[Z])_+\|_p, \quad \kappa \in [0, 1],$$

and it is a coherent risk measure.

Example 1.3.4 (Rockafellar and Uryasev [51, 52])**.** *The* ***average (or conditional) value at risk*** *is defined as*

$$AVaR_\alpha(Z) = \min_{t \in \mathbb{R}} \left\{ t + \frac{1}{\alpha} \mathbb{E}[(Z - t)_+] \right\} = \frac{1}{\alpha} \int_{1-\alpha}^{1} F_Z^{-1}(\beta) d\beta, \quad \alpha \in (0, 1).$$

This risk measure is a coherent risk measure, and we shall see in the next section that it is particularly important because it acts as a basic building block for Kusuoka representation of law-invariant risk measures.

Example 1.3.5 (Föllmer and Schied [26]). *The **entropic risk measure** is defined as*

$$\rho(Z) = \frac{1}{\gamma} \ln \mathbb{E}\left[e^{\gamma Z}\right], \quad \gamma > 0.$$

It is a convex risk measure. Note that it is not a coherent risk measure as it does not satisfy the positive homogeneity axiom.

1.4 Dual Representation of Coherent Risk Measures

One of the most remarkable properties of coherent risk measures is the following *dual representation.*

Theorem 1.4.1 (Dual Representation). *Assume that $\mathcal{Z} = L^p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and define q to be the conjugate of p, i.e. $1/p + 1/q = 1$. We also define*

$$\mathcal{P}_q = \left\{ Q \in \mathcal{P}(\Omega, \mathcal{F}) \mid \int_\Omega \left|\frac{dQ}{dP}\right|^q dP < +\infty \right\}.$$

A function $\rho : \mathcal{Z} \to \mathbb{R}$ is a coherent risk measure if and only if there exists a convex closed set $\mathcal{A} \subset \mathcal{P}_q$ such that

$$\rho(Z) = \max_{\mu \in \mathcal{A}} \int_\Omega Z(\omega) \mu(d\omega). \tag{1.1}$$

This dual representation theorem is a special case of conjugate duality in convex analysis [50]. In the risk measure context, it was initially proved in the finite-dimensional case in [3] and was later refined in several papers (e.g., [27, 56]). Theorem 1.4.1 is important as it is the foundation of many numerical methods for optimizing risk measures.

It turns out that if we additionally assume that the risk measure is law-invariant, we have an even more explicit representation theorem.

Definition 1.4.2. *A coherent risk measure ρ is called **law-invariant** (w.r.t. the underlying probability P) if $Z \sim_{\text{st}} W$ implies $\rho(Z) = \rho(W)$, where $\sim_{\text{st}}$ stands for equality in distribution, i.e., $P[Z \leq \eta] = P[W \leq \eta]$ for all $\eta \in \mathbb{R}$.*

We check readily that the Average Value at Risk (AVaR) presented in Example 1.3.4 is a law-invariant coherent risk measure. The following theorem, due to Kusuoka, states that all law-invariant coherent risk measure can be represented using AVaR.

Theorem 1.4.3 (Kusuoka [39])**.** *Assume that* $(\Omega, \mathcal{F}, P)$ *is atomless and* $\mathcal{Z} = L^p(\Omega, \mathcal{F}, P)$*. Then* $\rho : \mathcal{Z} \to \mathbb{R}$ *is a law-invariant coherent risk measure if and only if there exists a convex set* $\Lambda \subset \mathcal{P}\big((0,1]\big)$ *such that*

$$\rho(Z) = \sup_{\lambda \in \Lambda} \int_{[0,1)} AVaR_\alpha(Z)\lambda(d\alpha), \quad \forall Z \in \mathcal{Z}, \tag{1.2}$$

The Kusuoka representation theorem establishes a correspondence between the set of law-invariant coherent risk measures and the set of all convex sets of $\mathcal{P}\big((0,1]\big)$. We hence obtain a very handy way to construct law-invariant coherent risk measures: it suffices to define a convex set of $\mathcal{P}\big((0,1]\big)$ and define the risk measure via (1.2). In the special case where Λ contains only one element, the corresponding law-invariant coherent risk measure is called *spectral*.

Let us revisit some of the examples presented in the previous section and examine these representation theorems in an explicit fashion.

We start with the Average Value at Risk (AVaR) presented in Example 1.3.4. We check readily that it is a law-invariant coherent risk measure. In [56] it was proved that AVaR_α has a dual representation (1.1) with

$$\mathcal{A} = \left\{ u \;\middle|\; 0 \leq \frac{d\mu}{dP} \leq \frac{1}{\alpha},\, \mu(\Omega) = 1 \right\}.$$

AVaR_α has an obvious Kusuoka representation (1.2) with $\Lambda = \{\delta_\alpha\}$, where δ stands for the Dirac mass.

We then look at the mean-semideviation risk measure presented in Example 1.3.3. Again we check readily that it is a law-invariant coherent risk measure. It was also proved in [56] that the set $\mathcal{A}$ in the dual representation theorem is given by

$$\mathcal{A} = \left\{ \mu \;\middle|\; \frac{d\mu}{dP} = 1 + h - \int_\Omega h(\omega)P(d\omega),\, \|h\|_q \leq \kappa,\, h \geq 0 \right\}.$$

Moreover, in [46, Lemma 3.6] the authors proved that the Kusuoka representation of the first order mean-semideviation risk measure ($p = 1$) has the form

$$\Lambda = \{\lambda \in \mathcal{P}([0,1)) \mid \exists \alpha \in (0,1),\, \lambda = \kappa\alpha\delta_\alpha + (1-\kappa\alpha)\delta_1\},$$

where δ stands for the Dirac mass.

Chapter 2

Risk Measures Based on Observable Processes

In this chapter we introduce fundamental concepts and properties of dynamic risk measures for uncontrolled stochastic processes in discrete time. In Section 2.1 we set up our probabilistic framework, and in Sections 2.2 and 2.3 we revisit some important concepts existing in the literature. In Section 2.4, we introduce the new notion of stochastic conditional time consistency, which is a stronger requirement on the dynamic risk measure than the standard time consistency, and which is particularly useful for controlled stochastic processes. Based on this concept, we derive the structure of dynamic risk measures involving transition risk mappings: a family of static risk measures on the space of functions of a state.

2.1 Preliminaries

In all subsequent considerations, we work with a Borel subset $\mathcal{X}$ of a Polish space (a separable and complete metric space) and the product measurable space $\big(\Omega := \mathcal{X}^T, \mathcal{F} := \mathcal{B}(\mathcal{X})^T\big)$ where T is a positive integer and $\mathcal{B}(\mathcal{X})^T$ is the product σ-algebra of Borel sets. For an element of $\mathcal{X}^T$, we use $\{X_t\}_{t=1,\dots,T}$ to denote the discrete-time process of projections of X on the coordinate spaces. For $t = 1, \dots, T$, we also define $\mathcal{H}_t = \mathcal{X}^t$ to be the space of possible histories up to time t, and we use $h_t = (x_1, \dots, x_t)$ for a generic element of $\mathcal{H}_t$: a specific history up to time t. The random vector $(X_1, \dots, X_t)$ will be denoted by H_t. In applications, $\mathcal{X}$ is usually a finite or finite-dimensional space, but we use a more general setting to allow for $\mathcal{X}$ to be the space of probability measures (for instance the space of belief states in POMDP).

We assume that for all $t = 1, \dots, T-1$, the **transition kernels**, which describe the conditional distribution of X_{t+1}, given $X_1, \cdots, X_t$, are measurable functions

$$Q_t : \mathcal{X}^t \to \mathcal{P}(\mathcal{X}), \quad t = 1, \ldots, T-1, \tag{2.1}$$

where $\mathcal{P}(\mathcal{X})$ is the set of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. These kernels, along with the initial distribution of X_1, define a unique probability measure P on the product space $\mathcal{X}^T$ with the product σ-algebra.

For a stochastic system described above, we consider a sequence of random variables $\{Z_t\}_{t=1,\ldots,T}$ taking values in $\mathbb{R}$; we assume that lower values of Z_t are preferred (e.g., Z_t represents a "cost" at time t). We require $\{Z_t\}_{t=1,\ldots,T}$ to be bounded and adapted to $\{\mathcal{F}_t\}_{t=1,\ldots,T}$ - the natural filtration generated by the process X. In order to facilitate our discussion, we introduce the following spaces:

$$\mathcal{Z}_t = \left\{ Z : \mathcal{X}^T \to \mathbb{R} \,\middle|\, Z \text{ is } \mathcal{F}_t\text{-measurable and bounded} \right\}, \quad t = 1, \ldots, T. \tag{2.2}$$

It is then equivalent to say that $Z_t \in \mathcal{Z}_t$. We also introduce the spaces

$$\mathcal{Z}_{t,T} = \mathcal{Z}_t \times \cdots \times \mathcal{Z}_T, \quad t = 1, \ldots, T,$$

representing the space of sequences of costs starting from time t.

Since Z_t is $\mathcal{F}_t$-measurable, a measurable function $\phi_t : \mathcal{X}^t \to \mathbb{R}$ exists such that $Z_t = \phi_t(X_1, \ldots, X_t)$. With a slight abuse of notation, we still use Z_t to denote this function.

2.2 Dynamic Risk Measures

In this section, we quickly review some definitions and concepts related to conditional and dynamic risk measures. All relations (e.g., equality, inequality) between random variables are understood in the "everywhere" sense.

Definition 2.2.1. *A mapping $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$, where $1 \le t \le T$, is called a* ***conditional risk measure****, if it has the* monotonicity property*: for all $(Z_t, \ldots, Z_T)$ and $(W_t, \ldots, W_T)$ in $\mathcal{Z}_{t,T}$, if $Z_s \le W_s$, for all $s = t, \ldots, T$, then $\rho_{t,T}(Z_t, \ldots, Z_T) \le \rho_{t,T}(W_t, \ldots, W_T)$.*

Definition 2.2.2. *A conditional risk measure* $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$

(i) *is **normalized** if* $\rho_{t,T}(0, \ldots, 0) = 0$*;*

(ii) *is **translation-invariant** if for all* $(Z_t, \ldots, Z_T) \in \mathcal{Z}_{t,T}$*,*

$$\rho_{t,T}(Z_t, \ldots, Z_T) = Z_t + \rho_{t,T}(0, Z_{t+1}, \ldots, Z_T).$$

Compared to the properties introduced in Definition 1.3.1 for static risk measures, the normalization property is weaker than the positive homogeneity and the translation invariance is a dynamic version of translation equivariance. Throughout the chapter, we assume all conditional risk measures to be at least normalized. Translation-invariance is a fundamental property, which will also be frequently used; under normalization, it implies that $\rho_{t,T}(Z_t, 0, \ldots, 0) = Z_t$.

Definition 2.2.3. *A conditional risk measure* $\rho_{t,T}$ *has the **local property** if*

$$\mathbb{1}_A \rho_{t,T}(Z_t, \ldots, Z_T) = \rho_{t,T}(\mathbb{1}_A Z_t, \ldots, \mathbb{1}_A Z_T),$$

for all $(Z_t, \ldots, Z_T) \in \mathcal{Z}_{t,T}$ *and for all events* $A \in \mathcal{F}_t$*.*

The local property means that the conditional risk measure at time t restricted to any $\mathcal{F}_t$-event A is not influenced by the values that $Z_t, \ldots, Z_T$ take on A^c.

Definition 2.2.4. *A **dynamic risk measure*** $\rho = \{\rho_{t,T}\}_{t=1,\ldots,T}$ *is a sequence of conditional risk measures* $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{Z}_t$*. We say that* ρ *is normalized, translation-invariant, or has the local property, if all* $\rho_{t,T}$*,* $t = 1, \ldots, T$*, satisfy the respective conditions of Definitions 2.2.2 or 2.2.3.*

Note that the above definitions and time consistency in the following section can be introduced on a underlying probability space $(\Omega, \mathcal{F}, P)$ equipped with filtrations $\{\mathcal{F}_t\}_{t=1,\ldots,T}$, and cost spaces $\mathcal{Z}_t = \mathcal{L}_p(\Omega, \mathcal{F}_t, P)$, where Ω is not necessarily $\mathcal{X}^T$. Here we choose the natural model by starting from the process X and its transition kernels for two reasons:

- In this work we really want to focus on the history H_t that can happen.
- It is easy to extend this setup to the controlled case (Chapter 3) by allowing various distributions P for $(\mathcal{X}^T, \mathcal{B}(\mathcal{X})^T)$.

2.3 Time Consistency

The notion of time consistency can be formulated in different ways, with weaker or stronger assumptions; but the key idea is that if one sequence of costs, compared to another sequence, has the same current cost and lower measure in the future, then it should have lower current measure. In this and the next section, we discuss two formulations of time consistency: the (now) standard one, and our new proposal specially suited for process-based measures. We also show how the tower property (the recursive relation between $\rho_{t,T}$ and $\rho_{t+1,T}$ that the time consistency implies) improves with the more refined time consistency concept. The following definition of time consistency was employed in [57].

Definition 2.3.1. *A dynamic risk measure* $\{\rho_{t,T}\}_{t=1,\dots,T}$ *is* ***time-consistent*** *if for any* $1 \le t < T$ *and for all* $(Z_t, \dots, Z_T), (W_t, \dots, W_T) \in \mathcal{Z}_t$, *the conditions*

$$\begin{cases} Z_t = W_t, \\ \rho_{t+1,T}(Z_{t+1}, \dots, Z_T) \le \rho_{t+1,T}(W_{t+1}, \dots, W_T), \end{cases}$$

imply that $\rho_{t,T}(Z_t, \dots, Z_T) \le \rho_{t,T}(W_t, \dots, W_T)$.

In the extant literature, time consistency is usually defined for mappings $\rho_{t,T} : \mathcal{Z}_T \to \mathcal{Z}_t$, which evaluate at time t the risk of a final cost Z_T. In [4] it is required that $\rho_{t,T}(Z_T) = \rho_{t,T}\big(\rho_{s,T}(Z_T)\big)$ for all $1 \le t \le s \le T$. Our definition, restricted to sequences of form $(0, \dots, 0, Z_T)$, is equivalent to it, which can be easily verified by induction. In [60, Def. 6.79] a strong version of Definition 2.3.1 is introduced (with sharp inequalities).

It turns out that a translation-invariant and time-consistent dynamic risk measure can be decomposed into and then reconstructed from so-called *one-step conditional risk mappings.*

Theorem 2.3.2 ([57])**.** *A dynamic risk measure* $\{\rho_{t,T}\}_{t=1,\dots,T}$ *is translation-invariant and time-consistent if and only if there exist mappings* $\rho_t : \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, $t = 1, \dots, T-1$, *satisfying the monotonicity and normalization properties, called* ***one-step conditional***

***risk mappings**, such that for all $t = 1, \dots, T-1$,*

$$\rho_{t,T}(Z_t, \dots, Z_T) = Z_t + \rho_t\big(\rho_{t+1,T}(Z_{t+1}, \dots, Z_T)\big). \tag{2.3}$$

This relation is related to the *Koopmans equation* [36] for utility functions. The operators $A(Z_t, Z_{t+1}) = Z_t + \rho_t(Z_{t+1})$ generalize the concept of *aggregator* to measures of risk.

In general, time consistency does not imply the local property, unless additional conditions are satisfied.

Conceptually, the one-step conditional risk mappings play a similar role to one-step conditional expectations, and will be very useful when an analog of the tower property is involved. At this stage, without further refinement of assumptions, it remains a fairly abstract and general object that is hard to characterize. In [33], for the case of the expected utility model, ρ_t was a conditional expectation of a pointwise monotonic transformation of its argument. In [57], a more general, but seemingly special form of this one-step conditional risk mappings was imposed, which was well suited for Markovian applications, but it was unclear whether other forms of such mappings exist. In order to gain deeper understanding of these concepts, we introduce a stronger notion of time consistency, and we argue that any one-step conditional risk mapping is of the form postulated in [57]. To this end, we use the particular structure of the space $\big(\mathcal{X}^T, \mathcal{B}(\mathcal{X})^T\big)$ and the way a probability measure is defined on this space.

2.4 Stochastic Conditional Time Consistency and Transition Risk Mappings

We now refine the concept of time consistency for process-based risk measures.

Definition 2.4.1. *A dynamic risk measure $\big\{\rho_{t,T}\big\}_{t=1,\dots,T}$ is **stochastically conditionally time-consistent** with respect to $\{Q_t\}_{t=1,\dots,T-1}$ if for any $1 \le t \le T-1$, for any $h_t \in \mathcal{X}^t$, and for all $(Z_t, \dots, Z_T)$, $(W_t, \dots, W_T) \in \mathcal{Z}_{t,T}$, the conditions*

$$\begin{cases} Z_t(h_t) = W_t(h_t), \\ \big(\rho_{t+1,T}(Z_{t+1}, \dots, Z_T) \mid H_t = h_t\big) \preceq_{\text{st}} \big(\rho_{t+1,T}(W_{t+1}, \dots, W_T) \mid H_t = h_t\big), \end{cases} \tag{2.4}$$

imply

$$\rho_{t,T}(Z_t,\dots,Z_T)(h_t) \le \rho_{t,T}(W_t,\dots,W_T)(h_t), \tag{2.5}$$

where the relation $\preceq_{\text{st}}$ *is the conditional stochastic order understood as follows:*

$$Q_t(h_t)\Big(\big\{\, x \,|\, \rho_{t+1,T}(Z_{t+1},\dots,Z_T)(h_t,x) > \eta \,\big\}\Big) \\ \le Q_t(h_t)\Big(\big\{\, x \,|\, \rho_{t+1,T}(W_{t+1},\dots,W_T)(h_t,x) > \eta \,\big\}\Big), \quad \forall\, \eta \in \mathbb{R}.$$

When the choice of the underlying transition kernels is clear from the context, we will simply say that the dynamic risk measure is stochastically conditionally time-consistent.

Proposition 2.4.2. *If a dynamic risk measure* $\big\{\rho_{t,T}\big\}_{t=1,\dots,T}$ *is stochastically conditionally time-consistent and has the translation property, then it is time-consistent and has the local property.*

Proof. Suppose $\big\{\rho_{t,T}\big\}_{t=1,\dots,T}$ is stochastically conditionally time-consistent. We verify Definition 2.3.1. If $Z_t = W_t$ and $\rho_{t+1,T}(Z_{t+1},\dots,Z_T) \le \rho_{t+1,T}(W_{t+1},\dots,W_T)$ (pointwise) then (2.4) is true, for all h_t. Then Definition 2.4.1 implies that (2.5) is true, and thus $\big\{\rho_{t,T}\big\}_{t=1,\dots,T}$ is time-consistent.

Let us prove by induction on t from T down to 1 that $\rho_{t,T}$ have the local property. Clearly, $\rho_{T,T}$ does: if $A \in \mathcal{F}_T$, then Definition 2.2.2 yields

$$\mathbb{1}_A \rho_{T,T}(Z_T) = \mathbb{1}_A Z_T = \rho_{T,T}(\mathbb{1}_A Z_T).$$

Suppose $\rho_{t+1,T}$ satisfies the local property for some $1 \le t < T$, and consider any $A \in \mathcal{F}_t$, any $h_t \in \mathcal{X}^t$, and any $(Z_t,\dots,Z_T) \in \mathcal{Z}_{t,T}$. Two cases may occur.

- If $\mathbb{1}_A(h_t) = 0$, then $[\mathbb{1}_A Z_t](h_t) = 0$. The local property for $t+1$ yields:

  $$\big[\rho_{t+1,T}(\mathbb{1}_A Z_{t+1},\dots,\mathbb{1}_A Z_T)\big](h_t,\cdot) = \big[\mathbb{1}_A \rho_{t+1,T}(Z_{t+1},\dots,Z_T)\big](h_t,\cdot) = 0.$$

  By stochastic conditional time consistency,

  $$\rho_{t,T}(\mathbb{1}_A Z_t, \mathbb{1}_A Z_{t+1},\dots,\mathbb{1}_A Z_T)(h_t) = \rho_{t,T}(0,\dots,0)(h_t) = 0.$$

- If $\mathbb{1}_A(h_t) = 1$, then $[\mathbb{1}_A Z_t](h_t) = Z_t(h_t)$. The local property for $t+1$ implies that

$$\begin{aligned} & \big[\rho_{t+1,T}(\mathbb{1}_A Z_{t+1}, \dots, \mathbb{1}_A Z_T)\big](h_t, \cdot) \\ = \; & \big[\mathbb{1}_A \rho_{t+1,T}(Z_{t+1}, \dots, Z_T)\big](h_t, \cdot) \\ = \; & \rho_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, \cdot). \end{aligned}$$

  By stochastic conditional time consistency,

$$\rho_{t,T}(\mathbb{1}_A Z_t, \dots, \mathbb{1}_A Z_T)(h_t) = \rho_{t,T}(Z_t, \dots, Z_T)(h_t).$$

In both cases, $\rho_{t,T}(\mathbb{1}_A Z_t, \dots, \mathbb{1}_A Z_T)(h_t) = \big[\mathbb{1}_A \rho_{t,T}(Z_t, \dots, Z_T)\big](h_t)$. $\square$

The following examples illustrate the differences between the concepts of time consistency and stochastic conditional time consistency.

Example 2.4.3. *Suppose $\mathcal{X} = \{0, 1\}$, $T = 2$, $Q(x) = \{1/2, 1/2\}$ for both $x \in \mathcal{X}$. A random element Z_1 is a function of x_1, while Z_2 is a function of the pair (x_1, x_2). Consider the risk measure*

$$\begin{aligned} &\rho_{1,2}(Z_1, Z_2)(x_1) = Z_1(x_1) + Z_2(x_1, x_1), \\ &\rho_{2,2}(Z_2)(x_1, x_2) = Z_2(x_1, x_2). \end{aligned} \tag{2.6}$$

It is time-consistent: if $Z_1 = W_1$ and $Z_2 \le W_2$ then also $\rho_{1,2}(Z_1, Z_2) \le \rho_{1,2}(W_1, W_2)$. It also has the normalization, translation, and local properties. Let

$$\begin{cases} Z_1(x_1) = 0, \\ Z_2(x_1, x_2) = x_2; \end{cases}$$
$$\begin{cases} W_1(x_1) = 0, \\ W_2(x_1, x_2) = 1 - x_2. \end{cases}$$

The conditional distributions of Z_2 and W_2 are identical, and Definition 2.4.1 requires that $\rho_{1,2}(Z_1, Z_2) = \rho_{1,2}(W_1, W_2)$. But (2.6) yields $\rho_{1,2}(Z_1, Z_2)(x_1) = x_1$ and $\rho_{1,2}(W_1, W_2)(x_1) = 1 - x_1$. The measure (2.6) is not stochastically conditionally time-consistent. The reason is that it does not include the distribution of the next state in the risk evaluation. $\square$

Generally, the definition of dynamic risk measures in Section 2.2 and the definition of time consistency property in Section 2.3 are valid with any filtration on an underlying probability space $(\Omega, \mathcal{F}, P)$, instead of process-generated filtration. However, from now on, we only consider dynamic risk measures defined with a filtration generated by a *specific* process, because we are interested in those risk measures which can be evaluated on each specific history path. That is why we call these risk measures "process-based."

The following proposition shows that the stochastic conditional time consistency implies that the one-step risk mappings ρ_t can be equivalently represented by static law-invariant risk measures on $\mathcal{V}$, where

$$\mathcal{V} = \left\{ v : \mathcal{X} \to \mathbb{R} \,\middle|\, v \text{ is measurable and bounded} \right\}. \tag{2.7}$$

Let us first restate Definition 1.4.2 on law invariance and make the dependency on the underlying probability measure explicit.

Definition 2.4.4. *A measurable function $r : \mathcal{V} \to \mathbb{R}$ is* ***law-invariant with respect to the probability measure*** *q on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, if for all $v, w \in \mathcal{V}$*

$$\Big(v|q \sim_{\text{st}} w|q \Big) \Rightarrow \Big(r(v) = r(w) \Big),$$

where $\Big(v|q \sim_{\text{st}} w|q \Big)$ means that $q\big(\{v \le \eta\}\big) = q\big(\{w \le \eta\}\big)$ for all $\eta \in \mathbb{R}$.

We can now state the main result of this section.

Theorem 2.4.5. *A process-based dynamic risk measure $\big\{\rho_{t,T}\big\}_{t=1,\dots,T}$ is translation-invariant and stochastically conditionally time-consistent if and only if functionals σ_t : $\operatorname{graph}(Q_t) \times \mathcal{V} \to \mathbb{R}$, $t = 1, \dots, T-1$, exist, such that*

(i) *for all $t = 1, \dots, T-1$ and all $h_t \in \mathcal{X}^t$, the functional $\sigma_t(h_t, Q_t(h_t), \cdot)$ is a normalized, monotonic, and law-invariant risk measure on $\mathcal{V}$ with respect to the distribution $Q_t(h_t)$;*

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

$$\rho_{t,T}(Z_t, \dots, Z_T)(h_t) = Z_t(h_t) + \sigma_t\big(h_t, Q_t(h_t), \rho_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, \cdot)\big). \tag{2.8}$$

Moreover, for all $t = 1, \dots, T-1$, σ_t is uniquely determined by $\rho_{t,T}$ as follows: for every $h_t \in \mathcal{X}^t$ and every $v \in \mathcal{V}$,

$$\sigma_t(h_t, Q_t(h_t), v) = \rho_{t,T}(0, V, 0, \dots, 0)(h_t), \tag{2.9}$$

where $V \in \mathcal{Z}_{t+1}$ satisfies the equation $V(h_t,\cdot) = v(\cdot)$, and can be arbitrary elsewhere.

Proof.

- Assume $\{\rho_{t,T}\}_{t=1,\dots,T}$ is translation-invariant and stochastically conditionally time-consistent. We shall prove the existence of σ_t satisfying (2.8)–(2.9).

  Formula (2.9) defines a normalized and monotonic risk measure on the space $\mathcal{V}$. Define, for a fixed $h_t \in \mathcal{X}^t$,

$$v(x) = \rho_{t+1,T}(Z_{t+1},\dots,Z_T)(h_t,x), \quad \forall\, x \in \mathcal{X},$$

$$V(h_{t+1}) = \begin{cases} v(x), & \text{if } h_{t+1} = (h_t,x), \\ 0, & \text{otherwise.} \end{cases}$$

  By translation invariance and normalization,

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

  Thus, by the translation property and stochastic conditional time consistency,

$$\begin{aligned} \rho_{t,T}(Z_t,\dots,Z_T)(h_t) &= Z_t(h_t) + \rho_{t,T}(0,Z_{t+1},\dots,Z_T)(h_t) \\ &= Z_t(h_t) + \rho_{t,T}(0,V,0,\dots,0)(h_t) \\ &= Z_t(h_t) + \sigma_t(h_t,Q_t(h_t),v). \end{aligned}$$

  This chain of relations proves also the uniqueness of σ_t.

  We need only verify the postulated law invariance of $\sigma_t(h_t,Q_t(h_t),\cdot)$. If $V,V' \in \mathcal{Z}_{t+1}$ have the same conditional distribution, given h_t, then Definition 2.4.1 implies that $\rho_{t,T}(0,V,0,\dots,0)(h_t) = \rho_{t,T}(0,V',0,\dots,0)(h_t)$, and law invariance follows from (2.9).

- On the other hand, if such transition risk mappings exist, then $\{\rho_{t,T}\}_{t=1,\dots,T}$ is stochastically conditionally time-consistent by the monotonicity and law invariance of $\sigma(h_t,\cdot)$. In order to show the translation invariance of $\rho_{t,T}$, we can use (2.8) to obtain for any $t = 1,\dots,T-1$, and for all $h_t \in \mathcal{X}^t$ the following identity:

$$\begin{aligned} \rho_{t,T}(0,Z_{t+1},\dots,Z_T)(h_t) &= \sigma_t\big(h_t,Q_t(h_t),\rho_{t+1,T}(Z_{t+1},\dots,Z_T)(h_t,\cdot)\big) \\ &= \rho_{t,T}(Z_t,\dots,Z_T)(h_t) - Z_t(h_t). \end{aligned}$$

$\square$

Remark 2.4.6. *With a slight abuse of notation, we included the distribution $Q_t(h_t)$ as an argument of the transition risk mapping in view of the application to controlled processes.*

In the following examples, we apply common static risk measures (Examples 1.3.3-1.3.5) to σ_t to construct dynamic risk measures.

Example 2.4.7. *In the theory of risk-sensitive Markov decision processes, the following family of* entropic risk measures *is employed (see [29, 43, 18, 54, 10, 17, 20, 40, 7]):*

$$\rho_{t,T}(Z_t, \dots, Z_T) = \frac{1}{\gamma} \ln \left(\mathbb{E}\Big[\exp\Big(\gamma {\textstyle\sum_{s=t}^T} Z_s\Big) \mid \mathcal{F}_t \Big] \right), \quad t = 1, \dots, T, \quad \gamma > 0.$$

It is stochastically conditionally time-consistent, and corresponds to the transition risk mapping

$$\sigma_t(h_t, q, v) = \frac{1}{\gamma} \ln \left(\mathbb{E}_q[e^{\gamma v}] \right) = \frac{1}{\gamma} \ln \left(\int_{\mathcal{X}} e^{\gamma v(x)} \, q(dx) \right), \quad \gamma > 0. \tag{2.10}$$

In the construction of a dynamic risk measure, we use $q = Q_t(h_t)$. We could also make γ in (2.10) dependent on the time t, the current state x_t, or even the entire history h_t, and still obtain a stochastically conditionally time-consistent dynamic risk measure. If γ depends on t and x_t only, the mapping (2.10) corresponds to a Markov risk measure discussed in Sections 4.1 and 4.2. □

Example 2.4.8. *The following transition risk mapping satisfies the condition of Theorem 2.4.5 and corresponds to a stochastically conditionally time-consistent dynamic risk measure:*

$$\sigma_t(h_t, q, v) = \int_{\mathcal{X}} v(s) \, q(ds) + \varkappa_t(h_t) \left(\int_{\mathcal{X}} \left[\left(v(s) - \int_{\mathcal{X}} v(s') \, q(ds') \right)_+ \right]^p q(ds) \right)^{1/p}, \tag{2.11}$$

where $\varkappa_t : \mathcal{X}^t \to [0, 1]$ is a measurable function, and $p \in [1, +\infty)$. It is an analogue of the static mean–semideviation measure of risk (Example 1.3.3), whose consistency with stochastic dominance is well–known [44, 45]. In the construction of a dynamic risk measure, we use $q = Q_t(h_t)$. If $\varkappa_t$ depends on x_t only, the mapping (2.11) corresponds to a Markov risk measure (see Sections 4.1 and 4.2). □

Example 2.4.9. *The following transition risk mapping is derived from the Average Value at Risk [51]:*

$$\sigma_t(h_t, q, v) = \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha_t(h_t)} \int_{\mathcal{X}} \big(v(s) - \eta \big)_+ \, q(ds) \right\}, \tag{2.12}$$

where $\alpha_t(h_t)$ *is a measurable function with values in* $[\alpha_{\min}, \alpha_{\max}] \subset (0,1)$. *The mapping* (2.12) *satisfies the condition of Theorem 2.4.5; its consistency with stochastic dominance is well–known [46].* □

Our use of the stochastic dominance relation in the definition of stochastic conditional time consistency rules out some candidates for transition risk mappings.

Example 2.4.10. *Suppose* $\sigma_t(h_t, q, v) = v(x_1)$, *where* $x_1 \in \mathcal{X}$ *is a selected state. Such a mapping is a coherent measure of risk, as a function of the last argument, and may be law-invariant. In particular, it is law-invariant with* $\mathcal{X} = \{x_1, x_2\}$, $q(x_1) = 1/3$, $q(x_2) = 2/3$, $v(x_1) = 3$, $v(x_2) = 1$, $w(x_1) = 2$, $w(x_2) = 4$. *For this mapping, we have* $v \preceq_{\text{st}} w$ *under* q, *but* $\sigma_t(h_t, q, v) > \sigma_t(h_t, q, w)$, *and thus the condition of stochastic conditional time consistency is violated. This is due to the fact that the probability of reaching* x_1, *no matter how small, does not affect the value of the risk measure. We consciously exclude such cases, because in controlled systems, to be discussed in the next section, the second argument* (q) *is the only one that depends on our decisions. It should be included in the definition of our preferences, if practically meaningful results are to be obtained.* □

Chapter 3

Risk Measures for Controlled Stochastic Processes

We now extend the setting of Chapter 2 by allowing the transition kernels (2.1) to depend on control variables u_t.

3.1 The Model

We still work with the process $\{X_t\}_{t=1,\dots,T}$ on the space $\mathcal{X}^T$ and introduce a control space $\mathcal{U}$, which is assumed to be a Borel subset of a Polish space. At each time t, we observe the state x_t and then apply a control $u_t \in \mathcal{U}$. We assume that the admissible control sets and the transition kernels (conditional distributions of the next state) depend on all currently-known state and control values. More precisely, we make the following assumptions:

1. For all $t = 1, \dots, T$, we require that $u_t \in \mathcal{U}_t(x_1, u_1, \dots, x_{t-1}, u_{t-1}, x_t)$, where $\mathcal{U}_t : \mathcal{G}_t \rightrightarrows \mathcal{U}$ is a measurable multifunction, and $\mathcal{G}_1, \dots, \mathcal{G}_T$ are the sets of histories of all currently-known state and control values before applying each control:
$$\begin{cases} \mathcal{G}_1 = \mathcal{X}, \\ \mathcal{G}_{t+1} = \operatorname{graph}(\mathcal{U}_t) \times \mathcal{X} \subseteq (\mathcal{X} \times \mathcal{U})^t \times \mathcal{X}, \quad t = 1, \dots, T-1; \end{cases}$$
Here $\operatorname{graph}(\mathcal{U}_t) = \big\{(x_1, u_1, \dots, x_t, u_t) \in (\mathcal{X} \times \mathcal{U})^t : u_t \in \mathcal{U}_t(x_1, u_1, \dots, x_t)\big\}$.

2. For all $t = 1, \dots, T$, the control-dependent transition kernels
$$Q_t : \operatorname{graph}(\mathcal{U}_t) \to \mathcal{P}(\mathcal{X}), \quad t = 1, \dots, T-1, \tag{3.1}$$
are measurable, and for all $t = 1, \dots, T-1$, for all $(x_1, u_1, \dots, x_t, u_t) \in \operatorname{graph}(\mathcal{U}_t)$, $Q_t(x_1, u_1, \dots, x_t, u_t)$ describes the conditional distribution of X_{t+1}, given currently-known states and controls.

For this controlled process, a (deterministic) *history-dependent admissible policy* $\pi = (\pi_1, \dots, \pi_T)$ is a sequence of measurable selectors, called *decision rules*, $\pi_t : \mathcal{G}_t \to \mathcal{U}$ such that $\pi_t(g_t) \in \mathcal{U}_t(g_t)$ for all $g_t \in \mathcal{G}_t$. We can easily prove by induction on t that for an admissible policy π each π_t reduces to a measurable function on $\mathcal{X}^t$, as $u_s = \pi_s(h_s)$ for all $s = 1, \dots, t-1$. We are still using π_s to denote the decision rule, although it is a different function, formally; it will not lead to any misunderstanding. The set of admissible policies is

$$\begin{aligned} \mathit{\Pi} := \big\{ \, &\pi = (\pi_1, \dots, \pi_T) \mid \\ &\forall t,\ \pi_t(x_1, \dots, x_t) \in \mathcal{U}_t(x_1, \pi_1(x_1), \dots, x_{t-1}, \pi_{t-1}(x_1, \dots, x_{t-1}), x_t) \, \big\}. \end{aligned} \tag{3.2}$$

For any fixed policy $\pi \in \mathit{\Pi}$, the transition kernels can be rewritten as measurable functions from $\mathcal{X}^t$ to $\mathcal{P}(\mathcal{X})$:

$$Q_t^\pi : (x_1, \dots, x_t) \mapsto Q_t\big(x_1, \pi_1(x_1), \dots, x_t, \pi_t(x_1, \dots, x_t)\big),\ t = 1, \dots, T-1, \tag{3.3}$$

just like the transition kernels of the uncontrolled case given in (2.1), but indexed by π. Thus, for any policy $\pi \in \mathit{\Pi}$, we can consider $\{X_t\}_{t=1,\dots,T}$ as an "uncontrolled" process, on the probability space $\big(\mathcal{X}^T, \mathcal{B}(\mathcal{X})^T, P^\pi\big)$ with P^π defined by $\{Q_t^\pi\}_{t=1,\dots,T-1}$. The process $\{X_t\}$ is adapted to the policy-independent filtration $\{\mathcal{F}_t\}_{t=1,\dots,T}$. As before and throughout this chapter, $h_t \in \mathcal{X}^t$ stands for $(x_1, \dots, x_t)$.

We still use the same spaces $\mathcal{Z}_t$, $t = 1, \dots, T$, as defined in (2.2) for the costs incurred at each stage; these spaces also allow us to consider control-dependent costs as collections of policy-indexed costs in $\mathcal{Z}_{1,T}$. Thus, we are able to define and analyze (time-consistent) dynamic risk measures ρ^π for each fixed $\pi \in \mathit{\Pi}$, as in Chapter 2. Note that ρ^π are defined on the same spaces independently of π, because the filtration and the spaces $\mathcal{Z}_t$, $t = 1, \dots, T$, are not dependent on π; however, we do need to index the measures of risk by the policy π, because the transition kernels and, consequently, the probability measure on the space $\mathcal{X}^T$, depend on π.

3.2 Stochastic Conditional Time Consistency and Transition Risk Mappings

We need to compare risk levels among different policies, so a meaningful order among the risk measures ρ^π, with $\pi \in \Pi$, is needed. It turns out that our concept of stochastic conditional time consistency (Definition 2.4.1) can be extended to this setting.

Definition 3.2.1. *A family of process-based dynamic risk measures* $\big\{\rho^\pi_{t,T}\big\}_{t=1,\dots,T-1}^{\pi\in\Pi}$ *is **stochastically conditionally time-consistent** if for any* $\pi, \pi' \in \Pi$*, for any* $1 \le t < T$*, for all* $h_t \in \mathcal{X}^t$*, all* $(Z_t, \dots, Z_T) \in \mathcal{Z}_{t,T}$ *and all* $(W_t, \dots, W_T) \in \mathcal{Z}_{t,T}$*, the conditions*

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

imply

$$\rho^\pi_{t,T}(Z_t, \dots, Z_T)(h_t) \le \rho^{\pi'}_{t,T}(W_t, \dots, W_T)(h_t).$$

Remark 3.2.2. *As in Definition 2.4.1, the conditional stochastic order "*$\preceq_{\text{st}}$*" is understood as follows: for all* $\eta \in \mathbb{R}$ *we have*

$$\begin{aligned} Q^\pi_t(h_t)\Big(\big\{\, x \,|\, \rho^\pi_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, x) > \eta \,\big\}\Big) \\ \le Q^{\pi'}_t(h_t)\Big(\big\{\, x \,|\, \rho^{\pi'}_{t+1,T}(W_{t+1}, \dots, W_T)(h_t, x) > \eta \,\big\}\Big). \end{aligned}$$

This definition helps us build a connection among dynamic risk measures ρ^π, for $\pi \in \Pi$, as we explain it below. Before passing to the details, we can say in short that the same transition risk mappings as in the uncontrolled case are the only possible structures of such risk measures.

If a family of process-based dynamic risk measures $\big\{\rho^\pi_{t,T}\big\}_{t=1,\dots,T-1}^{\pi\in\Pi}$ is stochastically conditionally time-consistent, then for each fixed $\pi \in \Pi$ the process-based dynamic risk measure $\big\{\rho^\pi_{t,T}\big\}_{t=1,\dots,T-1}$ is stochastically conditionally time-consistent, as defined in Definition 2.4.1. By virtue of Proposition 2.4.5, for each $\pi \in \Pi$, there exist functionals $\sigma^\pi_t : \text{graph}(Q^\pi_t) \times \mathcal{V} \to \mathbb{R}$, $t = 1 \dots T-1$, such that for all $t = 1, \dots, T-1$, all $h_t \in \mathcal{X}^t$, the functional $\sigma^\pi_t(h_t, Q^\pi_t(h_t), \cdot\,)$ is a law-invariant risk measure on $\mathcal{V}$ with respect to the distribution $Q^\pi_t(h_t)$ and

$$\rho^\pi_{t,T}(Z_t, \dots, Z_T)(h_t) = Z_t(h_t) + \sigma^\pi_t\big(h_t, Q^\pi_t(h_t), \rho^\pi_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, \cdot)\big), \quad \forall h_t \in \mathcal{X}^t.$$

Consider any $\pi, \pi' \in \Pi$, $h_t \in \mathcal{X}^t$, and $(Z_t, \dots, Z_T) \in \mathcal{Z}_{t,T}$, $(W_t, \dots, W_T) \in \mathcal{Z}_{t,T}$ such that

$$\begin{cases} Z_t(h_t) = W_t(h_t), \\ Q_t^{\pi}(h_t) = Q_t^{\pi'}(h_t), \\ \rho^{\pi}_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, \cdot) = \rho^{\pi'}_{t+1,T}(W_{t+1}, \dots, W_T)(h_t, \cdot). \end{cases}$$

Then we have

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

where the relation $\sim_{\text{st}}$ means that both $\preceq_{\text{st}}$ and $\succeq_{\text{st}}$ are true; in other words, equality in law. Because of the stochastic conditional time consistency,

$$\rho^{\pi}_{t,T}(Z_t, \dots, Z_T)(h_t) = \rho^{\pi'}_{t,T}(W_t, \dots, W_T)(h_t),$$

whence

$$\sigma_t^{\pi}\big(h_t, Q_t^{\pi}(h_t), \rho^{\pi}_{t+1,T}(Z_{t+1}, \dots, Z_T)(h_t, \cdot)\big) \\ = \sigma_t^{\pi'}\big(h_t, Q_t^{\pi'}(h_t), \rho^{\pi'}_{t+1,T}(W_{t+1}, \dots, W_T)(h_t, \cdot)\big).$$

All three arguments of σ_t^{π} and $\sigma_t^{\pi'}$ are identical. Consequently, σ^{π} does not depend on π directly, and all dependence on π is carried by the controlled kernel Q_t^{π}. This is a highly desirable property, when we apply dynamic risk measures to a control problem. We summarize this important observation in the following theorem, which extends Theorem 2.4.5 to the case of controlled processes.

Theorem 3.2.3. *A family of process-based dynamic risk measures* $\big\{\rho^{\pi}_{t,T}\big\}_{t=1,\dots,T}^{\pi\in\Pi}$ *is translation-invariant and stochastically conditionally time-consistent if and only if there exist functionals*

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

such that:

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

$$\forall q^1, q^2 \in \big\{\, Q_t^{\pi}(h_t) : \pi \in \Pi \,\big\},\ \forall v^1, v^2 \in \mathcal{V}, \\ \big(v^1 \mid q^1\big) \preceq_{st} \big(v^2 \mid q^2\big) \implies \sigma_t(h_t, q^1, v^1) \le \sigma_t(h_t, q^2, v^2),$$

where $\big(v^1 \,|\, q^1\big) = q \circ v^{-1}$ *means "the distribution of* v *under* q*;"*

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

$$\rho^{\pi}_{t,T}(Z_t, \ldots, Z_T)(h_t) = Z_t(h_t) + \sigma_t\big(h_t, Q^{\pi}_t(h_t), \rho^{\pi}_{t+1,T}(Z_{t+1}, \ldots, Z_T)(h_t, \cdot)\big). \tag{3.4}$$

Moreover, for all $t = 1, \ldots, T-1$*,* σ_t *is uniquely determined by* $\rho_{t,T}$ *as follows: for every* $h_t \in \mathcal{X}^t$*, for every* $q \in \big\{\, Q^{\pi}_t(h_t) : \pi \in \Pi \,\big\}$*, and for every* $v \in \mathcal{V}$*,*

$$\sigma_t(h_t, q, v) = \rho^{\pi}_{t,T}(0, V, 0, \ldots, 0)(h_t), \tag{3.5}$$

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

Proof. We have shown the existence of $\{\sigma_t\}_{t=1,\ldots,T}$ satisfying (3.4) and (3.5) in the discussion preceding the theorem. We can verify the strong monotonicity with respect to stochastic dominance by (3.5) and Definition 3.2.1. □

It follows that the transition risk mappings of Examples 2.4.7, 2.4.8, and 2.4.9 are perfectly suitable transition risk mappings for controlled processes as well, provided that the corresponding parameters (γ, $\varkappa$, and α) depend on t and x_t only.

Chapter 4

Application to Controlled Markov Systems

Our results can be further specialized to the case when $\{X_t\}$ is a controlled Markov system, in which we assume the following conditions:

- The admissible control sets are measurable multifunctions of the current state, i.e., $\mathcal{U}_t : \mathcal{X} \rightrightarrows \mathcal{U}, \quad t = 1, \dots, T$;
- The dependence in the transition kernel (3.1) on the history is carried only through the last state and control: $Q_t : \text{graph}(\mathcal{U}_t) \to \mathcal{P}(\mathcal{X})$, $t = 1, \dots, T-1$;
- The step-wise costs are dependent only on the current state and control: $Z_t = c_t(x_t, u_t)$, $t = 1, \dots, T$, where $c_t : \text{graph}(\mathcal{U}_t) \to \mathbb{R}$, $t = 1, \dots, T$ are measurable bounded functions.

Let $\mathit{\Pi}$ be the set of admissible *history-dependent policies*:

$$\mathit{\Pi} := \big\{ \pi = (\pi_1, \dots, \pi_T) \,|\, \forall t,\ \pi_t(x_1, \dots, x_t) \in \mathcal{U}_t(x_t) \big\}.$$

To alleviate notation, for all $\pi \in \mathit{\Pi}$ and for all measurable $c = (c_1, \dots, c_T)$, we write

$$v_t^{c,\pi}(h_t) := \rho_{t,T}^{\pi}\big(c_t(X_t, \pi_t(H_t)), \dots, c_T(X_T, \pi_T(H_T))\big)(h_t).$$

The following result is a direct consequence of Theorem 3.2.3 in the Markovian case.

Corollary 4.0.1. *For a controlled Markov system, a family of process-based dynamic risk measures* $\big\{\rho_{t,T}^{\pi}\big\}_{t=1,\dots,T}^{\pi\in\mathit{\Pi}}$ *is translation-invariant and stochastically conditionally time-consistent if and only if functionals*

$$\sigma_t : \big\{\big(h_t, Q_t(x_t, u)\big) : h_t \in \mathcal{X}^t, u \in \mathcal{U}_t(x_t)\big\} \times \mathcal{V} \to \mathbb{R}, \quad t = 1 \dots T-1,$$

exist, such that

(i) *For all* $t = 1, \ldots, T-1$ *and all* $h_t \in \mathcal{X}^t$, $\sigma_t(h_t, \cdot, \cdot)$ *is normalized and strongly monotonic with respect to stochastic dominance on* $\{ Q_t(x_t, u) : u \in \mathcal{U}_t(x_t) \}$;

(ii) *For all* $\pi \in \Pi$, *for all bounded measurable* $c = (c_1, \ldots, c_T)$, *for all* $t = 1, \ldots, T-1$, *and for all* $h_t \in \mathcal{X}^t$,

$$v_t^{c,\pi}(h_t) = c_t(x_t, \pi_t(h_t)) + \sigma_t\Big(h_t, Q_t(x_t, \pi_t(h_t)), v_{t+1}^{c,\pi}(h_t, \cdot)\Big). \tag{4.1}$$

Proof. To verify the "if and only if" statement, we can show that (3.5) is true if σ_t satisfies (4.1) for all measurable bounded c. $\square$

4.1 Markov Risk Measures

Consider a Markov policy π composed of state-dependent measurable decision rules $\pi_t : \mathcal{X} \mapsto \mathcal{U}$, $t = 1, \ldots, T$. Because of the Markov property of the transition kernels, for a Markov policy π, the future evolution of the process $\{X_\tau\}_{\tau=t,\ldots,T}$ is solely dependent on the current state x_t, so is the distribution of the future costs $c_\tau(X_\tau, \pi_\tau(X_\tau))$, $\tau = t, \ldots, T$. Therefore, it is reasonable to assume that the dependence of the conditional risk measure on the history is also carried by the current state only.

Definition 4.1.1. *A family of process-based dynamic risk measures* $\{\rho_{t,T}^\pi\}_{t=1,\ldots,T}^{\pi\in\Pi}$ *for a controlled Markov system is* ***Markov*** *if for all Markov policies* $\pi \in \Pi$, *for all measurable* $c = (c_1, \ldots, c_T)$, *and for all* $h_t = (x_1, \ldots, x_t)$ *and* $h'_t = (x'_1, \ldots, x'_t)$ *in* $\mathcal{X}^t$ *such that* $x_t = x'_t$, *we have* $v_t^{c,\pi}(h_t) = v_t^{c,\pi}(h'_t)$.

Proposition 4.1.2. *Under translation invariance and stochastic conditional time consistency,* $\{\rho_{t,T}^\pi\}_{t=1,\ldots,T}^{\pi\in\Pi}$ *is Markov if and only if the dependence of* σ_t *on* h_t *is carried only by* x_t, *for all* $t = 1, \ldots, T-1$.

Proof. Suppose $\{\rho_{t,T}^\pi\}_{t=1,\ldots,T}^{\pi\in\Pi}$ is Markov. For all $t = 1, \ldots, T-1$, for all $h_t, h'_t \in \mathcal{X}^t$ such that $x_t = x'_t$, for all $u \in \mathcal{U}_t(x_t)$ and for all $v \in \mathcal{V}$, there exists a Markov $\pi \in \Pi$ such that $\pi_t(x_t) = u$. By setting $c = (0, \ldots, 0, c_{t+1}, 0, \ldots, 0)$ with $c_{t+1} : (x', u') \mapsto v(x')$, the Markov property of ρ^π implies that

$$\sigma_t(h_t, Q_t(x_t, u), v) = v_t^{c,\pi}(h_t) = v_t^{c,\pi}(h'_t) = \sigma_t(h'_t, Q_t(x_t, u), v).$$

Therefore, σ_t is indeed memoryless, that is, its dependence on h_t is carried by x_t only.

Suppose σ_t, $t = 1, \ldots, T-1$, are all memoryless. We prove by induction backward in time that for all $t = T, \ldots, 1$, $v_t^{c,\pi}(h_t) = v_t^{c,\pi}(h'_t)$ for all Markov π and all $h_t, h'_t \in \mathcal{X}^t$ such that $x_t = x'_t$. For $t = T$ we have: $v_T^{c,\pi}(h_T) = c_T(x_T, \pi_T(x_T)) = v_T^{c,\pi}(h'_T)$. We can just write it as $v_T^{c,\pi}(x_T)$. If this relation is true for some $t+1 \le T$, then for t we obtain

$$\begin{aligned} v_t^{c,\pi}(h_t) &= c_t(x_t, \pi_t(x_t)) + \sigma_t\big(x_t, Q_t(x_t, \pi_t(x_t)), v_{t+1}^{c,\pi}(h_t, \cdot)\big) \\ &= c_t(x_t, \pi_t(x_t)) + \sigma_t\big(x_t, Q_t(x_t, \pi_t(x_t)), v_{t+1}^{c,\pi}(\cdot)\big). \end{aligned}$$

The right hand side is a function of x_t, rather than h_t, and we can write the value of the risk measure as $v_t^{c,\pi}(x_t)$. By induction, the result holds true for all t. $\square$

Theorem 4.1.3. *For a controlled Markov system, a family of process-based dynamic risk measures $\{\rho_{t,T}^{\pi}\}_{t=1,\ldots,T}^{\pi\in\Pi}$ is translation-invariant, stochastically conditionally time-consistent, and Markov, if and only if there exist functionals*

$$\sigma_t : \big\{ \big(x, Q_t(x,u)\big) : x \in \mathcal{X}, u \in \mathcal{U}_t(x) \big\} \times \mathcal{V} \to \mathbb{R}, \quad t = 1 \ldots T-1,$$

where $\mathcal{V}$ is the set of bounded measurable functions on $\mathcal{X}$, such that:

(i) *For all $t = 1, \ldots, T-1$ and all $x \in \mathcal{X}$, $\sigma_t(x, \cdot, \cdot)$ is normalized and strongly monotonic with respect to stochastic dominance on $\{ Q_t(x,u) : u \in \mathcal{U}_t(x) \}$;*

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

$$v_t^{c,\pi}(h_t) = c_t(x_t, \pi_t(h_t)) + \sigma_t\Big(x_t, Q_t(x_t, \pi_t(h_t)), v_{t+1}^{c,\pi}(h_t, \cdot)\Big). \tag{4.2}$$

Theorem 4.1.3 provides us with a simple recursive formula (4.2) for the evaluation of risk of a Markov policy π:

- for final time T

$$v_T^{c,\pi}(x) = c_T(x, \pi_T(x)), \quad x \in \mathcal{X}$$

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

$$v_t^{c,\pi}(x) = c_t(x, \pi_t(x)) + \sigma_t\Big(x, Q_t(x, \pi_t(x)), v_{t+1}^{c,\pi}\Big), \quad x \in \mathcal{X}$$

It involves calculation of the values of functions $v_t^{c,\pi}(\cdot)$ on the state space $\mathcal{X}$.

4.2 Dynamic Programming

In this section, we fix the cost functions $c = (c_1, \ldots, c_T)$ and consider a family of dynamic risk measures $\{\rho^{\pi}_{t,T}\}^{\pi\in\Pi}_{t=1,\ldots,T}$ which is normalized, translation-invariant (Definition 2.2.2), stochastically conditionally time-consistent (Definition 3.2.1), and Markov (Definition 4.1.1). Our objective is to analyze the risk minimization problem:

$$\min_{\pi\in\Pi} v^{\pi}_1(x_1), \quad x_1 \in \mathcal{X}. \tag{4.3}$$

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

$$v^*_t(h_t) = \inf_{\pi\in\Pi_{t,T}(h_t)} v^{\pi}_t(h_t), \quad t = 1, \ldots, T, \quad h_t \in \mathcal{X}^t, \tag{4.4}$$

where $\Pi_{t,T}(h_t)$ is the set of feasible deterministic policies $\pi = \{\pi_t, \ldots, \pi_T\}$. As stated in Theorem 4.1.3, transition risk mappings $\{\sigma_t\}_{t=1,\ldots,T-1}$ exist, such that

$$v^{\pi}_t(h_t) = c_t(x_t, \pi_t(h_t)) + \sigma_t\Big(x_t, Q_t(x_t, \pi_t(h_t)), v^{\pi}_{t+1}(h_t, \cdot)\Big),$$
$$t = 1, \ldots, T-1,\ \pi \in \Pi,\ h_t \in \mathcal{X}^t. \tag{4.5}$$

Our intention is to prove that the value functions $v^*_t(\cdot)$ are **memoryless**, that is, for all $h_t = (x_1, \ldots, x_t)$ and $h'_t = (x'_1, \ldots, x'_t)$ such that $x_t = x'_t$, we have $v^*_t(h_t) = v^*_t(h'_t)$. In this case, with a slight abuse of notation, we shall simply write $v^*_t(x_t)$.

In order to formulate the main result of this section, we equip the space $\mathcal{P}(\mathcal{X})$ of probability measures on $\mathcal{X}$ with the topology of weak convergence.

Theorem 4.2.1. *Suppose a family of dynamic risk measures $\{\rho^{\pi}_{t,T}\}^{\pi\in\Pi}_{t=1,\ldots,T}$ is normalized, translation-invariant, stochastically conditionally time-consistent, and Markov. We assume the following conditions:*

(i) *The transition kernels $Q_t(\cdot,\cdot)$, $t = 1, \ldots, T$, are weakly continuous;*

(ii) *For every lower semicontinuous $v \in \mathcal{V}$ the transition risk mappings $\sigma_t(\cdot,\cdot,v)$, $t = 1, \ldots, T$, are lower semicontinuous;*

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

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

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

$$v_T^*(x) = \min_{u \in \mathcal{U}_T(x)} c_T(x,u), \quad x \in \mathcal{X}, \tag{4.6}$$

$$v_t^*(x) = \min_{u \in \mathcal{U}_t(x)} \Big\{ c_t(x,u) + \sigma_t\big(x, Q_t(x,u), v_{t+1}^*\big) \Big\}, \quad x \in \mathcal{X}, \quad t = T-1, \dots, 1. \tag{4.7}$$

Moreover, the Markov policy $\hat{\pi}$ given by the equations:

$$\hat{\pi}_T(x) \in \operatorname*{argmin}_{u \in \mathcal{U}_T(x)} c_T(x,u), \quad x \in \mathcal{X},$$

$$\hat{\pi}_t(x) \in \operatorname*{argmin}_{u \in \mathcal{U}_t(x)} \Big\{ c_t(x,u) + \sigma_t\big(x, Q_t(x,u), v_{t+1}^*\big) \Big\}, \quad x \in \mathcal{X}, \quad t = T-1, \dots, 1,$$

is optimal for problem (4.3)*.*

Proof. We prove the memoryless property of $v_t^*(\cdot)$ and construct the optimal Markov policy by induction backwards in time. For all $h_T \in \mathcal{X}^T$ we have

$$v_T^*(h_T) = \inf_{\pi \in \Pi} c_T(x_T, \pi_T(h_T)) = \inf_{u \in \mathcal{U}_T(x_T)} c_T(x_T, u). \tag{4.8}$$

Since $c_T(\cdot,\cdot)$ is lower semicontinuous, it is a *normal integrand*, that is, its epigraphical mapping $x \mapsto \{(u,\alpha) \in \mathcal{U} \times \mathbb{R} : c_T(x,u) \le \alpha\}$ is closed-valued and measurable [49, Def. 14.1, Ex. 14.31]. Due to assumption (iv), the mapping

$$\bar{c}_T(x,u) = \begin{cases} c_T(x,u) & \text{if } u \in \mathcal{U}_T(x), \\ +\infty & \text{otherwise}, \end{cases}$$

is a normal integrand as well. By virtue of [49, Thm. 14.37], the infimum in (4.8) is attained and is a measurable function of x_T. Hence, $v_T^*(\cdot)$ is measurable and memoryless. By assumptions (iii) and (iv) and Berge's theorem, it is also lower semicontinuous (see, *e.g.*, [5, Thm. 1.4.16]). Moreover, the optimal solution mapping $\Psi_T(x) = \big\{ u \in \mathcal{U}_T(x) : c_T(x,u) = v_T^*(x) \big\}$ is measurable and has nonempty and closed values. Therefore, a measurable selector $\hat{\pi}_T$ of Ψ_T exists [38], [5, Thm. 8.1.3].

Suppose $v_{t+1}^*(\cdot)$ is memoryless and lower semicontinuous, and Markov decision rules $\{\hat{\pi}_{t+1}, \dots, \hat{\pi}_T\}$ exist such that

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

Then for any $h_t \in \mathcal{X}^t$ we have

$$v_t^*(h_t) = \inf_{\pi\in\Pi} v_t^\pi(h_t) = \inf_{\pi\in\Pi} \left\{c_t(x_t,\pi_t(h_t)) + \sigma_t\big(x_t, Q_t(x_t,\pi_t(h_t)), v_{t+1}^\pi(h_t,\cdot)\big)\right\}.$$

On the one hand, since $v_{t+1}^\pi(h_t,\cdot) \geq v_{t+1}^*(\cdot)$ and σ_t is non-decreasing with respect to the last argument, we obtain

$$\begin{aligned} v_t^*(h_t) &\geq \inf_{\pi\in\Pi} \Big\{c_t(x_t,\pi_t(h_t)) + \sigma_t\big(x_t, Q_t(x_t,\pi_t(h_t)), v_{t+1}^*\big)\Big\} \\ &= \inf_{u\in\mathcal{U}_t(x_t)} \Big\{c_t(x_t,u) + \sigma_t\big(x_t, Q_t(x_t,u), v_{t+1}^*\big)\Big\}. \end{aligned} \tag{4.9}$$

By assumptions (i)–(iii), the mapping $(x,u) \mapsto c_t(x,u) + \sigma_t\big(x, Q_t(x,u), v_{t+1}^*\big)$ is lower semicontinuous. Invoking [49, Thm. 14.37] and assumption (iv) again, exactly as in the case of $t = T$, we conclude that the optimal solution mapping

$$\Psi_t(x) = \Big\{u \in \mathcal{U}_t(x) \;\big|\; c_t(x,u) + \sigma_t\big(x, Q_t(x,u), v_{t+1}^*\big) = \\ \inf_{u\in\mathcal{U}_t(x)} \left\{c_t(x,u) + \sigma_t\big(x, Q_t(x,u), v_{t+1}^*\big)\right\}\Big\}$$

is measurable and has nonempty and closed values; hence, a measurable selector $\hat{\pi}_t$ of Ψ_t exists [5, Thm. 8.1.3]. Substituting this selector into (4.9), we obtain

$$v_t^*(h_t) \geq c_t(x_t,\hat{\pi}_t(x_t)) + \sigma_t\big(x_t, Q_t(x_t,\hat{\pi}_t(x_t)), v_{t+1}^{\{\hat{\pi}_{t+1},\ldots,\hat{\pi}_T\}}\big) = v_t^{\{\hat{\pi}_t,\ldots,\hat{\pi}_T\}}(x_t).$$

In the last equation, we used (4.5) and the fact that the decision rules $\hat{\pi}_t,\ldots,\hat{\pi}_T$ are Markov.

On the other hand,

$$v_t^*(h_t) = \inf_{\pi\in\Pi} v_t^\pi(h_t) \leq v_t^{\{\hat{\pi}_t,\ldots,\hat{\pi}_T\}}(x_t).$$

Therefore, $v_t^*(h_t) = v_t^{\{\hat{\pi}_t,\ldots,\hat{\pi}_T\}}(x_t)$ is measurable, memoryless, and

$$\begin{aligned} v_t^*(x_t) &= \min_{u\in\mathcal{U}_t(x_t)} \Big\{c_t(x_t,u) + \sigma_t\big(x_t, Q_t(x,u), v_{t+1}^*\big)\Big\} \\ &= c_t(x_t,\hat{\pi}_t(x_t)) + \sigma_t\big(x_t, Q_t(x_t,\hat{\pi}_t(x_t)), v_{t+1}^*\big). \end{aligned}$$

By assumptions (ii), (iii), (iv), and Berge's theorem, $v_t^*(\cdot)$ is lower semicontinuous (see, e.g. [5, Thm. 1.4.16]). This completes the induction step. $\square$

Remark 4.2.2. *If we replace semicontinuity with continuity in assumptions (ii)–(iv), then the value functions v_t^*, $t = 1, \dots, T$, will be continuous. The proof is identical.*

Let us verify the weak lower semicontinuity assumption (ii) of the mean–semideviation transition risk mapping of Example 2.4.8. To make the mapping Markovian, we assume that the parameter $\varkappa$ depends on x only, that is,

$$\sigma(x, q, v) = \int_{\mathcal{X}} v(s)\; q(ds) + \varkappa(x) \left(\int_{\mathcal{X}} \left[\left(v(s) - \int_{\mathcal{X}} v(s')\; q(ds') \right)_+ \right]^p q(ds) \right)^{1/p}. \quad (4.10)$$

As before, $p \in [1, \infty)$. For simplicity, we skip the subscript t of σ and $\varkappa$.

Lemma 4.2.3. *Suppose $\varkappa(\cdot)$ is continuous. Then for every lower semicontinuous function v, the mapping $(x, q) \mapsto \sigma(x, q, v)$ in (4.10) is lower semicontinuous.*

Proof. Let $q_k \to q$ weakly and $x_k \to x$. For all $s \in \mathcal{X}$ we have the inequality

$$\begin{aligned} 0 &\le \left[v(s) - \int_{\mathcal{X}} v(s')\; q(ds') \right]_+ \\ &\le \left[v(s) - \int_{\mathcal{X}} v(s')\; q_k(ds') \right]_+ + \left[\int_{\mathcal{X}} v(s')\; q_k(ds') - \int_{\mathcal{X}} v(s')\; q(ds') \right]_+ . \end{aligned}$$

By the triangle inequality for the norm in $\mathcal{L}_p(\mathcal{X}, \mathcal{B}(\mathcal{X}), q_k)$,

$$\begin{aligned} &\left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\; q(ds') \right]_+^p q_k(ds) \right)^{1/p} \\ &\le \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\; q_k(ds') \right]_+^p q_k(ds) \right)^{1/p} + \left[\int_{\mathcal{X}} v(s')\; q_k(ds') - \int_{\mathcal{X}} v(s')\; q(ds') \right]_+ . \end{aligned}$$

Adding $\int_{\mathcal{X}} v(s)\, q(ds)$ to both sides, we obtain

$$\begin{aligned} &\int_{\mathcal{X}} v(s)\; q(ds) + \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\; q(ds') \right]_+^p q_k(ds) \right)^{1/p} \\ &\le \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\; q_k(ds') \right]_+^p q_k(ds) \right)^{1/p} + \max \left\{ \int_{\mathcal{X}} v(s)\; q_k(ds),\; \int_{\mathcal{X}} v(s)\; q(ds) \right\}. \end{aligned}$$

By the lower semicontinuity of v and weak convergence of q_k to q, we have

$$\int_{\mathcal{X}} v(s)\; q(ds) \le \liminf_{k \to \infty} \int_{\mathcal{X}} v(s)\; q_k(ds),$$

that is, for every $\varepsilon > 0$, we can find k_ε such that for all $k \ge k_\varepsilon$

$$\int_{\mathcal{X}} v(s)\; q_k(ds) \ge \int_{\mathcal{X}} v(s)\; q(ds) - \varepsilon.$$

Therefore, for these k we obtain

$$\begin{aligned}&\int_{\mathcal{X}} v(s)\, q(ds) + \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\, q(ds')\right]_+^p q_k(ds)\right)^{1/p}\\ &\le \int_{\mathcal{X}} v(s)\, q_k(ds) + \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\, q_k(ds')\right]_+^p q_k(ds)\right)^{1/p} + \varepsilon.\end{aligned}$$

Taking the "lim inf" of both sides, and using the weak convergence of q_k to q and the lower semicontinuity of the functions integrated, we conclude that

$$\begin{aligned}&\int_{\mathcal{X}} v(s)\, q(ds) + \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\, q(ds')\right]_+^p q(ds)\right)^{1/p}\\ &\le \liminf_{k\to\infty} \left\{\int_{\mathcal{X}} v(s)\, q_k(ds) + \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\, q_k(ds')\right]_+^p q_k(ds)\right)^{1/p}\right\} + \varepsilon.\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, the last relation proves the lower semicontinuity of σ in the case when $\varkappa(x) \equiv 1$. The case of a continuous $\varkappa(x) \in [0,1]$ can be now easily analyzed by noticing that $\sigma(x,q,v)$ is a convex combination of the expected value, and the risk measure of the last displayed relation:

$$\begin{aligned}\sigma(x,q,v) &= \big(1 - \varkappa(x)\big) \int_{\mathcal{X}} v(s)\, q(ds)\\ &\quad + \varkappa(x) \left\{\int_{\mathcal{X}} v(s)\, q(ds) + \left(\int_{\mathcal{X}} \left[v(s) - \int_{\mathcal{X}} v(s')\, q(ds')\right]_+^p q(ds)\right)^{1/p}\right\}.\end{aligned}$$

As both components are lower semicontinuous in (x,q), so is their sum. □

We can also verify the weak continuity of the Average-Value-at-Risk transition risk mapping of Example 2.4.9. To make the mapping Markovian, we assume that the parameter α depends on x only, that is,

$$\sigma(x,q,v) = \min_{\eta\in\mathbb{R}} \left\{\eta + \frac{1}{\alpha(x)} \int_{\mathcal{X}} \big(v(s) - \eta\big)_+\, q(ds)\right\}, \tag{4.11}$$

For simplicity, we skip the subscript t of σ and α.

Lemma 4.2.4. *Suppose $\alpha(\cdot)$ is continuous and takes values in $[\alpha_{\min}, \alpha_{\max}] \subset (0,1)$. Then for every continuous function v, the mapping $(x,q) \mapsto \sigma(x,q,v)$ in (4.11) is continuous.*

Proof. Consider the function

$$(q,\eta) \mapsto \int_{\mathcal{X}} \big(v(s) - \eta\big)_+\, q(ds). \tag{4.12}$$

Suppose $q_k \to q$ weakly and $\eta_k \to \eta$. We have

$$\int_{\mathcal{X}} \big(v(s)-\eta\big)_+ \, q_k(ds) \le \int_{\mathcal{X}} \big(v(s)-\eta_k\big)_+ \, q_k(ds) + |\eta_k - \eta|.$$

Taking the "lim inf" of both sides and using the weak convergence of q_k to q and the lower semicontinuity of the functions integrated, we conclude that

$$\begin{aligned} & \int_{\mathcal{X}} \big(v(s)-\eta\big)_+ \, q(ds) \\ \le \; & \liminf_{k\to\infty} \int_{\mathcal{X}} \big(v(s)-\eta\big)_+ \, q_k(ds) \\ \le \; & \liminf_{k\to\infty} \int_{\mathcal{X}} \big(v(s)-\eta_k\big)_+ \, q_k(ds). \end{aligned}$$

Thus the function (4.12) is lower semicontinuous. It follows that the function being minimized with respect to η in (4.11) is jointly lower semicontinuous with respect to (x, q, η). Since $q_k \to q$ weakly, the collection $\{q_k\}$ is *tight* (Prohorov's theorem; see, *e.g.*, [11, Sec. 1.6]). Since $v(\cdot)$ is continuous, the measures $q_k \circ v^{-1}$ are tight as well. Hence, a bounded interval $C \subset \mathbb{R}$ exists such that all α-quantiles of all $q_k \circ v^{-1}$ and $q \circ v^{-1}$ are contained in C, for all $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. Therefore, we can restrict η to C in (4.11), without affecting the values of $\sigma(x_k, q_k, v)$ and $\sigma(x, q, v)$. By Berge's theorem, the optimal value in (4.11) is continuous (see, *e.g.*, [5, Thm. 1.4.16]). $\square$

4.3 Application to Multistage Stochastic Programming with Decision-Dependent Probabilities

In this section, we consider a multistage stochastic programming problem with three ingredients in the model. The first one is the history-dependent process $\{I_t\}_{t=1,\dots,T}$ with each I_t taking values in a Borel space $\mathcal{I}_t$ (a Borel subset of a Polish space). We can interpret $\{I_t\}$ as the process of exogeneous random inputs (data process).

The second one is the "physical" state process $\{Y_t\}_{t=1,\dots,T}$ on the Borel space $\mathcal{Y}$, and the third is the control process $\{U_t\}_{t=1,\dots,T}$ on the Borel control space $\mathcal{U}$. The processes are related by the equation:

$$Y_{t+1} = f_t(Y_t, U_t, I_{t+1}), \quad t = 1, \dots, T-1, \tag{4.13}$$

in which $f_t : \mathcal{Y} \times \mathcal{U} \times \mathcal{I}_{t+1} \to \mathcal{Y}$ are measurable functions. We assume that $Y_1 \in \mathcal{Y}$ is deterministic and known.

The distribution of the process $\{I_t\}$ is defined by the controlled transition kernels $K_t : \mathcal{I}_1 \times \cdots \times \mathcal{I}_t \times \mathcal{U} \to \mathcal{P}(\mathcal{I}_{t+1})$,

$$\mathbb{P}[I_{t+1} \in B] = K_t(i_1, \dots, i_t, u_t)(B), \quad \forall\, B \in \mathcal{B}(\mathcal{I}_{t+1}).$$

We can now proceed as in Section 3.1, substituting X by (I, Y) and Q by K.

1. The information revealed at time t is i_t and y_t, and then we make decision u_t in the admissible control set $\mathcal{U}_t(i_1, \dots, i_t, y_t)$.

2. The set of admissible policies is

$$\Pi = \left\{ \pi = (\pi_1, \dots, \pi_T) \,\middle|\, \quad \forall t,\ \pi_t(i_1, \dots, i_t, y_t) \in \mathcal{U}_t(i_1, \dots, i_t, y_t) \right\}.$$

3. For a fixed policy $\pi \in \Pi$, the physical state Y_t is formally a function of $i_1, \dots, i_t$. Therefore, the resulting transition kernels can be formally written as

$$K_t^\pi : (i_1, \dots, i_t) \mapsto K_t\big(i_1, \pi_1(i_1), \dots, i_t, \pi_t(i_1, \dots, i_t, y_t(i_1, \dots, i_t, \pi))\big) \in \mathcal{P}(\mathcal{I}_{t+1}),$$
$$t = 1 \dots T - 1.$$

Example 4.3.1. *In a special case, $\{I_t\}$ may be a process on a scenario tree, where I_1 is deterministic (the root node), and each set $\mathcal{I}_t$ is the set of possible histories $(i_1, \dots, i_t)$ which are represented as nodes at level t of the tree. The variable I_{t+1}, given $i_1, \dots, i_t$, has a finite number of possible realizations: a set of successor nodes of the node at level t representing the history $(i_1, \dots, i_t)$. In this case, $K_t^\pi(i_1, \dots, i_t)$ is the conditional probability distribution of the next I_{t+1} on this finite set. Our model allows for these conditional probabilities to depend on the control u_t at time t.* □

It follows that the process

$$\big\{X_t = (I_1, \dots, I_t, Y_t)\big\}_{t=1,\dots,T}$$

is a controlled Markov process with values in the spaces

$$\mathcal{X}_t = \mathcal{I}_1 \times \cdots \times \mathcal{I}_t \times \mathcal{Y}, \quad t = 1, \dots, T,$$

because of $\big\{(I_1, \dots, I_t)\big\}_{t=1,\dots,T}$ being Markov, and because of (4.13).

If we assume in addition that stage-wise costs have the form $Z_t = c_t(Y_t, U_t)$, we can apply the results of Chapter 4 (in particular Theorem 4.1.3) with the notions of H_t, h_t, Q_t, $\mathcal{Z}_t$, ρ, $\mathit{\Pi}$, v_t associated with the process X. The policy evaluation equation (4.2) in this case takes on the form

$$v_t^{c,\pi}(h_t) = c_t(y_t, \pi_t(h_t)) + \sigma_t\Big(x_t, Q_t(x_t, \pi_t(h_t)), v_{t+1}^{c,\pi}(h_t, \cdot)\Big) \tag{4.14}$$

for any $\pi \in \mathit{\Pi} = \big\{\pi = (\pi_1, \ldots, \pi_T) \,|\, \forall t,\ \pi_t(x_1, \ldots, x_t) \in \mathcal{U}_t(x_t)\big\}$. As for a fixed policy $\pi \in \mathit{\Pi}$, H_t is function of $(I_1 \ldots I_t)$, the transition kernel $Q_t(x_t, \pi_t(h_t))$ is solely determined by $K_t^\pi(i_1, \ldots, i_t)$. Consequently, if we only consider the histories h_t that can happen, (4.14) becomes

$$\begin{aligned} w_t^{c,\pi}(i_1, \ldots, i_t) = c_t(y_t, \pi_t(i_1, \ldots, i_t)) \\ + \sigma_t\Big(i_1, \ldots, i_t, y_t, K_t^\pi(i_1, \ldots, i_t), w_{t+1}^{c,\pi}(i_1, \ldots, i_t, \cdot)\Big), \ \forall \pi \in \mathit{\Pi}, \end{aligned} \tag{4.15}$$

where

$$w_t^{c,\pi}(i_1, \ldots, i_t) = v_t^{c,\pi}(H_t^\pi(i_1, \ldots, i_t)).$$

In this equation, y_t is considered as a dependent variable, which can be evaluated for a given policy π as a function of the history $i_1, \ldots, i_t$. Thus, the policy evaluation can be carried out on the space of functions of $(i_1, \ldots, i_t)$.

In the dynamic programming equation (4.6) we must take into account the entire state X_t. The equation takes on the form:

$$\begin{aligned} w_t^*(i_1, \ldots, i_t, y_t) = \min_{u \in \mathcal{U}_t(i_1, \ldots, i_t, y_t)} \Big\{ c_t(y_t, u) \\ + \sigma_t\Big(i_1, \ldots, i_t, y_t, K_t(i_1, \ldots, i_t, u), i \mapsto w_{t+1}^*\big(i_1, \ldots, i_t, i, f_t(y_t, u, i)\big)\Big). \end{aligned}$$

The reason is that the dynamic programming equation is solved backwards, and there is no way to know the value of y_t. For each y_t, however, the last argument of σ_t is a function of the new information i_{t+1} only, because the next y_{t+1} follows from (y_t, u, i_{t+1}) via (4.13).

In the scenario tree case, this result extends the dynamic programming equations of [60, sec. 6.8.4] in two ways: it is derived from the assumptions about the stochastic conditional time consistency and the Markovian structure of the risk measure, and it covers the case of decision-dependent probabilities.

Chapter 5

Partially Observable Markov Decision Processes

5.1 The Model

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

In order to develop our subsequent theory, it is essential to define the model in a clear and rigorous way. This section follows [6, Ch. 5], which the readers are encouraged to consult for more details.

The state space of the model is defined as $\mathcal{X} \times \mathcal{Y}$ where $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ are two Borel spaces (Borel subsets of Polish spaces). From the modeling perspective, $x \in \mathcal{X}$ is the part of the state that we can observe at each step, while $y \in \mathcal{Y}$ is unobservable. The measurable space that we will work with is then given by $\Omega = (\mathcal{X} \times \mathcal{Y})^T$ endowed with the canonical product σ-field $\mathcal{F}$, and we use x_t and y_t to denote the canonical projections at time t.

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

The control space is given by a Borel space $\mathcal{U}$, and since only $\mathcal{X}$ is observable, the set of admissible controls at step t is given by a measurable multifunction $\mathcal{U}_t : \mathcal{X} \rightrightarrows \mathcal{U}$ with nonempty values. The transition kernel at time t is

$$K_t : \operatorname{graph}(\mathcal{U}_t) \times \mathcal{Y} \to \mathcal{P}(\mathcal{X} \times \mathcal{Y}),$$

where $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is the space of probability measures on $\mathcal{X} \times \mathcal{Y}$. In other words, if the state at time t is (x, y) and we apply control u, the distribution of the next state is $K_t(\,\cdot\,,\cdot\mid x, y, u)$.

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

$$\mathcal{H}_t = \mathcal{X}^t = \underbrace{\mathcal{X} \times \cdots \times \mathcal{X}}_{t \text{ times}}$$

and

$$\begin{cases} \mathcal{G}_1 = \mathcal{X}, \\ \mathcal{G}_t = \{(g_{t-1}, u_{t-1}) \mid g_{t-1} \in \mathcal{G}_{t-1},\, u_{t-1} \in \mathcal{U}_t(x_{t-1})\} \times \mathcal{X}, \quad t \geq 2, \end{cases} \tag{5.1}$$

to respectively denote the spaces of possible histories h_t and g_t. Here we make distinction of g_t and h_t because we should make decision of u_t based on g_t as the past controls $u_1, \ldots, u_{t-1}$ are also taken into consideration when estimating the conditional distribution of Y_t (see Section 5.3). We still write $H_t = (X_1, \ldots, X_t)$.

For this controlled process, a (deterministic) *history-dependent admissible policy* $\pi = (\pi_1, \ldots, \pi_T)$ is a sequence of measurable decision rules $\pi_t : \mathcal{G}_t \to \mathcal{U}$ such that $\pi_t(g_t) \in \mathcal{U}_t(x_t)$ for all $g_t \in \mathcal{G}_t$ (such a policy exists, due to the measurable selector theorem of [38]). We can easily prove by induction on t that for such an admissible policy π, each π_t reduces to a measurable function of $h_t = (x_1, x_2, \ldots, x_t)$, as $u_s = \pi_s(x_1, \ldots, x_s)$ for all $s = 1, \ldots, t-1$. We are still using π_s to denote the decision rule, although it is a different function, formally; it will not lead to any misunderstanding. Therefore the set of admissible policies is

$$\mathit{\Pi} = \left\{ \pi = (\pi_1, \ldots, \pi_T) \,\middle|\, \pi_t(x_1, \ldots, x_t) \in \mathcal{U}_t(x_t),\ t = 1, \ldots, T \right\}.$$

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

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

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

For any $Z \in \mathcal{Z}_t$, there is a measurable and bounded functional $\overline{Z} : (\mathcal{X} \times \mathcal{Y})^t \to \mathbb{R}$

such that $Z = \overline{Z}(X_1, Y_1, \dots, X_t, Y_t)$. With a slight abuse of notation, we still use Z to denote this function.

5.2 Risk Filters for Partially Observable Systems

5.2.1 Dynamic risk filters

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

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

We have $\mathcal{S}_t \subset \mathcal{Z}_t$, and any element $S \in \mathcal{S}_t$ can also be considered as a measurable bounded functional on $\mathcal{X}^t$; with slight abuse of notation, $S = S(H_t)$. All equality and inequality relations between random variables are understood in the "everywhere" sense.

Definition 5.2.1. *A mapping $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{S}_t$, where $1 \le t \le T$, is called a **conditional risk evaluator**, if it satisfies the* monotonicity property*: for all $(Z_t, \dots, Z_T)$ and $(W_t, \dots, W_T)$ in $\mathcal{Z}_{t,T}$, if $Z_s \le W_s$ for all $s = t, \dots, T$, then*

$$\rho_{t,T}(Z_t, \dots, Z_T) \le \rho_{t,T}(W_t, \dots, W_T).$$

Definition 5.2.2. *A conditional risk evaluator $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{S}_t$*

(i) *is **normalized** if $\rho_{t,T}(0, \dots, 0) = 0$;*

(ii) *is **translation invariant** if $\forall (Z_t, \dots, Z_T) \in \mathcal{S}_t \times \mathcal{Z}_{t+1,T}$,*
$\rho_{t,T}(Z_t, \dots, Z_T) = Z_t + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T)$;

(iii) *is **decomposable** if a mapping $\rho_t : \mathcal{Z}_t \to \mathcal{S}_t$ exists such that:*

$$\begin{cases} \rho_t(Z_t) = Z_t, \quad \forall Z_t \in \mathcal{S}_t, \\ \rho_{t,T}(Z_t, \dots, Z_T) = \rho_t(Z_t) + \rho_{t,T}(0, Z_{t+1}, \dots, Z_T), \quad \forall (Z_t, \dots, Z_T) \in \mathcal{Z}_{t,T}. \end{cases}$$

The concept of conditional risk evaluator reduces to the conditional risk measure (Definition 2.2.1), if the arguments $Z_t, \dots, Z_T$ are in the spaces $\mathcal{S}_t \times \cdots \times \mathcal{S}_T$; and the translation invariance property in the above definition is defined in the same way as in Definition 2.2.2, while the spaces of arguments $Z_{t+1}, \dots, Z_T$ are larger. The decomposability property is a similar but stronger assumption compared to the translation invariance.

Throughout the chapter, we assume all conditional risk evaluators to be at least normalized. The properties (i) and (iii) of Definition 5.2.2 imply that

$$\rho_t(Z_t) = \rho_{t,T}(Z_t, 0, \dots, 0). \tag{5.2}$$

We also redefine the local property (see Definition 2.2.3 for the conditional risk measures) specifically for the conditional risk evaluators.

Definition 5.2.3. *A conditional risk evaluator $\rho_{t,T}$ has the* ***local property*** *if for any event $A \in \mathcal{F}_t^X$ and all $(Z_t, \dots, Z_T) \in \mathcal{Z}_{t,T}$ we have*

$$\mathbb{1}_A \rho_{t,T}(Z_t, \dots, Z_T) = \rho_{t,T}(\mathbb{1}_A Z_t, \dots, \mathbb{1}_A Z_T).$$

The local property means that the conditional risk evaluator at time t restricted to any $\mathcal{F}_t^X$-event A is not influenced by the values that $Z_t, \dots, Z_T$ take on A^c.

Definition 5.2.4. *A* risk filter $\{\rho_{t,T}\}_{t=1,\dots,T}$ *is a sequence of conditional risk evaluators $\rho_{t,T} : \mathcal{Z}_{t,T} \to \mathcal{S}_t$. We say that it is normalized, translation-invariant, decomposable, or has the local property, if all $\rho_{t,T}$, $t = 1, \dots, T$, satisfy the respective conditions of Definitions 5.2.2 or 5.2.3.*

5.2.2 Stochastic conditional time consistency

For a POMDP defined in Section 5.1, we have to use a family of risk filters $\{\rho^{\pi}_{t,T}\}^{\pi \in \Pi}_{t=1,\dots,T}$, because the policy affects the probability measure on the space $\Omega = (\mathcal{X} \times \mathcal{Y})^T$. When two policies π and π' are compared, even if the resulting costs were pointwise equal, $\rho^{\pi}_{t,T}(Z_t, \dots, Z_T)$ and $\rho^{\pi'}_{t,T}(Z_t, \dots, Z_T)$ should not be necessarily equal, because the probability measures P^{π} and $P^{\pi'}$ could be different. We extend the concept stochastic conditional time consistency (Definition 3.2.1) that allows us to relate the whole family of risk filters.

Definition 5.2.5. *A family of risk filters* $\{\rho^{\pi}_{t,T}\}^{\pi\in\Pi}_{t=1,\dots,T}$ *is* ***stochastically conditionally time-consistent*** *if for any* $\pi,\pi'\in\Pi$*, for any* $1\le t<T$*, for all* $h_t\in\mathcal{X}^t$*, all* $(Z_{t+1},\dots,Z_T)\in\mathcal{Z}_{t+1,T}$ *and all* $(W_{t+1},\dots,W_T)\in\mathcal{Z}_{t+1,T}$*, the condition*

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

implies

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

Remark 5.2.6. *The conditional stochastic order "$\preceq_{\text{st}}$" is understood as follows: for all* $\eta\in\mathbb{R}$ *we have*

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

Proposition 5.2.7. *A family of risk filters* $\{\rho^{\pi}_{t,T}\}^{\pi\in\Pi}_{t=1,\dots,T}$ *that is normalized, decomposable, and stochastically conditionally time-consistent has the local property if and only if all* ρ^{π}_t *(Definition 5.2.2) satisfy the local property:* $\rho^{\pi}_t(\mathbb{1}_A Z_t)=\mathbb{1}_A\rho^{\pi}_t(Z_t)$ *for all* $A\in\mathcal{F}^X_t$*,* $Z_t\in\mathcal{Z}_t$*,* $\pi\in\Pi$*, and* $t=1,\dots,T$*.*

Proof. The "only if" part is obvious because of (5.2); we need to prove that if all ρ^{π}_t satisfy the local property, then for any $\pi\in\Pi$, for any $t=1,\dots,T$, and any $A\in\mathcal{F}^X_t$,

$$\mathbb{1}_A\rho^{\pi}_{t,T}(Z_t,\dots,Z_T)=\rho^{\pi}_{t,T}(\mathbb{1}_A Z_t,\dots,\mathbb{1}_A Z_T). \tag{5.3}$$

We use induction on t from T down to 1. At the final time,

$$\rho^{\pi}_{T,T}(\mathbb{1}_A Z_t)=\rho^{\pi}_T(\mathbb{1}_A Z_t)=\mathbb{1}_A\rho^{\pi}_T(Z_t).$$

Suppose $\rho^{\pi}_{t+1,T}$ has the local property. Then by decomposability we have

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

and

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

As $\rho^{\pi}_t(\mathbb{1}_A Z_t)=\mathbb{1}_A\rho^{\pi}_t(Z_t)$, to verify (5.3) we need to show that

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

For any $h_t \in \mathcal{X}^T$ we have

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

The local property of ρ^{π}_{t+1} yields

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

so by stochastic conditional time consistency,

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

Thus,

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

which proves (5.3).

$\square$

The following theorem shows that the stochastic conditional time consistency implies that one-step risk mappings can be represented by static law-invariant risk measures on

$$\mathcal{V} = \left\{ v : \mathcal{X} \to \mathbb{R} \,\middle|\, v \text{ is measurable and bounded} \right\}. \tag{5.4}$$

Recall that a measurable function $r : \mathcal{V} \to \mathbb{R}$ is said to be *monotonic*, *normalized* and *translation invariant* if it satisfies the Definition 1.3.1. It is said to be *law-invariant* with respect to a probability measure q on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ if it satisfies Definition 2.4.4.

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

$$Q^{\pi}_t(h_t)(C) = \mathbb{P}^{\pi}[X_{t+1} \in C \mid H_t = h_t], \quad \forall C \in \mathcal{B}(\mathcal{X}). \tag{5.5}$$

Later in Section 5.3 we will show that Q^{π}_t can be computed in a recursive way with the help of belief states and Bayes operators.

We can now state the main result of this section.

Theorem 5.2.8. *A family of risk filters* $\left\{\rho^{\pi}_{t,T}\right\}^{\pi\in\Pi}_{t=1,\ldots,T}$ *is normalized, translation invariant, and stochastically conditionally time-consistent if and only if transition risk mappings*

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

exist, such that

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

$$\forall q^1, q^2 \in \left\{ Q^{\pi}_t(h_t) : \pi \in \Pi \right\},\ \forall v^1, v^2 \in \mathcal{V},$$
$$(v^1 \mid q^1) \preceq_{\mathrm{st}} (v^2 \mid q^2) \implies \sigma_t(h_t, q^1, v^1) \le \sigma_t(h_t, q^2, v^2),$$

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

(ii) *For all* $\pi \in \Pi$*, for all* $t = 1\ldots T-1$*, for all* $(Z_t,\ldots,Z_T) \in \mathcal{S}_t \times \mathcal{Z}_{t+1,T}$*, and for all* $h_t \in \mathcal{X}^t$*,*

$$\rho^{\pi}_{t,T}(Z_t, Z_{t+1},\ldots,Z_T)(h_t) = Z_t + \sigma_t(h_t, Q^{\pi}_t(h_t), \rho^{\pi}_{t+1,T}(Z_{t+1},\ldots,Z_T)(h_t,\cdot)). \tag{5.6}$$

Moreover, for all $t = 1\ldots T-1$*,* σ_t *is uniquely determined by* $\{\rho^{\pi}_{t,T}\}^{\pi\in\Pi}$ *as follows: for every* $h_t \in \mathcal{X}^t$*, for every* $q \in \left\{ Q^{\pi}_t(h_t) : \pi \in \Pi \right\}$*, and for every* $v \in \mathcal{V}$*,*

$$\sigma_t(h_t, q, v) = \rho^{\pi}_{t,T}(0, V, 0, \ldots, 0)(h_t),$$

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

Proof.

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

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

we obtain, by translation invariance and normalization,

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

Thus, by translation invariance and stochastic conditional time consistency,

$$\begin{aligned} & \rho^{\pi}_{t,T}(Z_t,\dots,Z_T)(h_t) \\ = \;& Z_t(h_t) + \rho^{\pi}_{t,T}(0,Z_{t+1},\dots,Z_T)(h_t) \\ = \;& Z_t(h_t) + \rho^{\pi}_{t,T}(0,W,0,\dots,0)(h_t) \\ = \;& Z_t(h_t) + \sigma^{\pi}_t(h_t,q,w). \end{aligned} \tag{5.7}$$

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

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

$$(v^1 \mid Q^{\pi_1}_t(h_t)) \preceq_{\text{st}} (v^2 \mid Q^{\pi_2}_t(h_t)), \tag{5.8}$$

where $v^1, v^2 \in \mathcal{V}$ and $h_t \in \mathcal{X}^t$. Define $V^1(h_t,\cdot) = v^1(\cdot)$ and $V^2(h_t,\cdot) = v^2(\cdot)$. Then Definition 5.2.5 implies that

$$\rho^{\pi_1}_{t,T}(0,V^1,0,\dots,0)(h_t) \le \rho^{\pi_2}_{t,T}(0,V^2,0,\dots,0)(h_t).$$

This combined with (5.7) yields

$$\sigma^{\pi_1}_t(h_t,Q^{\pi_1}_t(h_t),v^1) \le \sigma^{\pi_2}_t(h_t,Q^{\pi_2}_t(h_t),v^2). \tag{5.9}$$

Suppose $Q^{\pi_1}_t(h_t) = Q^{\pi_2}_t(h_t)$ and $v^1 = v^2$. Then both $\preceq_{\text{st}}$ and $\succeq_{\text{st}}$ are true in (5.8) and thus (5.9) becomes an equality. This proves that in fact σ^{π}_t does not depend on π, and all dependence on π is carried by the controlled kernel Q^{π}_t. Moreover, the function $\sigma_t(h_t,\cdot,\cdot)$ is indeed strongly monotonic with respect to stochastic dominance.

- On the other hand, if such transition risk mappings σ_t exist, then $\{\rho_{t,T}\}^{\pi\in\Pi}_{t=1,\dots,T}$ is stochastically conditionally time-consistent by the monotonicity and law invariance of $\sigma_t(h_t,\cdot,\cdot)$. We can now use (5.6) to obtain for any $t = 1,\dots,T-1$, and for all $h_t \in \mathcal{X}^t$ the translation invariance of $\rho^{\pi}_{t,T}$.

$\square$

Let us stress that in Theorem 5.2.8 we assume the translation invariance and therefore $Z_t \in \mathcal{S}_t$ in formula (5.6) (Z_t is observed at time t). For general $Z_t \in \mathcal{Z}_t$ we have the following corollary.

Corollary 5.2.9. *If a family of risk filters* $\left\{\rho^{\pi}_{t,T}\right\}^{\pi\in\Pi}_{t=1,\dots,T}$ *is normalized, decomposable, and stochastically conditionally time-consistent, then for any* $\pi \in \Pi$, $t = 1\dots T-1$, *for all* $(Z_t,\dots,Z_T) \in \mathcal{Z}_{t,T}$ *and all* $h_t \in \mathcal{X}^t$,

$$\rho^{\pi}_{t,T}(Z_t, Z_{t+1},\dots,Z_T)(h_t) = \rho^{\pi}_t(Z_t) + \sigma_t(h_t, Q^{\pi}_t(h_t), \rho^{\pi}_{t+1,T}(Z_{t+1},\dots,Z_T)(h_t,\cdot)), \quad (5.10)$$

with the transition risk mappings σ_t *defined in Theorem 5.2.8. Moreover, the functionals* $\left\{\rho^{\pi}_t\right\}^{\pi\in\Pi}_{t=1,\dots,T}$, *are monotonic, normalized, and translation invariant risk measures.*

Proof. The derivation is identical, just in the chain of relations (5.7) we use decomposability instead of translation invariance. The properties of the functionals ρ^{π}_t follow directly from the corresponding properties of the risk filter. □

Theorem 5.2.8 and Corollary 5.2.9 can be considered as the counterpart of Theorem 3.2.3 in the POMDP case. They allow us to characterize and construct stochastically conditionally time-consistent risk filters defined on space of cost functions adapted to $\mathcal{F}^{X,Y}$, by static risk measures defined on the space of the observed state $\mathcal{X}$.

Corollary 5.2.10. *If a family of risk filters* $\left\{\rho^{\pi}_{t,T}\right\}^{\pi\in\Pi}_{t=1,\dots,T}$ *is normalized, decomposable and stochastically conditionally time-consistent, then for any* $\pi \in \Pi$, $t = 1,\dots,T-1$, *for all* $(Z_t,\dots,Z_T) \in \mathcal{Z}_{t,T}$ *and all* $h_t \in \mathcal{X}^t$,

$$\rho^{\pi}_{t,T}(Z_t, Z_{t+1},\dots,Z_T)(h_t) = \rho^{\pi}_{t,T}(W_t, W_{t+1},\dots,W_T)(h_t),$$

where $W_t \in \mathcal{S}_t$ *is defined as* $W_t = \rho^{\pi}_t(Z_t)$, *for* $t = 1,\dots,T$.

Proof. The result can be verified from (5.10) by induction on t from T down to 1.

□

5.3 Bayes Operator and Belief States

5.3.1 Bayes operator

At each time t, the conditional distribution of the next observable state $Q_t^\pi(h_t)$ defined in (5.5) can be easily computed if we know the conditional distribution of the current unobservable state, called the **belief state**:

$$\varXi_t^\pi(h_t) \in \mathcal{P}(\mathcal{Y}) : \ \varXi_t^\pi(h_t)(D) = \mathbb{P}^\pi[Y_t \in D \,|\, H_t = h_t], \quad \forall D \in \mathcal{B}(\mathcal{Y}), \tag{5.11}$$

as we have

$$Q_t^\pi(h_t) = \int_{\mathcal{Y}} K_t^X(x_t, y, \pi_t(h_t)) \ \varXi_t^\pi(h_t)(dy), \tag{5.12}$$

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

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

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

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

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

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

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

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

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

Example 5.3.1 (Bayes operator with kernels given by density functions)**.** *Assume that each transition kernel $K_t(x, y, u)$ has a density $q_t(\cdot, \cdot \mid x, y, u)$ with respect to a finite product measure $\mu_X \otimes \mu_Y$ on $\mathcal{X} \times \mathcal{Y}$. Then the Bayes operator has the form*

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

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

Thus, we can calculate Q_t and $\varXi_t$ (defined in (5.5) and (5.11)) based on $g_t = (x_1, u_1, \ldots, x_{t-1}, u_{t-1}, x_t) \in \mathcal{G}_t$ (5.1) recursively with the help of Bayes operators:

- The initial belief $\varXi_1(x_1)$ is the conditional distribution of Y_1 given $X_1 = x_1$;
- The Bayes operator provides us the following formula to update the belief states:

  $$\varXi^{\pi}_{t+1}(h_{t+1}) = \varPhi_t\big(x_t, \varXi^{\pi}_t(h_t), \pi_t(h_t), x_{t+1}\big),$$

  and, by induction on t,

  $$\varXi^{\pi}_t(h_t) = \varXi_t\big(x_1, \pi_1(x_1), \ldots, x_{t-1}, \pi_{t-1}(h_{t-1}), x_t\big) = \varXi_t\big(G^{\pi}_t(h_t)\big);$$

- The conditional distribution of X_{t+1} can be calculated by (5.12), and

  $$Q^{\pi}_t(h_t) = Q_t\big(x_1, \pi_1(x_1), \ldots, x_t, \pi_t(h_t)\big) = Q_t\big(G^{\pi}_t(h_t), \pi_t(h_t)\big).$$

5.3.2 Markov problems

For this section, we make the following additional assumptions:

Assumption 5.3.2. *The costs $Z_1, \ldots, Z_T$ are only dependent on the current states and controls, that is, they have the following form*

$$Z^{\pi}_t = Z_t(X_t, Y_t, \pi_t(H_t)), \quad \forall t = 1, \ldots, T, \quad \forall \pi \in \varPi. \tag{5.13}$$

Assumption 5.3.3. *For any $t = 1, \ldots, T$ and all $\pi \in \Pi$ the operators $\{\rho_t^\pi\}^{\pi \in \Pi}$ (introduced in Defintion 5.2.2) are law-invariant in the following sense:*

$$\left[\rho_t^\pi(Z_t^\pi)\right](h_t) = \left[\rho_t^{\pi'}(Z_t^{\pi'})\right](h'_t)$$

for all $\pi, \pi' \in \Pi$ and all $h_t, h'_t \in \mathcal{H}_t$ such that

$$\mathbb{P}^\pi\left[Z_t^\pi \le \eta \mid H_t = h_t\right] = \mathbb{P}^{\pi'}\left[Z_t^{\pi'} \le \eta \mid H_t = h'_t\right], \quad \forall \eta \in \mathbb{R}.$$

Under these assumptions we are able to express the immediate risk of a cost Z_t as a function of the extended state.

Proposition 5.3.4. *Under Assumptions 5.3.2 and 5.3.3, a bounded measurable function $r_t : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \times \mathcal{U} \to \mathbb{R}$ exists, such that for all $t = 1, \ldots, T$, all $h_t \in H_t$, and all $\pi \in \Pi$,*

$$\rho_t^\pi(Z_t(X_t, Y_t, \pi_t(H_t)))(h_t) = r_t(x_t, \Xi_t^\pi(h_t), \pi_t(h_t)),$$

Proof. For any $t = 1, \ldots, T$, all $h_t, h'_t \in \mathcal{H}_t$, all $\pi, \pi' \in \Pi$, if $x_t = x'_t$, $\Xi_t^\pi(h_t) = \Xi_t^{\pi'}(h'_t)$, and $\pi_t(h_t) = \pi'_t(h'_t)$, then

$$\mathbb{P}^\pi\left[Z_t(X_t, Y_t, \pi_t(H_t)) \le \eta \mid H_t = h_t\right] = \mathbb{P}^{\pi'}\left[Z_t(X_t, Y_t, \pi'_t(H_t)) \le \eta \mid H_t = h'_t\right].$$

By Assumption 5.3.3, the following equality holds:

$$\rho_t^\pi(Z_t(X_t, Y_t, \pi_t(H_t)))(h_t) = \rho_t^{\pi'}(Z_t(X_t, Y_t, \pi'_t(H_t)))(h'_t).$$

This means that $\rho_t^\pi(Z_t(X_t, Y_t, \pi_t(H_t)))(h_t)$ is in fact a function of x_t, $\Xi_t^\pi(h_t)$, and $\pi_t(h_t)$. The boundedness of r_t follows from the boundedness of $\rho_t^\pi(Z_t)$.

□

Proposition 5.3.4 allows us to substitute $Z_1^\pi, \ldots, Z_T^\pi$ of the form (5.13) with

$$W_t^\pi = r_t(x_t, \Xi_t^\pi(h_t), \pi_t(h_t)), \quad t = 1, \ldots, T, \tag{5.14}$$

and the conditional risk evaluator keeps the same values. For example, in expected value models, we have $r_t(x, \xi, u) = \int_{\mathcal{Y}} c_t(x, y, u)\, \xi(dy)$, where $c_t : \mathcal{X} \times \mathcal{Y} \times \mathcal{U} \to \mathbb{R}$ is the running cost function, but more general functionals can be used here instead of the expectation with respect to the belief state. We call the functions $r_t(\cdot,\cdot,\cdot)$ the **running risk functions**. In the case when the running cost does not depend on the unobservable state, *i.e.*, $Z^\pi_t = c_t(X_t, \pi_t(H_t))$ in (5.13), we simply have $r_t(x_t, \Xi^\pi_t(h_t), \pi_t(h_t)) = c_t(x_t, \pi_t(h_t))$ and the running risk function reduces to the running cost function.

The transition risk mappings of Examples 2.4.7, 2.4.8, and 2.4.9 satisfy the conditions of Theorem 5.6 corresponding to stochastically conditionally time-consistent risk filters.

5.3.3 Markov risk measures

We say that a policy $\pi \in \Pi$ is *Markov* if each decision rule $\pi_t(\cdot)$ depends only on the current observed state x_t and the current belief state ξ_t.

Definition 5.3.5. *In POMDP, a policy $\pi \in \Pi$ is **Markov** if $\pi_t(h_t) = \pi_t(h'_t)$ for all $t = 1, \ldots, T$ and all h_t, $h'_t \in \mathcal{X}^t$ such that $x_t = x'_t$ and $\Xi^\pi_t(h_t) = \Xi^\pi_t(h'_t)$.*

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

$$v^\pi_t(h_t) := \rho^\pi_{t,T}\big(r_t(X_t, \Xi^\pi_t, \pi_t(H_t)), \ldots, r_T(X_T, \Xi^\pi_T, \pi_T(H_T))\big)(h_t). \tag{5.15}$$

Definition 5.3.6. *A family of risk filters $\{\rho^\pi_{t,T}\}^{\pi\in\Pi}_{t=1,\ldots,T}$ for a POMDP is* Markov *if for all Markov policies $\pi \in \Pi$, for all bounded measurable $r = (r_1, \ldots, r_T)$, and for all $h_t = (x_1, \ldots, x_t)$ and $h'_t = (x'_1, \ldots, x'_t)$ in $\mathcal{X}^t$ such that $x_t = x'_t$ and $\Xi^\pi_t(h_t) = \Xi^\pi_t(h'_t)$, we have*

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

Proposition 5.3.7. *A normalized, translation invariant, and stochastically conditionally time-consistent family of risk filters $\{\rho^\pi_{t,T}\}^{\pi\in\Pi}_{t=1,\ldots,T}$ is Markov if and only if the dependence of σ_t on h_t is carried by $(x_t, \Xi^\pi_t(h_t))$ only, for all $t = 1, \ldots, T-1$.*

Proof. Fix $t = 1, \dots, T-1$ and $w \in \mathcal{V}$. Let $\pi \in \mathit{\Pi}$ be an arbitrary policy. Consider $h_t, h_t' \in \mathcal{X}^t$ such that $x_t = x_t'$, $\mathit{\Xi}_t^\pi(h_t) = \mathit{\Xi}_t^\pi(h_t') = \xi_t$, and $Q_t^\pi(h_t) = Q_t^\pi(h_t')$. By the measurable selector Theorem [38], a Markov policy $\lambda \in \mathit{\Pi}$ exists, such that $\pi_t(h_t) = \lambda_t(x_t, \xi_t)$ (for the fixed t and h_t).

By setting $r = (0, \dots, 0, r_{t+1}, 0, \dots, 0)$ with $r_{t+1}(x', \xi', u') \equiv w(x')$, we obtain from the Markov property in Definition 5.3.6:

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

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

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

□

The following theorem summarizes our observations.

Theorem 5.3.8. *A family of risk filters* $\{\rho_{t,T}^\pi\}_{t=1,\dots,T}^{\pi \in \mathit{\Pi}}$ *for a POMDP is normalized, translation-invariant, stochastically conditionally time-consistent, and Markov if and only if transition risk mappings*

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

exist, such that

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

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

$$v_t^\pi(h_t) = r_t(x_t, \xi_t, \pi_t(h_t)) + \sigma_t\big(x_t, \mathit{\Xi}_t^\pi(h_t), Q_t^\pi(h_t), v_{t+1}^\pi(h_t, \cdot)\big). \tag{5.16}$$

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

Corollary 5.3.9. *Under the conditions of Theorem 5.3.8, for any Markov policy π, the function (5.15) depends on r, $\pi_t,\dots,\pi_T$, and (x_t,ξ_t) only, and the following relation is true:*

$$
\begin{aligned}
&v_t^{\pi_t,\dots,\pi_T}(x_t,\xi_t) = r_t(x_t,\xi_t,\pi_t(x_t,\xi_t)) + \\
&\sigma_t\Big(x_t,\xi_t,\int_{\mathcal{Y}} K_t^X(x_t,y,\pi_t(x_t,\xi_t))\,\xi_t(dy), x' \mapsto v_{t+1}^{\pi_{t+1},\dots,\pi_T}(x',\varPhi_t(x_t,\xi_t,\pi_t(x_t,\xi_t),x'))\Big).
\end{aligned}
\tag{5.17}
$$

Proof. We use induction backward in time. For $t = T$ we have $v_T^{\pi}(h_T) = r_T(X_T,\xi_T,\pi_T(X_T,\xi_T))$ and our assertion is true. If it is true for $t+1$, formula (5.16) reads

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

Substitution of (5.12) proves our assertion.

□

5.4 Dynamic Programming

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

$$
\min_{\pi\in\varPi} v_1^{\pi}(x_1,\varXi_1(x_1)), \quad x_1 \in \mathcal{X}.
$$

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

$$
v_t^*(g_t) = \inf_{\pi\in\varPi_{t,T}(g_t)} v_t^{\pi}(h_t), \quad g_t = (x_1,u_1,\dots,u_{t-1},x_t) \in \mathcal{G}_t, \quad t = 1,\dots,T, \tag{5.18}
$$

where $\mathcal{G}_t$ is defined in (5.1) and $\varPi_{t,T}(g_t)$ is the set of feasible deterministic policies π such that $\pi_s(h_s) = u_s$, $\forall s = 1\dots t-1$. By Theorem 5.3.8, transition risk mappings $\big\{\sigma_t\big\}_{t=1,\dots,T-1}$ exist, such that equations (5.16) hold.

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

We also assume that the kernels $K_t(x,y,u)$ have densities $q_t(\cdot,\cdot \mid x,y,u)$ with respect to a finite product measure $\mu_X \otimes \mu_Y$ on $\mathcal{X} \times \mathcal{Y}$, as in Example 5.3.1. In this case,

$$\left[\int_{\mathcal{Y}} K_t^X(x,y,u)\,\xi(dy)\right](dx') = \left[\int_{\mathcal{Y}}\int_{\mathcal{Y}} q_t(x',y' \mid x,y,u)\,\xi(dy)\,\mu_Y(dy')\right]\mu_X(dx'). \tag{5.19}$$

Our main result is that the value functions (5.18) are *memoryless*, that is, they depend on (x_t,ξ_t) only, and that they satisfy a generalized form of a dynamic programming equation. The equation also allows us to identify the optimal policy.

We remark that we cannot mechanically apply earlier results in Section 4.2 on fully observable Markov models and new techniques are required to prove the result. The difficulty is in the composite nature of the transition risk mappings, where the Bayes operator features.

Theorem 5.4.1. *We assume the following conditions:*

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

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

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

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

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

$$v_T^*(x,\xi) = \min_{u\in\mathcal{U}_T(x)} r_T(x,\xi,u), \quad x\in\mathcal{X}, \quad \xi\in\mathcal{P}(\mathcal{X}),$$

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

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

$$\hat{\pi}_T(x,\xi) \in \operatorname*{argmin}_{u\in\mathcal{U}_T(x)} r_T(x,\xi,u), \quad x\in\mathcal{X}, \quad \xi\in\mathcal{P}(\mathcal{Y}), \tag{5.20}$$

$$\hat{\pi}_t(x,\xi) \in \operatorname*{argmin}_{u\in\mathcal{U}_t(x)} \Big\{ r_t(x,\xi,u) + \sigma_t\Big(x,\xi,\int_{\mathcal{Y}} K_t^X(x,y,u)\,\xi(dy), x' \mapsto v_{t+1}^*\big(x',\Phi_t(x,\xi,u,x')\big)\Big)\Big\},$$
$$x\in\mathcal{X},\quad \xi\in\mathcal{P}(\mathcal{Y}),\quad t=T-1,\dots,1. \tag{5.21}$$

Proof. For all $g_T \in \mathcal{G}_T$ we have

$$v_T^*(g_T) = \inf_{\pi_T(g_T)} r_T(x_T,\xi_T,\pi_T(h_T)) = \inf_{u\in\mathcal{U}_T(x_T)} r_T(x_T,\xi_T,u). \tag{5.22}$$

By assumptions (iii) and (iv), owing to the Berge theorem (see [5, Theorem 1.4.16]), the infimum in (5.22) is attained and is a lower semicontinuous function of (x_T,ξ_T). Hence, v_T^* is memoryless. Moreover, the optimal solution mapping $\Psi_T(x,\xi) = \big\{u\in\mathcal{U}_T(x) : r_T(x,\xi,u) = v_T^*(x,\xi)\big\}$ has nonempty and closed values and is measurable. Therefore, a measurable selector $\hat{\pi}_T$ of Ψ_T exists (see, [38], [5, Thm. 8.1.3]), and

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

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

$$v_{t+1}^*(g_{t+1}) = v_{t+1}^*(x_{t+1},\xi_{t+1}) = v_{t+1}^{\{\hat{\pi}_{t+1},\dots,\hat{\pi}_T\}}(x_{t+1},\xi_{t+1}),\quad \forall g_{t+1}\in\mathcal{G}_{t+1}.$$

Then for any $g_t\in\mathcal{G}_t$ formula (5.16), after substituting (5.12), yields

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

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

$$\begin{aligned} &v_t^*(g_t) \\ &\ge \inf_{\pi\in\Pi_{t,T}(g_t)} \Big\{ r_t(x_t,\xi_t,\pi_t(h_t)) + \\ &\qquad \sigma_t\Big(x_t,\xi_t,\int_{\mathcal{Y}} K_t^X(x_t,y,\pi_t(h_t))\,\xi_t(dy), x'\mapsto v_{t+1}^*(x',\Phi_t(x_t,\xi_t,\pi_t(h_t),x'))\Big)\Big\} \\ &= \inf_{u\in\mathcal{U}_t(x_t)} \Big\{ r_t(x_t,\xi_t,u) + \\ &\qquad \sigma_t\Big(x_t,\xi_t,\int_{\mathcal{Y}} K_t^X(x_t,y,u)\,\xi_t(dy), x'\mapsto v_{t+1}^*(x',\Phi_t(x_t,\xi_t,u,x'))\Big)\Big\}. \end{aligned} \tag{5.23}$$

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

$$(x,\xi,u) \mapsto \sigma_t\Big(x,\xi,\int_{\mathcal{Y}} K_t^X(x,y,u)\,\xi(dy), x' \mapsto v_{t+1}^*\big(x',\Phi_t(x,\xi,u,x')\big)\Big). \tag{5.24}$$

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

First, we verify that the mapping $(x,\xi,u) \mapsto \int_{\mathcal{Y}} K_t^X(x,y,u)\,\xi(dy)$ appearing in the third argument of σ_t is weakly continuous. By formula (5.19), for any bounded continuous function $f : \mathcal{X} \to \mathbb{R}$ we have

$$\begin{aligned} &\int_{\mathcal{X}} f(x') \left[\int_{\mathcal{Y}} K_t^X(x^{(k)},y,u^{(k)})\;\xi^{(k)}(dy)\right](dx') \\ &= \int_{\mathcal{X}} f(x') \left[\int_{\mathcal{Y}}\int_{\mathcal{Y}} q_t(x',y' \mid x^{(k)},y,u^{(k)})\,\xi^{(k)}(dy)\,\mu_Y(dy')\right]\mu_X(dx') \end{aligned} \tag{5.25}$$

By assumption (i),

$$\lim_{k\to\infty} \int_{\mathcal{Y}} \big[q_t(x',y' \mid x^{(k)},y,u^{(k)}) - q_t(x',y' \mid x,y,u)\big]\xi^{(k)}(dy) = 0, \tag{5.26}$$

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

$$y \mapsto \int_{\mathcal{X}} f(x') \;\int_{\mathcal{Y}} q_t(x',y' \mid x,y,u)\;\mu_Y(dy')\;\mu_X(dx') \tag{5.27}$$

is continuous. Therefore, combining (5.25) and (5.26), we obtain the chain of equations:

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

The last by one equation follows from the weak convergence of $\xi^{(k)}$ to ξ and from the continuity of the function (5.27). Thus, the third argument of σ_t in (5.24) is continuous with respect to (x,ξ,u).

Let us examine the last argument of σ_t in (5.24). By (5.26), for every continuous bounded function $f(\cdot)$ on $\mathcal{Y}$, and for each fixed $x' \in \mathcal{X}$,

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

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

$$\int_{\mathcal{Y}} \int_{\mathcal{Y}} q_t(x', y' \mid x, y, u)\; \xi(dy)\; \mu_Y(dy') > 0. \tag{5.28}$$

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

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

Since $v^*_{t+1}(\cdot, \cdot)$ is lower-semicontinuous and $\varPhi_t(\cdot, \cdot, \cdot, x')$ is continuous, whenever condition (5.28) is satisfied, we infer that

$$V(x') \le \liminf_{k\to\infty}\; V^{(k)}(x'),$$

at all $x' \in \mathcal{X}$ at which (5.28) holds. As v^*_{t+1} and $\varPhi_t$ are measurable, both V and $\liminf_{k\to\infty} V^{(k)}$ are measurable as well.

By Theorem 5.3.8, the mapping σ_t is preserving the stochastic order $\preceq_{\text{st}}$ of the last argument with respect to the measure $\int_{\mathcal{Y}} K^X_t(x, y, u)\; \xi(dy)$. Since

$$\left(\int_{\mathcal{Y}} K^X_t(x, y, u)\; \xi(dy)\right) \left\{ x' \in \mathcal{X} : \int_{\mathcal{Y}} \int_{\mathcal{Y}} q_t(x', y' \mid x, y, u)\; \xi(dy)\; \mu_Y(dy') = 0 \right\} = 0,$$

the value of $\liminf_{k\to\infty}\; V^{(k)}(x')$ at the set of x' at which (5.28) is violated, is irrelevant. Consequently, by assumption (ii), with the view at the already established continuity of the third argument, we obtain the following chain of relations:

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

Consequently, the mapping (5.24) is lower semicontinuous.

Using assumptions (ii) and (iv) and invoking the Berge theorem again (see, e.g., [5, Theorem 1.4.16]), we deduce that the infimum in (5.23) is attained and is a lower semicontinuous function of (x_t,ξ_t). Moreover, the optimal solution mapping, that is, the set of $u\in\mathcal{U}_T(x)$ at which the infimum in (5.23) is attained, is nonempty, closed-valued, and measurable. Therefore, a minimizer $\hat{\pi}_t$ in (5.23) exists and is a measurable function of (x_t,ξ_t) (see, *e.g.*, [38], [5, Thm. 8.1.3]). Substituting this minimizer into (5.23), we obtain

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

In the last equation, we used Corollary 5.3.9. On the other hand, we have

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

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

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

This completes the induction step.

□

The most essential assumption of Theorem 5.4.1 is assumption (ii) of the lower semicontinuity of the transition risk mappings $\sigma_t(\cdot,\cdot,\cdot,\cdot)$. If these mappings are derived from convex or coherent risk measures, their lower semicontinuity with respect to the last argument follows from the corresponding property of the risk measure. In particular, [56, Cor. 3.1] derives continuity from monotonicity on Banach lattices. The semicontinuity with respect to the third argument, the probability measure, is a more complex issue. Lemmas 4.2.3 and 4.2.4, verified this condition for two popular risk measures: the Average Value at Risk and the mean-semideviation measure. Similar remarks apply to the assumption (iii) about the running risk functions. The assumptions (i) and (iv) are the same as in the utility models of [8].

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

Chapter 6

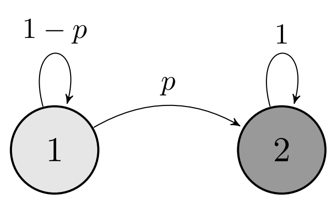
Illustration: Machine Deterioration

6.1 Description of the Process

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

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

The dynamics is Markovian, with the following transition graph for the "continue" control:



and the following transition matrices $K^{[u]}$, $u \in \{0, 1\}$:

$$
\begin{aligned}
K^{[0]} &= \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}, \\
K^{[1]} &= \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}.
\end{aligned}
\tag{6.1}
$$

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

$$\begin{aligned} \mathbb{P}\big[x_{t+1} \le C \,\big|\, y_t = i, u_t = 0\big] &= \int_{-\infty}^{C} f_i(x)\,dx, \quad i = 1,2, \\ \mathbb{P}\big[x_{t+1} \le C \,\big|\, y_t = i, u_t = 1\big] &= \int_{-\infty}^{C} f_1(x)\,dx, \quad i = 1,2. \end{aligned} \tag{6.2}$$

Assumption 6.1.1. *The functions f_1 and f_2 are uniformly bounded and the conditional distribution of x_{t+1} given that the machine is in "good" condition is stochastically smaller than the conditional distribution of x_{t+1} given that the machine is in "bad" condition,* i.e.,

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

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

Thus the relations (6.1) and (6.2) define $\big\{x_t, y_t\big\}_{t=1,\ldots,T+1}$ as a partially observable Markov process controlled by $\big\{u_t\big\}_{t=1,\ldots,T}$. Based on observations $(x_1,\ldots,x_t)$, the belief state $\xi_t \in [0,1]$ denotes the conditional probability that $y_t = 1$. We can update the posterior belief state as follows:

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

where Φ is the Bayes operator,

$$\Phi(\xi, x') = \frac{(1-p)\xi f_1(x')}{\xi f_1(x') + (1-\xi) f_2(x')}. \tag{6.3}$$

We assume that the initial probability $\xi_0 \in [0,1]$ is known; then $\xi_1(x_1) = \Phi(\xi_0, x_1)$. Directly from (6.3) we see that $\Phi(0,\cdot\,) = 0$, $\Phi(1,\cdot\,) = 1-p$, and $\Phi(\,\cdot\,, x')$ is nondecreasing.

6.2 Risk Modeling

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

$$\begin{cases} r_t(x_t, u_t) = R \cdot u_t + x_t, \quad t = 1,\ldots,T; \\ r_{T+1}(x_{T+1}) = x_{T+1}. \end{cases} \tag{6.4}$$

We denote the history of observations by $h_t = (x_1, \dots, x_t)$ and the set of all history-dependent policies by

$$\mathit{\Pi} := \big\{\, \pi = (\pi_1, \dots, \pi_T) \,|\, \forall t,\ \pi_t(x_1, \dots, x_t) \in \{0, 1\} \,\big\}.$$

We want to evaluate the costs (6.4) for any $\pi \in \mathit{\Pi}$, and find an optimal policy. The risk-neutral approach is to evaluate the conditional expectations of the sum of future costs:

$$\mathbb{E}\bigg[\sum_{\tau=t}^{T} \big(R \cdot \pi_\tau(h_\tau) + x_\tau\big) + x_{T+1} \,\bigg|\, h_t\bigg], \quad h_t \in \mathbb{R}^t, \quad t = 1, \dots, T+1, \quad \pi \in \mathit{\Pi}.$$

As shown in Theorem 5.3.8, construction of Markovian risk measures that replace the above expectations, is equivalent to specifying transition risk mappings

$$\sigma_t : \mathbb{R} \times \mathcal{P}(\mathbb{R}) \times \mathcal{P}(\mathbb{R}) \times \mathcal{V} \to \mathbb{R},$$

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

$$\sigma_t(x, \xi, q, v) = \sigma(q, v).$$

Remark 6.2.1. *For a probability measure $q \in \mathcal{P}(\mathbb{R})$ that has f as the density function, with slight abuse of notation, we also write $\sigma(f, \cdot)$ instead of $\sigma(q, \cdot)$.*

6.3 Value and Policy Monotonicity

In this section, we assume that the transition risk mapping $\sigma : \mathcal{P}(\mathbb{R}) \times \mathcal{V} \to \mathbb{R}$ satisfies all assumptions of Theorem 5.4.1. Then the optimal value functions v_t^*, $t = 1, \dots, T+1$ are memoryless and satisfy the following dynamic programming equations:

$$v_t^*(x, \xi) = x + \min \begin{Bmatrix} R + \sigma\big(f_1, x' \mapsto v_{t+1}^*(x', 1-p)\big); \\ \sigma\big(\xi f_1 + (1-\xi) f_2, x' \mapsto v_{t+1}^*(x', \mathit{\Phi}(\xi, x'))\big) \end{Bmatrix},$$

$$x \in \mathbb{R}, \quad \xi \in [0, 1], \quad t = 1, \dots, T, \tag{6.5}$$

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

Directly from (6.5) we see that $v_t^*(x,\xi) = x + w_t^*(\xi)$, $t = 1,\dots,T+1$. The dynamic programming equations (6.5) simplify as follows:

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

$$\xi \in [0,1], \quad t = 1,\dots,T, \quad (6.6)$$

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

Theorem 6.3.1. *If $\frac{f_1}{f_2}$ is non-increasing, then the functions $w_t^* : [0,1] \to \mathbb{R}$, $t = 1,\dots,T+1$ are non-increasing.*

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

1. $\xi_1 f_1 + (1-\xi_1) f_2 \succeq_{\text{st}} \xi_2 f_1 + (1-\xi_2) f_2$, because $f_1 \preceq_{\text{st}} f_2$.

2. For all x', we have $x' + w_{t+1}^*(\Phi(\xi_1, x')) \ge x' + w_{t+1}^*(\Phi(\xi_2, x'))$, as w_{t+1}^* is non-increasing and $\Phi(\cdot, x')$ is non-decreasing.

3. the mapping $x' \mapsto x' + w_{t+1}^*(\Phi(\xi, x'))$ is non-decreasing for all ξ. To show that, it is sufficient to establish that $x' \mapsto \Phi(\xi, x')$ is non-increasing, and this can be seen from the formula (for $0 \le p < 1$):

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

Thus

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

which completes the induction step.

□

The monotonicity assumption on $\frac{f_1}{f_2}$ is in fact a sufficient (but not necessary) condition for $f_1 \preceq_{\text{st}} f_2$. We illustrate this issue by the following examples.

Example 6.3.2 (Exponentially distributed costps)**.** *We have*

$$f_i(x) = \frac{1}{\theta_i} \exp\left(-\frac{x - m_i}{\theta_i}\right) \mathbb{1}_{\{x \geq m_i\}}, \quad i = 1, 2,$$

with m_1, m_2, θ_1, $\theta_2 \geq 0$. *To have* $f_1 \preceq_{\mathrm{st}} f_2$, *we must have* $m_1 \leq m_2$ *and* $\theta_1 \geq \theta_2$. *Thus the non-increasing property of* $\frac{f_1}{f_2}$ *is equivalent to* $f_1 \preceq_{\mathrm{st}} f_2$, *as*

$$\frac{f_1(x)}{f_2(x)} = \begin{cases} +\infty, & \text{for } x < m_2, \quad \text{if } m_1 < m_2; \\ 0, & \text{for } x < m_1, \quad \text{if } m_2 < m_1; \\ \frac{\theta_2}{\theta_1} \exp\left(\frac{x - m_2}{\theta_2} - \frac{x - m_1}{\theta_1}\right), & \text{for } x \geq \max(m_1, m_2). \end{cases}$$

Example 6.3.3 (Normally distributed costs)**.** *Consider the normal distributions truncated to* $[0, +\infty)$

$$f_i(x) = \frac{\exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)}{\int_0^\infty \exp\left(-\frac{1}{2}\left(\frac{z-\mu_i}{\sigma_i}\right)^2\right) dz}, \quad i = 1, 2,$$

where $\sigma_1, \sigma_2 > 0$ *and* $\mu_1, \mu_2 \in \mathbb{R}$. *We have*

$$\frac{f_1(x)}{f_2(x)} = const \cdot \exp\left(-\frac{1}{2}\left(\frac{x - \mu_1}{\sigma_1}\right)^2 + \frac{1}{2}\left(\frac{x - \mu_2}{\sigma_2}\right)^2\right),$$

so the derivative of $\ln \frac{f_1}{f_2}$ *is* $-\frac{x-\mu_1}{\sigma_1^2} + \frac{x-\mu_2}{\sigma_2^2}$. *A necessary and sufficient condition for* $\frac{f_1}{f_2}$ *being non-increasing on* $[0, +\infty)$ *is*

$$\sigma_1 \leq \sigma_2 \quad \text{and} \quad \frac{\mu_1}{\sigma_1^2} \leq \frac{\mu_2}{\sigma_2^2}.$$

Example 6.3.4 (Uniformly distributed costs)**.** *We set*

$$f_i = \frac{1}{M_i - m_i} \mathbb{1}_{[m_i, M_i]}, \quad i = 1, 2.$$

To have $f_1 \preceq_{\mathrm{st}} f_2$, *we must have* $m_1 \leq m_2$ *and* $M_1 \leq M_2$, *and then*

$$\frac{f_1(x)}{f_2(x)} = \begin{cases} +\infty, & \text{for } m_1 \leq x \leq m_2; \\ \frac{M_2 - m_2}{M_1 - m_1}, & \text{for } m_2 < x \leq M_1; \\ 0, & \text{for } x \geq M_2, \end{cases}$$

is non-increasing.

From Theorem 6.3.1 we obtain the following threshold property of the policy.

Theorem 6.3.5. *Under the assumptions of Theorem 6.3.1, there exist thresholds* $\xi_t^* \in [0,1]$, $t = 1,\dots,T$ *such that the policy*

$$u_t^* = \begin{cases} 0 & \text{if } \xi_t > \xi_t^*, \\ 1 & \text{if } \xi_t \le \xi_t^*, \end{cases}$$

is optimal.

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

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

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

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

and replacement is optimal for ζ as well.

□

6.4 Numerical Illustration

In this section, we solve the problem in the special case where f_1 and f_2 are density functions $\mathbb{U}(m_1, M_1)$ and $\mathbb{U}(m_2, M_2)$ with $m_1 \le m_2 \le M_1 \le M_2$, as in Example 6.3.4. Then the Bayes operator is piece-wise constant with respect to x':

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

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

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

We use the mean-semideviation transition risk mapping of Example 2.4.8, with $p = 1$ and constant $\varkappa$, i.e,

$$\sigma(q, v) = \mathbb{E}_q[v] + \varkappa \mathbb{E}_q\left[(v - \mathbb{E}_q(v))^+\right].$$

It is strongly monotonic with respect to stochastic order and lower semi-continuous with respect to (q, v). Then the dynamic programming equations (6.6) for $t = 1, \dots, T$ become:

$$\begin{aligned} w_t^*(\xi) = \min \Big\{ & R + E_t^*(1) + \mathbb{E}_{f_1}\big(x' \mapsto x' + w_{t+1}^*(1-p) - E_t^*(1)\big)_+; \\ & E_t^*(\xi) + \mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w_{t+1}^*(\varPhi(\xi, x')) - E_t^*(\xi)\big)_+ \Big\}, \end{aligned} \tag{6.7}$$

where

$$\begin{aligned} E_t^*(\xi) \quad &:= \quad \mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w_{t+1}^*(\varPhi(\xi, x'))\big) \\ &= \quad q_1(\xi)\left(\frac{m_1 + m_2}{2} + w_{t+1}^*(1-p)\right) \\ &\qquad + q_2(\xi)\left(\frac{m_2 + M_1}{2} + w_{t+1}^*(\hat{\phi}(\xi))\right) \\ &\qquad + q_3(\xi)\left(\frac{M_1 + M_2}{2} + w_{t+1}^*(0)\right). \end{aligned} \tag{6.8}$$

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

$$\begin{aligned} &\mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w_{t+1}^*(\varPhi(\xi, x')) - E_t^*(\xi)\big)_+ \\ &\quad = q_1(\xi)\, \theta\big(m_1, m_2, E_t^*(\xi) - w_{t+1}^*(1-p)\big) \\ &\qquad + q_2(\xi)\, \theta\big(m_2, M_1, E_t^*(\xi) - w_{t+1}^*(\hat{\phi}(\xi))\big) \\ &\qquad + q_3(\xi)\, \theta\big(M_1, M_2, E_t^*(\xi) - w_{t+1}^*(0)\big), \end{aligned} \tag{6.9}$$

where, for $a_1 \le a_2$,

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

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

1. Compute $E^*_t(1)$ and $E^*_t(\xi)$ by (6.8).

2. Compute $\mathbb{E}_{f_1}\big(x' \mapsto x' + w^*_{t+1}(1-p) - E^*_t(1)\big)_+$ and $\mathbb{E}_{\xi f_1 + (1-\xi) f_2}\big(x' \mapsto x' + w^*_{t+1}(\varPhi(\xi, x')) - E^*_t(\xi)\big)_+$ by (6.9).

3. Compute $w^*_t(\xi)$ using the dynamic programming equation (6.7).

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

In Figure 6.1, we present the value functions and optimal policies for an example with $m_1 = 0$, $m_2 = 80$, $M_1 = 100$, $M_2 = 500$, $p = 0.2$, $T = 6$, $R = 50$, and $\varkappa = 0.9$. In Figure 6.2, we display the distribution of the total cost obtained by simulating 100,000 runs of the system with both policies.

We see that the application of the risk-averse model increases the threshold values ξ^*_t of the optimal policies and results in a significantly less dispersed distribution of the total cost.

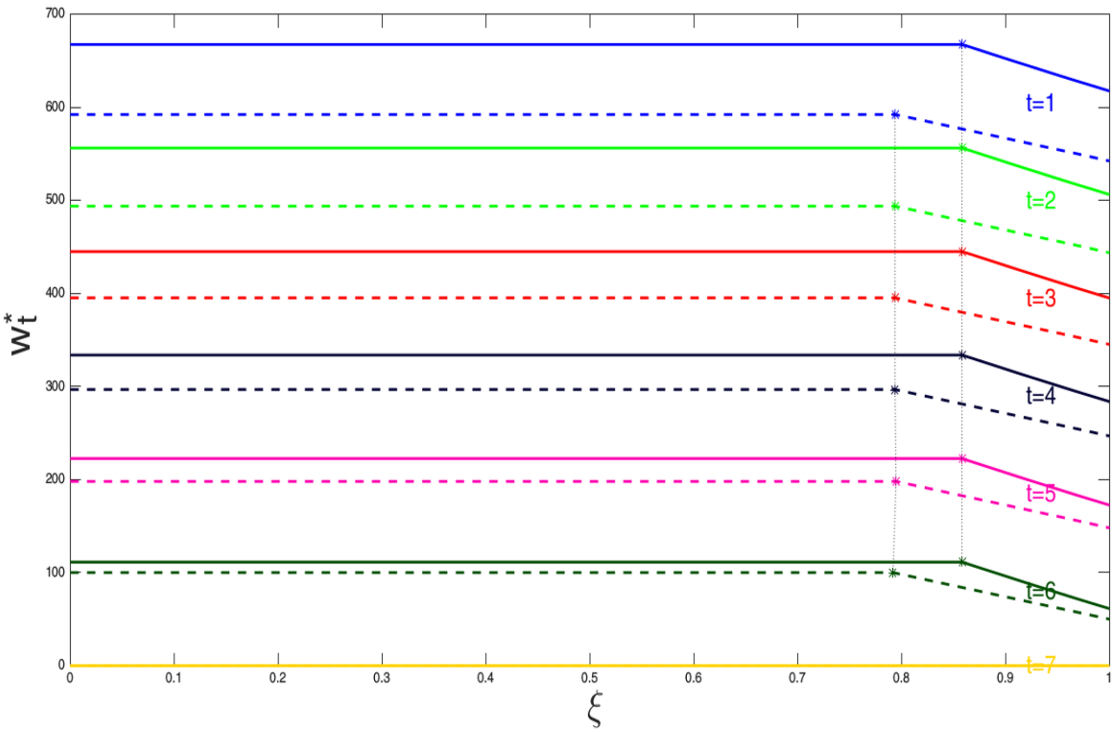


Figure 6.1: The value function for the risk-neutral model (dashed) and the risk-averse model (solid). The stars denote the critical value ξ^* below which replacement is optimal.

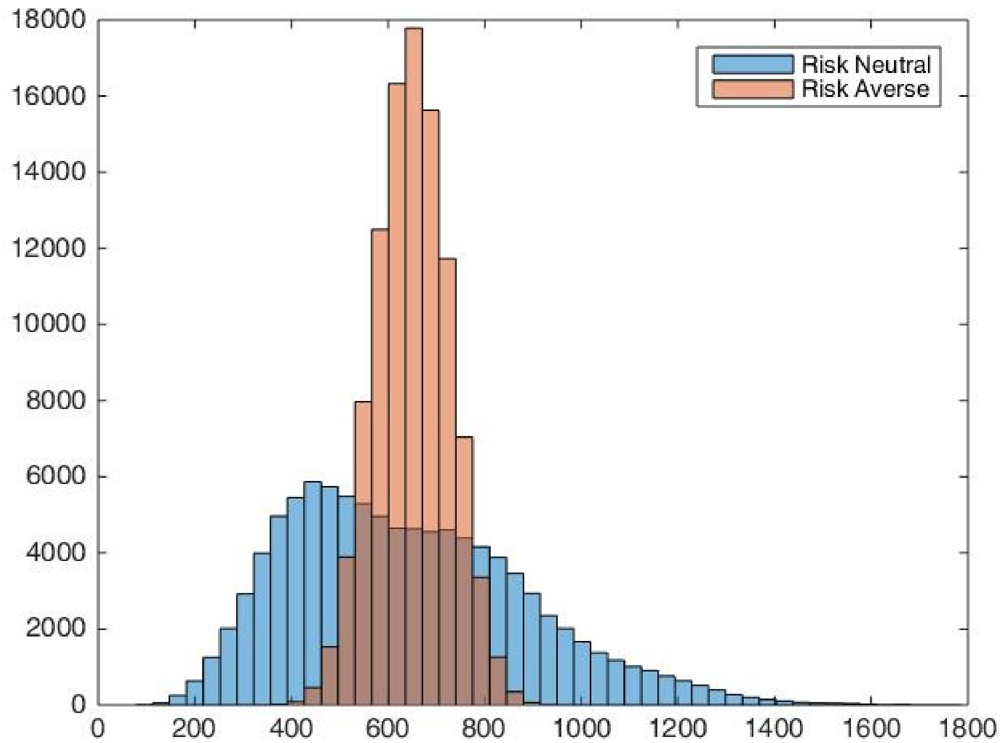


Figure 6.2: Empirical distribution of the total cost for the risk-neutral model (blue) and the risk-averse model (orange).

Conclusion

In this thesis, we developed theoretical foundations of the theory of dynamic risk measures for controlled stochastic processes, and the results were successfully applied and specialized to Markov decision processes (MDP) and partially observable Markov decision processes (POMDP).

We began with introducing a new class of dynamic risk measures for general discrete-time stochastic processes, which we call process-based. Under the uncontrolled setting, we introduced the notion of *stochastic conditional time consistency*, which is stronger than the classical notion of time consistency. We proved under general assumptions that a dynamic risk measure is stochastically conditionally time-consistent if and only if it can be represented by a collection of static law-invariant risk measures on the space of functions of the state of the base process. (These static law-invariant risk measures are called "transition risk mappings".) This full characterization allows us to construct such dynamic risk measures from static risk measures, while the static risk measures are mathematical objects that are much better understood and analyzed in the literature. Under the controlled setting, due to the presence of different control policies, we modified the notion of stochastic conditional time consistency so that it connects different policies into one unified property. In other words, the dynamic risk measures for all admissible policies share the same collection of transition risk mappings that is policy-independent.

The results were first specialized to Markov decision problems (MDP), where we evaluate different policies according to a process-based dynamic risk measure. We discovered that under the assumption that risk measures are memoryless under Markov policies, the history dependence of the transition risk mappings is reduced to the current state only. With the help of this special form of characterization, the dynamic programming equations governing the value function was derived. As a special case of MDP, for multistage stochastic programming problems with decision-dependent probabilities, we derived the form of risk measures on a scenario tree and we also derived the associated dynamic programming equations.

We then specialized our theory to partially observable Markov decision problems (POMDP). Due to the fact that part of the state is unobservable, the agent in POMDP needs to optimize the objective function while making "best guesses" on the unobservable part of the state. One of the major difficulties is that the cost process is adapted to the filtration generated by the full state process, while the risk measure needs to be assessed according to the filtration generated by the observable state process. To deal with this difficulty, we introduced the concept of a *risk filter*. We postulated the property of stochastic conditional time consistency adapted to such a filter. Our main result was that the risk filters can be equivalently modeled by special forms of transition risk mappings: static risk measures on the space of functions defined on the observable part of the state only. We also derived dynamic programming equations under this setting.

Finally, our theory on POMDP was applied to the machine deterioration problem. The problem was formulated as a cost minimization problem, where in each step we can choose to continue using the machine or to replace it. We can observe the operating cost, while the gradual deterioration of the machine is not directly observable. Our numerical results showcased the power of our risk-averse theory, in the sense that the dispersion of the total cost under the optimal policy noticeably decreased compared to the risk-neutral case.

References

[1] A. Arlotto, N. Gans, and J. M. Steele. Markov decision problems where means bound variances. *Operations Research*, 62(4):864–875, 2014.

[2] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Thinking coherently: Generalised scenarios rather than var should be used when calculating regulatory capital. *RISK-LONDON-RISK MAGAZINE LIMITED-*, 10:68–71, 1997.

[3] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[4] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, David Heath, and Hyejin Ku. Coherent multiperiod risk adjusted values and bellmans principle. *Annals of Operations Research*, 152(1):5–22, 2007.

[5] J.-P. Aubin and H. Frankowska. *Set-valued analysis*. Birkhäuser, Boston, MA, 2009.

[6] N. Bäuerle and U. Rieder. *Markov decision processes with applications to finance*. Universitext. Springer, Heidelberg, 2011.

[7] N. Bäuerle and U. Rieder. More risk-sensitive Markov decision processes. *Mathematics of Operations Research*, 39(1):105–120, 2013.

[8] N. Bäuerle and U. Rieder. Partially observable risk-sensitive Markov decision processes. *Mathematics of Operations Research*, 2017.

[9] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, New York-London, 1978.

[10] T. Bielecki, D. Hernández-Hernández, and S. R. Pliska. Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management. *Mathematical Methods of Operations Research*, 50(2):167–188, 1999.

[11] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, 2013.

[12] Ö. Çavus and A. Ruszczyński. Computational methods for risk-averse undiscounted transient Markov models. *Operations Research*, 62(2):401–417, 2014.

[13] Ö. Çavus and A. Ruszczyński. Risk-averse control of undiscounted transient Markov models. *SIAM Journal on Control and Optimization*, 52(6):3935–3966, 2014.

[14] Z. Chen, G. Li, and Y. Zhao. Time-consistent investment policies in Markovian markets: a case of mean-variance analysis. *J. Econom. Dynam. Control*, 40:293–316, 2014.

[15] P. Cheridito, F. Delbaen, and M. Kupper. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11:57–106, 2006.

[16] P. Cheridito and M. Kupper. Composition of time-consistent dynamic monetary risk measures in discrete time. *International Journal of Theoretical and Applied Finance*, 14(01):137–162, 2011.

[17] S. P. Coraluppi and S. I. Marcus. Risk-sensitive and minimax control of discrete-time, finite-state Markov decision processes. *Automatica*, 35(2):301–309, 1999.

[18] P. Dai Pra, L. Meneghini, and W. J. Runggaldier. Explicit solutions for multivariate, discrete-time control problems under uncertainty. *Systems & Control Letters*, 34(4):169–176, 1998.

[19] E. V. Denardo and U. G. Rothblum. Optimal stopping, exponential utility, and linear programming. *Math. Programming*, 16(2):228–244, 1979.

[20] G. B. Di Masi and Ł. Stettner. Risk-sensitive control of discrete-time Markov processes with infinite horizon. *SIAM J. Control Optim.*, 38(1):61–78, 1999.

[21] E. A. Feinberg, P. O. Kasyanov, and M. Z. Zgurovsky. Partially observable total-cost Markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research*, 41(2):656–681, 2016.

[22] E. Fernández-Gaucherand and S. I. Marcus. Risk-sensitive optimal control of hidden Markov models: Structural results. *IEEE Transactions on Automatic Control*, 42(10):1418–1422, 1997.

[23] J. A. Filar, L. C. M. Kallenberg, and H.-M. Lee. Variance-penalized Markov decision processes. *Math. Oper. Res.*, 14(1):147–161, 1989.

[24] H. Föllmer and I. Penner. Convex risk measures and the dynamics of their penalty functions. *Statistics & Decisions*, 24(1/2006):61–96, 2006.

[25] Hans Föllmer and Alexander Schied. Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447, 2002.

[26] Hans Föllmer and Alexander Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

[27] Marco Frittelli and Emanuela Rosazza Gianin. Putting order in risk measures. *Journal of Banking & Finance*, 26(7):1473–1486, 2002.

[28] K. Hinderer. *Foundations of non-stationary dynamic programming with discrete time parameter*. Lecture Notes in Operations Research and Mathematical Systems, Vol. 33. Springer-Verlag, Berlin-New York, 1970.

[29] R. A. Howard and J. E. Matheson. Risk-sensitive Markov decision processes. *Management Sci.*, 18:356–369, 1971/72.

[30] M. R. James, J. S. Baras, and R. J. Elliott. Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems. *IEEE transactions on automatic control*, 39(4):780–792, 1994.

[31] S. C. Jaquette. Markov decision processes with a new optimality criterion: discrete time. *Ann. Statist.*, 1:496–505, 1973.

[32] S. C. Jaquette. A utility criterion for Markov decision processes. *Management Sci.*, 23(1):43–49, 1975/76.

[33] A. Jaśkiewicz, J. Matkowski, and A. S. Nowak. Persistently optimal policies in stochastic dynamic programming with generalized discounting. *Mathematics of Operations Research*, 38(1):108–121, 2013.

[34] A. Jobert and L. C. G. Rogers. Valuations and dynamic convex risk measures. *Mathematical Finance*, 18(1):1–22, 2008.

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

[36] T. C. Koopmans. Stationary ordinal utility and impatience. *Econometrica*, pages 287–309, 1960.

[37] M. Kupper and W. Schachermayer. Representation results for law invariant time consistent functions. *Mathematics and Financial Economics*, 2(3):189–210, 2009.

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

[39] Shigeo Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.

[40] S. Levitt and A. Ben-Israel. On modeling risk in Markov decision processes. In *Optimization and related topics (Ballarat/Melbourne, 1999)*, volume 47 of *Appl. Optim.*, pages 27–40. Kluwer Acad. Publ., Dordrecht, 2001.

[41] K. Lin and S. I. Marcus. Dynamic programming with non-convex risk-sensitive measures. In *American Control Conference (ACC), 2013*, pages 6778–6783. IEEE, 2013.

[42] S. Mannor and J. N. Tsitsiklis. Algorithmic aspects of mean-variance optimization in Markov decision processes. *European J. Oper. Res.*, 231(3):645–653, 2013.

[43] S. I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernández, S. Coraluppi, and P. Fard. Risk sensitive Markov decision processes. In *Systems and control in the twenty-first century (St. Louis, MO, 1996)*, volume 22 of *Progr. Systems Control Theory*, pages 263–279. Birkhäuser, Boston, MA, 1997.

[44] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research*, 116(1):33–50, 1999.

[45] W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean–semideviation models. *Mathematical Programming*, 89(2):217–232, 2001.

[46] WLodzimierz Ogryczak and Andrzej Ruszczynski. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, 13(1):60–78, 2002.

[47] G.Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore, 2007.

[48] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and Their Applications*, 112:185–200, 2004.

[49] R T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer, Berlin, 1998.

[50] R Tyrrell Rockafellar. *Conjugate duality and optimization*. SIAM, 1974.

[51] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[52] R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

[53] B. Roorda, J. M. Schumacher, and J. Engwerda. Coherent acceptability measures in multiperiod models. *Mathematical Finance*, 15(4):589–612, 2005.

[54] W. J. Runggaldier. Concepts and methods for discrete and continuous time control under uncertainty. *Insurance: Mathematics and Economics*, 22(1):25–39, 1998.

[55] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31:544–561, 2006.

[56] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–542, 2006.

[57] Andrzej Ruszczyński. Risk-averse dynamic programming for markov decision processes. *Mathematical programming*, 125(2):235–261, 2010.

[58] G. Scandolo. *Risk Measures in a Dynamic Setting*. PhD thesis, Università degli Studi di Milano, Milan, Italy, 2003.

[59] A. Shapiro. Time consistency of dynamic risk measures. *Operations Research Letters*, 40(6):436–439, 2012.

[60] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on stochastic programming: modeling and theory (2nd ed.* SIAM, 2011.

[61] Y. Shen, W. Stannat, and K. Obermayer. Risk-sensitive Markov control processes. *SIAM Journal on Control and Optimization*, 51(5):3652–3672, 2013.

[62] S. Weber. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, 16(2):419–441, 2006.

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