RISK-AVERSE CONTROL OF UNDISCOUNTED TRANSIENT MARKOV MODELS[*]

ÖZLEM ÇAVUŞ[†] AND ANDRZEJ RUSZCZYŃSKI[‡]

Abstract. We use Markov risk measures to formulate a risk-averse version of the undiscounted total cost problem for a transient controlled Markov process. Using the new concept of a multikernel, we derive conditions for a system to be risk transient, that is, to have finite risk over an infinite time horizon. We derive risk-averse dynamic programming equations satisfied by the optimal policy and we describe methods for solving these equations. We illustrate the results on an optimal stopping problem and an organ transplantation problem.

Key words. dynamic risk measures, Markov risk measures, multikernels, stochastic shortest path, optimal stopping, randomized policy

AMS subject classification. 90C40

DOI. 10.1137/13093902X

1. Introduction. The optimal control problem for transient Markov processes is a classical model in operations research (see Veinott [50], Pliska [35], Bertsekas and Tsitsiklis [7], Hernández-Lerma and Lasserre [19], and the references therein). The research is focused on the expected total undiscounted cost model, with increased state and control space generality.

Our objective is to consider a risk-averse model. So far, risk-averse problems for transient Markov models were based on the arrival probability criteria (see, e.g., Nie and Wu [27] and Ohtsubo [29]) and utility functions (see Denardo and Rothblum [12] and Patek [33]). We plan to use the recent theory of dynamic risk measures (see Scandolo [45], Fritelli and Scandolo [16], Riedel [37], Ruszczyński and Shapiro [42, 44], Cheridito, Delbaen and Kupper [8], Artzner et al. [3], Klöppel and Schweizer [23], Pflug and Römisch [34], and the references therein) to develop and solve new risk-averse formulations of the stochastic optimal control problem for transient Markov models. Specific examples of such models are stochastic shortest path problems (Bertsekas and Tsitsiklis [7]) and optimal stopping problems (cf. Çınlar [11], Dynkin and Yushkevich [13, 14], Puterman [36]).

A systematic approach to Markov decision problems with coherent dynamic measures of risk was initiated by Ruszczyński [41], who considered risk-averse finite horizon and discounted infinite horizon models. This was further extended to nonconvex criteria by Lin and Marcus in [26]. Shen, Stannat, and Obermayer [48] considered risk-sensitive discounted and average cost models where the coherence assumptions were relaxed.

Some applications of stochastic shortest path problems concerned with expected performance criteria are given in the survey paper by White [52] and the references therein. However, in many practical problems, the expected values may not be ap-

[†]Department of Industrial Engineering, Bilkent University, Ankara, Turkey (ozlem.cavus@bilkent. edu.tr).

[‡]Department of Management Science and Information Systems, Rutgers University, Piscataway, NJ 08854 (rusz@rutgers.edu).

propriate to measure performance, because they implicitly assume that the decision maker is risk neutral. Below, we provide examples of such real-life problems which were modeled before as a discrete-time Markov decision process with expected value as the objective function. Alagoz et al. [1] suggested a discounted, infinite horizon, and absorbing Markov decision process model to find the optimal time of liver transplantation for a risk-neutral patient under the assumption that the liver is transplanted from a living donor. However, referring to Chew and Ho [9], they state that the risk neutrality of the patient is not a realistic assumption. Kurt and Kharoufeh [25] proposed a discounted, infinite horizon Markov decision process model for optimal replacement time of a system under Markovian deterioration and Markovian environment. So and Thomas [49] employed a discrete-time Markov decision process to model profitability of credit cards.

Our theory of risk-averse control problems for transient models applies to these and many other models. Our results complement and extend the results of Ruszczyński [41], where infinite-horizon *discounted* models were considered. We consider *undiscounted* models for *transient* Markov systems. The paper is organized as follows.

In section 2, we quickly review some basic concepts of controlled Markov models. In section 3, we adapt and extend our earlier theory of Markov risk measures. In section 4, we introduce and analyze the concept of a *multikernel* (a multivalued kernel), which is essential for our theory. General assumptions and technical issues associated with measurability of decision rules are discussed in section 5. Section 6 is devoted to the analysis of a finite horizon model. The main model with infinite horizon and dynamic risk measures is analyzed in section 7. We introduce in it the concept of a *risk-transient model* and develop equations for evaluating policies in such models. In section 8, we derive risk-averse versions of dynamic programming equations for risk-transient models. Section 9 compares randomized and deterministic polices. Finally, section 10 illustrates our results on risk-averse versions of an optimal stopping problem of Karlin [22] and of the organ transplantation problem of Alagoz et al. [1].

2. Controlled Markov processes. We quickly review the main concepts of controlled Markov models and we introduce relevant notation (for details, see [15, 18, 19]). Let $\mathscr{X}$ be a state space, and $\mathscr{U}$ a control space. We assume that $\mathscr{X}$ and $\mathscr{U}$ are Borel spaces (Borel subsets of Polish spaces), with Borel σ-algebras $\mathscr{B}(\mathscr{X})$ and $\mathscr{B}(\mathscr{U})$. A control set is a measurable multifunction $U : \mathscr{X} \rightrightarrows \mathscr{U}$; for each state $x \in \mathscr{X}$ the set $U(x) \subseteq \mathscr{U}$ is a nonempty set of possible controls at x. A controlled transition kernel Q is a measurable mapping from the graph of U to the set $\mathscr{P}(\mathscr{X})$ of probability measures on $\mathscr{X}$ (equipped with the topology of weak convergence); $Q(x, u)$ is a probability measure on $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$, for all $x \in \mathscr{X}$ and $u \in U(x)$.

The cost of transition from x to y, when control u is applied, is represented by $c(x, u, y)$, where $c : \mathscr{X} \times \mathscr{U} \times \mathscr{X} \to \mathbb{R}$. Only $u \in U(x)$ and those $y \in \mathscr{X}$ to which transition is possible matter here, but it is convenient to consider the function $c(\cdot, \cdot, \cdot)$ as defined on the product space.

A *stationary controlled Markov process* is defined by a state space $\mathscr{X}$, a control space $\mathscr{U}$, a control set U, a controlled transition kernel Q, and a cost function c.

For $t = 1, 2, \ldots$, we define the space of state and control histories up to time t as $\mathscr{H}_t = \mathrm{graph}(U)^{t-1} \times \mathscr{X}$. Each history is a sequence $h_t = (x_1, u_1, \ldots, x_{t-1}, u_{t-1}, x_t) \in \mathscr{H}_t$.

We denote by $\mathscr{P}(\mathscr{U})$ and $\mathscr{P}(U(x))$ the sets of probability measures on $\mathscr{U}$ and $U(x)$, respectively. A *randomized policy* is a sequence of measurable functions $\pi_t : \mathscr{H}_t \to \mathscr{P}(\mathscr{U})$, $t = 1, 2, \ldots$, such that $\pi_t(h_t) \in \mathscr{P}(U(x_t))$ for all $h_t \in \mathscr{H}_t$. In words, the

distribution of the control u_t is supported on a subset of the set of feasible controls $U(x_t)$. A *Markov policy* is a sequence of measurable functions $\pi_t : \mathscr{X} \to \mathscr{P}(\mathscr{U})$, $t = 1, 2, \ldots$, such that $\pi_t(x) \in \mathscr{P}(U(x))$ for all $x \in \mathscr{X}$. The function $\pi_t(\cdot)$ is called the *decision rule* at time t. A Markov policy is *stationary* if there exists a function $\pi : \mathscr{X} \to \mathscr{P}(\mathscr{U})$ such that $\pi_t(x) = \pi(x)$ for all $t = 1, 2, \ldots$, and all $x \in \mathscr{X}$. Such a policy and the corresponding decision rule are called *deterministic*, if for every $x \in \mathscr{X}$ there exists $u(x) \in U(x)$ such that the measure $\pi(x)$ is supported on $\{u(x)\}$. In this paper, we focus on deterministic policies.

Consider the canonical sample space $\Omega = \mathscr{X}^\infty$ with the product σ-algebra $\mathscr{F}$. Let P_1 be the initial distribution of the state $x_1 \in \mathscr{X}$. Suppose we are given a deterministic policy $\Pi = \{\pi_t\}_{t=1}^\infty$. The Ionescu Tulcea theorem (see, e.g., [6]) states that there exists a unique probability measure P^Π on $(\Omega, \mathscr{F})$ such that for every measurable set $B \subset \mathscr{X}$ and all $h_t \in \mathscr{H}_t$, $t = 1, 2, \ldots$,

$$P^\Pi(x_1 \in B) = P_1(B),$$
$$P^\Pi(x_{t+1} \in B \,|\, h_t) = Q\big(B \,|\, x_t, \pi_t(h_t)\big).$$

To simplify our notation, from now on we assume that the initial state x_1 is fixed. It will be obvious how to modify our results for a random initial state. For a stationary decision rule π, we write Q^π to denote the corresponding transition kernel.

Our interest is in *transient* Markov models. We assume that some *absorbing state* $x_A \in \mathscr{X}$ exists such that $Q\big(\{x_A\} | x_A, u\big) = 1$ and $c(x_A, u, x_A) = 0$ for all $u \in U(x_A)$. Thus, after the absorbing state is reached, no further costs are incurred.[1] To analyze such Markov models, it is convenient to consider the effective state space $\widetilde{\mathscr{X}} = \mathscr{X} \setminus \{x_A\}$, and the effective controlled substochastic kernel $\widetilde{Q}$ whose arguments are restricted to $\widetilde{\mathscr{X}}$ and whose values are nonnegative measures on $\widetilde{\mathscr{X}}$, so that $\widetilde{Q}\big(B | x, u\big) = Q\big(B | x, u\big)$, for all Borel sets $B \subset \widetilde{\mathscr{X}}$, all $x \in \widetilde{\mathscr{X}}$, and all $u \in U(x)$.

Our point of departure is the *expected total cost problem*, which is to find a policy $\Pi = \{\pi_t\}_{t=1}^\infty$ so as to minimize the expected cost until absorption:

$$\min_\Pi \; \mathbb{E}^\Pi \left[\sum_{t=1}^\infty c(x_t, u_t, x_{t+1}) \right].$$

Here $\mathbb{E}^\Pi\big[\,\cdot\,\big]$ denotes the expected value with respect to the measure P^Π. Under appropriate assumptions, the problem has a solution in the form of a stationary Markov policy (see, e.g., [19, section 9.6]). The optimal policy can be found by solving appropriate dynamic programming equations.

Our intention is to introduce risk aversion to the problem, and to replace the expected value operator by a dynamic risk measure. We do not assume that the costs are nonnegative, and thus our approach applies also, among others, to stochastic longest path problems and optimal stopping problems with positive rewards.

3. Markov risk measures. Suppose T is a fixed time horizon. Each policy $\Pi = \{\pi_1, \pi_2, \ldots\}$ results in a cost sequence $Z_t = c(x_{t-1}, u_{t-1}, x_t)$, $t = 2, \ldots, T+1$, on the probability space $(\Omega, \mathscr{F}, P^\Pi)$. We define the σ-subalgebras $\mathscr{F}_t$ on $\mathscr{X}^t$, and vector spaces $\mathscr{Z}_t^\Pi$ of $\mathscr{F}_t$-measurable random variables on Ω, $t = 1, \ldots, T$.

---

[1]The case of a larger class of absorbing states easily reduces to the case of one absorbing state.

To evaluate the risk of this sequence we use a dynamic time-consistent risk measure of the following form:

$$
\begin{aligned}
J_T(\Pi, x_1) = \rho_1^{\Pi}\Big(& c(x_1, \pi_1(x_1), x_2) + \rho_2^{\Pi}\Big(c(x_2, \pi_2(x_2), x_3) + \cdots \\
& + \rho_{T-1}^{\Pi}\big(c(x_{T-1}, \pi_{T-1}(x_{T-1}), x_T) \\
& + \rho_T^{\Pi}(c(x_T, \pi_T(x_T), x_{T+1}))\cdots\Big)\Big).
\end{aligned}
$$

(3.1)

Here, $\rho_t^{\Pi} : \mathscr{Z}_{t+1}^{\Pi} \to \mathscr{Z}_t^{\Pi}$, $t = 1, \ldots, T$, are one-step conditional risk measures. Ruszczyński [41, section 3] derives the nested formulation (3.1) from general properties of monotonicity and time consistency of dynamic measures of risk.

It is convenient to introduce vector spaces $\mathscr{Z}_{t,\theta}^{\Pi} = \mathscr{Z}_t^{\Pi} \times \mathscr{Z}_{t+1}^{\Pi} \times \cdots \times \mathscr{Z}_\theta^{\Pi}$, where $1 \le t \le \theta \le T+1$ and the conditional risk measures $\rho_{t,\theta}^{\Pi} : \mathscr{Z}_{t,\theta}^{\Pi} \to \mathscr{Z}_t^{\Pi}$ is defined as follows:

$$
(3.2) \qquad \rho_{t,\theta}^{\Pi}(Z_t, \ldots, Z_\theta) = Z_t + \rho_t^{\Pi}\Big(Z_{t+1} + \rho_{t+1}^{\Pi}\big(Z_{t+2} + \cdots + \rho_{\theta-1}^{\Pi}(Z_\theta)\cdots\big)\Big).
$$

As indicated in [41], the fundamental difficulty of formulation (3.1) is that at time t the value of $\rho_t^{\Pi}(\cdot)$ is $\mathscr{F}_t$-measurable and is allowed to depend on the entire history h_t of the process. In order to overcome this difficulty, in [41, section 4] a new construction of a one-step conditional measure of risk was introduced. Its arguments are functions on the state space $\mathscr{X}$, rather than on the probability space Ω. We adapt this construction to our case, with a slightly more general form of the cost function.

Let $\mathscr{V} = \mathscr{L}_p(\mathscr{X}, \mathscr{B}, P_0)$, where $\mathscr{B}$ is the σ-field of Borel sets on $\mathscr{X}$, P_0 is some reference probability measure on $\mathscr{X}$, and $p \in [1, \infty)$. It is convenient to think of the dual space $\mathscr{V}'$ as the space of signed measures m on $(\mathscr{X}, \mathscr{B})$, which are absolutely continuous with respect to P_0, with densities (Radon–Nikodym derivatives) lying in the space $\mathscr{L}_q(\mathscr{X}, \mathscr{B}, P_0)$, where $1/p + 1/q = 1$. We make the following general assumption.

(G0) For all $x \in \mathscr{X}$ and $u \in U(x)$ the probability measure $Q(x, u)$ is an element of $\mathscr{V}'$.

In the case of finite state and control spaces P_0 may be the uniform measure; in other cases P_0 should be chosen in such a way that condition (G0) is satisfied. The existence of the measure P_0 is essential for the pairing of $\mathscr{V}$ and its dual space $\mathscr{V}'$, as discussed below.

We consider the set of probability measures in $\mathscr{V}'$:

$$
\mathscr{M} = \{m \in \mathscr{V}' : m(\mathscr{X}) = 1,\ m \ge 0\}.
$$

We also assume that the spaces $\mathscr{V}$ and $\mathscr{V}'$ are endowed with topologies that make them paired topological vector spaces with the bilinear form

$$
\langle \varphi, m \rangle = \int_{\mathscr{X}} \varphi(y)\, m(dy), \quad \varphi \in \mathscr{V}, \quad m \in \mathscr{V}'.
$$

The space $\mathscr{V}'$ (and thus $\mathscr{M}$) will be endowed with the weak* topology. We may endow $\mathscr{V}$ with the strong (norm) topology, or with the weak topology.

DEFINITION 3.1. *A measurable function* $\sigma : \mathscr{V} \times \mathscr{X} \times \mathscr{M} \to \mathbb{R}$ *is a* transition risk mapping *if for every* $x \in \mathscr{X}$ *and every* $m \in \mathscr{M}$, *the function* $\varphi \mapsto \sigma(\varphi, x, m)$ *is a coherent measure of risk on* $\mathscr{V}$.

Recall that $\sigma(\cdot)$ is a coherent measure of risk on $\mathscr{V}$ (we skip the other two arguments for brevity), if (see [2])

(A1) $\sigma(\alpha\varphi + (1-\alpha)\psi) \leq \alpha\sigma(\varphi) + (1-\alpha)\sigma(\psi) \; \forall \, \alpha \in (0,1), \; \varphi, \psi \in \mathscr{V}$;

(A2) If $\varphi \leq \psi$ then $\sigma(\varphi) \leq \sigma(\psi) \; \forall \, \varphi, \psi \in \mathscr{V}$;

(A3) $\sigma(a + \varphi) = a + \sigma(\varphi) \; \forall \, \varphi \in \mathscr{V}, \; a \in \mathbb{R}$;

(A4) $\sigma(\beta\varphi) = \beta\sigma(\varphi) \; \forall \, \varphi \in \mathscr{V}, \; \beta \geq 0$.

Example 3.1. Consider the first-order mean–semideviation risk measure analyzed by Ogryczak and Ruszczyński [30, 31], and Ruszczyński and Shapiro [43, Example 4.2], [44, Example 6.1]), but with the state and the underlying probability measure as its arguments. We define

$$(3.3) \qquad \sigma(\varphi, x, m) = \langle \varphi, m \rangle + \kappa \langle (\varphi - \langle \varphi, m \rangle)_+, m \rangle,$$

where $\kappa \in [0,1]$. We can verify directly that conditions (A1)–(A4) are satisfied. In a more general setting, $\kappa : \mathscr{X} \to [0,1]$ may be a measurable function.

Example 3.2. Another important example is the average value at risk (see, inter alia, Ogryczak and Ruszczyński [32, section 4], Pflug and Römisch [34, sections 2.2.3, 3.3.4], Rockafellar and Uryasev [39], Ruszczyński and Shapiro [43, Example 4.3], [44, Example 6.2]), which has the following transition risk counterpart:

$$\sigma(\varphi, x, m) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha} \langle (\varphi - \eta)_+, m \rangle \right\}, \quad \alpha \in (0,1).$$

Again, the conditions (A1)–(A4) can be verified directly. In a more general setting $\alpha : \mathscr{X} \to [\alpha_{\min}, \alpha_{\max}] \subset (0,1)$ may be a measurable function.

We shall use the property of *law invariance* of a transition risk mapping. For a function $\varphi \in \mathscr{V}$ and a probability measure $\mu \in \mathscr{M}$ we can define the distribution function $F_\varphi^\mu : \mathbb{R} \to [0,1]$ as follows:

$$F_\varphi^\mu(\eta) = \mu\big\{ y \in \mathscr{X} : \varphi(y) \leq \eta \big\}.$$

DEFINITION 3.2. *A transition risk mapping* $\sigma : \mathscr{V} \times \mathscr{X} \times \mathscr{M} \to \mathbb{R}$ *is* law invariant *if for all* $\varphi, \psi \in \mathscr{V}$ *and all* $\mu, \nu \in \mathscr{M}$ *such that* $F_\varphi^\mu \equiv F_\psi^\nu$, *we have* $\sigma(\varphi, x, \mu) = \sigma(\psi, x, \nu)$ *for all* $x \in \mathscr{X}$.

The concept of law invariance corresponds to a similar concept for coherent measures of risk, but here we additionally need to take into account the variability of the probability measure. The transition risk mappings of Examples 3.1 and 3.2 are law invariant.

The concept of law invariance is important in the context of Markov decision processes, where the model essentially defines the distribution of the state process for every policy Π. It also greatly simplifies the analysis of specific problems, as illustrated in section 10.1.

Transition risk mappings allow for convenient formulation of risk-averse preferences for controlled Markov processes, where the cost is evaluated by formula (3.1). Consider a controlled Markov process $\{x_t\}$ with a deterministic Markov policy $\Pi = \{\pi_1, \pi_2, \dots\}$. For a fixed time t and a measurable function $g : \mathscr{X} \times \mathscr{U} \times \mathscr{X} \to \mathbb{R}$ the value of $Z_{t+1} = g(x_t, u_t, x_{t+1})$ is a random variable. We assume that g is w-*bounded*, that is,

$$\big| g(x, u, y) \big| \leq C\big(w(x) + w(y) \big) \quad \forall \, x \in \mathscr{X}, \; u \in U(x), \; y \in \mathscr{X}$$

for some constant $C > 0$ and for the some *weight (bounding) function* $w : \mathscr{X} \to [1, \infty)$, $w \in \mathscr{V}$ (see, [5, section 2.4], [19, section 7.2], and [51] for the role of weight functions in

Markov decision processes). Then Z_{t+1} is an element of $\mathscr{Z}_{t+1}^{\Pi}$. Let $\rho_t^{\Pi} : \mathscr{Z}_{t+1}^{\Pi} \to \mathscr{Z}_t^{\Pi}$ be a family of conditional risk measures satisfying (A1)–(A4) for every deterministic policy Π. By definition, $\rho_t^{\Pi}\big(g(x_t, u_t, x_{t+1})\big)$ is an element of $\mathscr{Z}_t^{\Pi}$, that is, it is an $\mathscr{F}_t$-measurable function on $(\Omega, \mathscr{F})$. In the definition below, we restrict it to depend on the past only via the current state x_t.

DEFINITION 3.3. *A family of one-step conditional risk measures* $\rho_t^{\Pi} : \mathscr{Z}_{t+1}^{\Pi} \to \mathscr{Z}_t^{\Pi}$ *is a* Markov risk measure *with respect to the controlled Markov process* $\{x_t\}$ *if there exists a law invariant transition risk mapping* $\sigma : \mathscr{V} \times \mathscr{X} \times \mathscr{M} \to \mathbb{R}$ *such that for all w-bounded measurable functions* $g : \mathscr{X} \times \mathscr{U} \times \mathscr{X} \to \mathbb{R}$ *and for all feasible deterministic Markov policies* Π *we have*

$$(3.4) \qquad \rho_t^{\Pi}\big(g(x_t, \pi_t(x_t), x_{t+1})\big) = \sigma\big(g(x_t, \pi_t(x_t), \cdot), x_t, Q(x_t, \pi_t(x_t))\big) \quad a.s.$$

Observe that the right-hand side of formula (3.4) is parametrized by x_t, and thus it defines a special $\mathscr{F}_t$-measurable function of ω, whose dependence on the past is carried only via the state x_t. The quantifier "a.s." means "almost surely with respect to the measure P^{Π}."

4. Stochastic multikernels. In order to analyze Markov measures of risk, we need to introduce the concept of a multikernel.

DEFINITION 4.1. *A multikernel* *is a measurable multifunction* $\mathfrak{M}$ *from* $\mathscr{X}$ *to the space of regular measures on* $(\mathscr{X}, \mathscr{B}(X))$. *It is* stochastic *if its values are sets of probability measures. It is* substochastic *if* $0 \leq M(B|x) \leq 1$ *for all* $M \in \mathfrak{M}(x)$, $B \in \mathscr{B}(\mathscr{X})$, *and* $x \in \mathscr{X}$. *It is* convex (*closed*) *if for all* $x \in \mathscr{X}$ *its value* $\mathfrak{M}(x)$ *is a convex (closed) set.*

The concept of a multikernel is thus a multivalued generalization of the concept of a kernel. A measurable selector of a stochastic multikernel $\mathfrak{M}$ is a stochastic kernel M such that $M(x) \in \mathfrak{M}(x)$ for all $x \in \mathscr{X}$. We symbolically write $M \lessdot \mathfrak{M}$ to indicate that M is a measurable selector of $\mathfrak{M}$.

Recall that a composition $M_1 M_2$ of (sub)stochastic kernels M_1 and M_2 is given by the formula

$$(4.1) \qquad \big[M_1 M_2\big]\big(B\big|x\big) = \int_{\mathscr{X}} M_2(B|y)\, M_1(dy|x), \quad \mathscr{B} \in \mathscr{B}(\mathscr{X}), \quad x \in \mathscr{X}.$$

It is also a (sub)stochastic kernel. Multikernels, in particular substochastic multikernels, can be composed in a similar fashion.

DEFINITION 4.2. *If* $\mathfrak{M}_1$ *and* $\mathfrak{M}_2$ *are multikernels, then their* composition $\mathfrak{M}_1 \mathfrak{M}_2$ *is defined as follows:*

$$\big[\mathfrak{M}_1 \mathfrak{M}_2\big]\big(B\big|x\big) = \Big\{ \big[M_1 M_2\big]\big(B\big|x\big) : \ M_i \lessdot \mathfrak{M}_i, \ i = 1, 2 \Big\}.$$

It follows from Definition 4.2 that a composition of (sub)stochastic multikernels is a (sub)stochastic multikernel. We may compose a substochastic multikernel $\mathfrak{M}$ with itself several times, to obtain its "power":

$$(\mathfrak{M})^k = \underbrace{\mathfrak{M}\, \mathfrak{M}\, \cdots\, \mathfrak{M}}_{k \text{ times}}.$$

Multikernels can be *added* by employing the Minkowski sum of their values:

$$\big[\mathfrak{M}_1 + \mathfrak{M}_2\big](x) = \mathfrak{M}_1(x) + \mathfrak{M}_2(x) = \big\{\mu : \mu = \mu_1 + \mu_2,\ \mu_i \in \mathfrak{M}_i(x),\ i = 1, 2\big\}, \quad x \in \mathscr{X}.$$

The sum of stochastic multikernels is a multikernel with nonnegative values.

The concept of a multikernel and the composition operation arise in a natural way in the context of Markov risk measures. If $\sigma(\cdot,\cdot,\cdot)$ is a transition risk mapping, then the function $\sigma(\cdot,x,m)$ is lower semicontinuous for all $x \in \mathscr{X}$ and $m \in \mathscr{M}$ (see Ruszczyński and Shapiro [43, Proposition 3.1]). Then it follows from [43, Theorem 2.2] that for every $x \in \mathscr{X}$ and $m \in \mathscr{M}$ a closed convex set $\mathscr{A}(x,m) \subset \mathscr{M}$ exists, such that for all $\varphi \in \mathscr{V}$ we have

$$(4.2) \qquad \sigma(\varphi,x,m) = \max_{\mu \in \mathscr{A}(x,m)} \langle \varphi, \mu \rangle.$$

In fact, we also have

$$(4.3) \qquad \mathscr{A}(x,m) = \partial_\varphi \sigma(0,x,m),$$

that is, $\mathscr{A}(x,m)$ is the subdifferential of $\sigma(\cdot,x,m)$ at 0 (for the foundations of conjugate duality theory, see [38]). In many cases, the multifunction $\mathscr{A} : \mathscr{X} \times \mathscr{M} \rightrightarrows \mathscr{M}$ can be described analytically.

Example 4.1. For the mean-semideviation model of Example 3.1, following the derivations of [43, Example 4.2], we have
$$(4.4)$$
$$\mathscr{A}(x,m) = \left\{ \mu \in \mathscr{M} : \exists \big(h \in \mathscr{L}_\infty(\mathscr{X},\mathscr{B},P_0) \big) \frac{d\mu}{dm} = 1 + h - \langle h, m \rangle, \ \|h\|_\infty \leq \kappa, \ h \geq 0 \right\}.$$

Similar formulas can be derived for higher-order measures.

Example 4.2. For the conditional average value at risk of Example 3.2, following the derivations of [43, Example 4.3], we obtain

$$(4.5) \qquad \mathscr{A}(x,m) = \left\{ \mu \in \mathscr{M} : \frac{d\mu}{dm} \leq \frac{1}{\alpha} \right\}.$$

Consider formula (3.4) with $u_t = \pi_t(x_t)$. Using the representation (4.2) we can express the Markov risk measure as follows:

$$(4.6) \quad \rho_t^\Pi \big(g(x_t,\pi_t(x_t),x_{t+1}) \big) = \max_{\mu \in \mathscr{A}\big(x_t,Q(x_t,\pi_t(x_t))\big)} \int_\mathscr{X} g(x_t,\pi_t(x_t),y)\, \mu(dy) \quad \text{a.s.}$$

Suppose policy Π is stationary and $\pi_t = \pi$ for all t. For every $x \in \mathscr{X}$ we can define the set of probability measures

$$(4.7) \qquad \mathfrak{M}^\pi(x) = \mathscr{A}\big(x,Q(x,\pi(x))\big), \quad x \in \mathscr{X}.$$

The multifunction $\mathfrak{M}^\pi : \mathscr{X} \rightrightarrows \mathscr{P}(\mathscr{X})$, assigning to each $x \in \mathscr{X}$ the set $\mathfrak{M}^\pi(x)$, is a closed convex stochastic multikernel. We call it a *risk multikernel*, associated with the transition risk mapping $\sigma(\cdot,\cdot,\cdot)$, the controlled kernel Q, and the decision rule π. Its measurable selectors $M^\pi \prec \mathfrak{M}^\pi$ are transition kernels.

It follows that formula (4.6) for stationary policies Π can be rewritten as follows:

$$(4.8) \qquad \rho_t^\Pi \big(g(x_t,\pi_t(x_t),x_{t+1}) \big) = \max_{M \in \mathfrak{M}^\pi(x_t)} \int_\mathscr{X} g(x_t,\pi_t(x_t),y)\, M(dy).$$

In the risk-neutral case we have

$$\rho_t^\Pi \big(g(x_t,\pi_t(x_t),x_{t+1}) \big) = \mathbb{E}^\Pi \big[g(x_t,\pi_t(x_t),x_{t+1}) \big| x_t \big]$$
$$= \int_\mathscr{X} g(x_t,\pi_t(x_t)y)\, Q\big(dy\big| x_t,\pi(x_t)\big).$$

The comparison of the last two displayed equations reveals that in the risk-neutral case we have

$$\mathfrak{M}^\pi(x) = \big\{Q(x, \pi(x))\big\}, \quad x \in \mathscr{X}, \tag{4.9}$$

that is, the risk multikernel $\mathfrak{M}^\pi$ is single valued, and its only selector is the kernel $Q(\cdot, \pi(\cdot))$. In the risk-averse case, the risk multikernel $\mathfrak{M}^\pi$ is a closed convex-valued multifunction, whose measurable selectors are transition kernels. It is evident that properties of this multifunction are germane for our analysis. We return to this issue in section 7, where we calculate some examples of risk multikernels.

Remark 4.1. If $m \in \mathscr{A}(x, m)$ for all $x \in \mathscr{X}$ and $m \in \mathscr{M}$, then it follows from (4.7) that $Q(\cdot, \pi(\cdot))$ is a measurable selector of $\mathfrak{M}^\pi$. Moreover, it follows from (4.2) that for any function $\varphi \in \mathscr{V}$ we have

$$\rho_t^\Pi\big(\varphi(x_{t+1})\big) \geq \int_{\mathscr{X}} \varphi(y)\, Q\big(dy\big|x_t, \pi(x_t)\big) = \mathbb{E}^\Pi\big[\varphi(x_{t+1})\big|x_t\big].$$

It follows that the dynamic risk measure (3.1) is bounded from below by the expected value of the total cost. The condition $m \in \mathscr{A}(x, m)$ is satisfied by the measures of risk in Examples 4.1 and 4.2.

Interestingly, uncertain transition matrices were used by Nilim and El Ghaoui in [28] to increase robustness of control rules for Markov models. There is also an intriguing connection to Markov games (see, e.g., [17, 21]). In our theory, controlled multikernels arise in a natural way in the analysis of risk-averse preferences.

5. General assumptions. Semicontinuity and measurability. We call the controlled kernel Q *setwise (strongly) continuous* if for all Borel sets $B \subset \mathscr{X}$ and all convergent sequences $\{(x_k, u_k)\}$, $k = 1, 2, \ldots,$

$$\lim_{k \to \infty} Q(B|x_k, u_k) = Q(B|x, u),$$

where $x = \lim_{k \to \infty} x_k$ and $u = \lim_{k \to \infty} u_k$. We call Q *weakly* * continuous* if for all functions $v \in \mathscr{V}$

$$\lim_{k \to \infty} \int_{\mathscr{X}} v(y)\, Q(dy|x_k, u_k) = \int_{\mathscr{X}} v(y)\, Q(dy|x, u).$$

Under condition (G0), setwise and weak* continuity concepts are equivalent, because the set of piecewise constant functions is dense in $\mathscr{V}$.

In the product space $\mathscr{X} \times \mathscr{M}$ we always consider the product topology of strong convergence in $\mathscr{X}$ and weak* convergence in $\mathscr{M}$. In all our analyses we make the following assumptions:

(G1) The transition kernel $Q(\cdot, \cdot)$ is setwise continuous.
(G2) The multifunction $\mathscr{A}(\cdot, \cdot) \equiv \partial_\varphi \sigma(0, \cdot, \cdot)$ is lower semicontinuous.
(G3) The function $c(\cdot, \cdot, \cdot)$ is measurable, w-bounded, and $c(\cdot, \cdot, y)$ is lower semi-continuous for all $y \in \mathscr{X}$.
(G4) The multifunction $U(\cdot)$ is measurable and compact valued.

We need the following semicontinuity property of a transition risk mapping.

PROPOSITION 5.1. *Suppose* (G0)–(G3) *and let* $v \in \mathscr{V}$. *Then the mapping* $(x, u) \mapsto \sigma\big(c(x, u, \cdot) + v(\cdot), x, Q(x, u)\big)$ *is lower semicontinuous on* graph(U).

Proof. Let $\varphi(x, u, y) = c(x, u, y) + v(y)$. Consider the dual representation (4.2) of the transition risk mapping

$$\sigma(\varphi(x, u, \cdot), x, Q(x, u)) = \max_{\mu \in \mathscr{A}(x, Q(x, u))} \int_{\mathscr{X}} \varphi(x, u, y)\, \mu(dy). \tag{5.1}$$

By (G0), (G1), and (G2), the multifunction $(x, u) \mapsto \mathscr{A}(x, Q(x, u))$ is lower semicontinuous. Owing to condition (G3), the function $(x, u, \mu) \mapsto \int_{\mathscr{X}} \varphi(x, u, y) \, \mu(dy)$ is lower semicontinuous on $\mathrm{graph}(U) \times \mathscr{M}$. The assertion follows now from [4, Theorem 1.4.16], whose proof remains valid in our setting as well. $\square$

Some comments on the assumptions of Proposition 5.1 are in order. Continuity assumptions of the kernel Q are standard in the theory of risk-neutral Markov control processes (see, e.g., [18, App. C], [46]). If the transition risk mapping $\sigma(\cdot, \cdot, \cdot)$ is continuous, then its subdifferential (4.3) is upper semicontinuous. However, in Proposition 5.1 we assume *lower* semicontinuity of the mapping $(x, m) \mapsto \partial_\varphi \sigma(0, x, m)$, which is not trivial and should be verified for each case.

Example 5.1. Let us verify the lower semicontinuity assumption for the multifunction $\mathscr{A}$ given in (4.4). Consider an arbitrary $\mu \in \mathscr{A}(x, m)$ and suppose $x_k \to x$, $m_k \to m$, as $k \to \infty$. We need to find $\mu_k \in \mathscr{A}(x_k, m_k)$ such that $\mu_k \to \mu$. Let h be the function for which, according to (4.4), $\frac{d\mu}{dm} = 1 + h - \int h(z) \, m(dz)$. We define the μ_k by specifying their Radon–Nikodym derivatives: $\frac{d\mu_k}{dm_k} = 1 + h - \int h(z) \, m_k(dz)$. By construction, $\mu_k \in \mathscr{A}(x_k, m_k)$. Then, for any function $v \in \mathscr{V}$ we obtain

$$
\int_{\mathscr{X}} v(y) \, \mu_k(dy) = \int_{\mathscr{X}} v(y) \left(1 + h(y) - \int_{\mathscr{X}} h(z) \, m_k(dz) \right) m_k(dy)
$$
$$
= \int_{\mathscr{X}} v(y) \big(1 + h(y) \big) \, m_k(dy) - \int_{\mathscr{X}} h(z) \, m_k(dz) \int_{\mathscr{X}} v(y) \, m_k(dy).
$$

As $m_k \to m$, we conclude that for all $v \in \mathscr{V}$

$$
\lim_{k \to \infty} \int_{\mathscr{X}} v(y) \, \mu_k(dy)
$$
$$
= \int_{\mathscr{X}} v(y) \big(1 + h(y) \big) \, m(dy) - \int_{\mathscr{X}} h(z) \, m(dz) \int_{\mathscr{X}} v(y) \, m(dy) = \int_{\mathscr{X}} v(y) \, \mu(dy),
$$

which is the weak* convergence of μ_k to μ.

In the following result we use the concept of a *normal integrand*, that is, a function $f : \mathscr{X} \times U \to \mathbb{R} \cup +\infty$ such that that its epigraphical mapping

$$
x \mapsto \{ (u, \alpha) \in \mathscr{U} \times \mathbb{R} : f(x, u) \leq \alpha \}
$$

is a closed-valued and measurable multifunction (see Rockafellar and Wets [40, section 14.D]).

PROPOSITION 5.2. *Suppose* (G0)–(G4) *and let* $v \in \mathscr{V}$. *Then the function*

$$
\psi(x) = \inf_{u \in U(x)} \sigma \big(c(x, u, \cdot) + v(\cdot), x, Q(x, u) \big), \quad x \in \mathscr{X},
$$

is measurable and w-bounded, and a measurable selector $\pi \lessdot U$ exists, such that

$$
\psi(x) = \sigma \big(c(x, \pi(x), \cdot) + v(\cdot), x, Q(x, \pi(x)) \big) \quad \forall\, x \in \mathscr{X}.
$$

Proof. Owing to Proposition 5.1, the function $(x, u) \mapsto \sigma(c(x, u, \cdot) + v(\cdot), x, Q(x, u))$ is lower semicontinuous, and is thus a normal integrand [40, Ex. 14.31]. Consider the function $f : \mathscr{X} \times U \to \mathbb{R} \cup +\infty$ defined as follows:

$$
f(x, u) = \begin{cases} \sigma \big(c(x, u, \cdot) + v(\cdot), x, Q(x, u) \big) & \text{if } u \in U(x), \\ +\infty & \text{otherwise.} \end{cases}
$$

Due to (G4), it is a normal integrand as well. It follows from [40, Thm. 14.37] that the function $\psi(x) = \inf_u f(x, u)$ is measurable and that the optimal solution mapping $\Psi(x) = \{u \in \mathscr{U} : \psi(x) = f(x, u)\}$ is measurable. By (G4), the set $U(x)$ is compact, and thus $\Psi(x) \neq \emptyset$ for all $x \in \mathscr{X}$. Ψ is also compact valued. By virtue of [24], a measurable selector $\pi \lessdot \Psi$ exists. Let us recall the dual representation (5.1) again:

$$\psi(x) = \max_{\mu \in \mathscr{A}(x, Q(x, \pi(x)))} \int_{\mathscr{X}} \varphi(x, \pi(x), y) \, \mu(dy)$$

with $\varphi(x, u, y) = c(x, u, y) + v(y)$. As the set $\mathscr{A}(x, Q(x, \pi(x)))$ contains only probability measures, and the function $\varphi(\cdot, \cdot, \cdot)$ is w-bounded, the function $\psi(\cdot)$ is w-bounded as well. $\quad\square$

6. Finite horizon problem. We consider the Markov model at times $1, 2, \ldots, T+1$ under deterministic policies $\Pi = \{\pi_1, \pi_2, \ldots, \pi_T\}$. The cost at the last stage is given by a function $v_{T+1}(x_{T+1})$. Consider the problem

$$\tag{6.1} \min_{\Pi} \ J_T(\Pi, x_1),$$

where $J_T(\Pi, x_1)$ is defined by formula (3.1), with Markov conditional risk measures ρ_t^{Π}, $t = 1, \ldots, T$:

$$\tag{6.2} J_T(\Pi, x_1) = \rho_1^{\Pi}\bigg(c(x_1, u_1, x_2) + \rho_2^{\Pi}\Big(c(x_2, u_2, x_3) + \cdots$$
$$+ \rho_T^{\Pi}\big(c(x_T, u_T, x_{T+1}) + v_{T+1}(x_{T+1}) \big) \cdots \Big) \bigg).$$

In the formula above, we have $u_t = \pi_t(x_t)$, $t = 1, \ldots, T$. We assume that every one-step measure has the form (3.4), with some transition risk mapping $\sigma(\cdot, \cdot, \cdot)$.

THEOREM 6.1. *Assume that the general conditions* (G0)–(G4) *are satisfied, and that the function $v_{T+1}(\cdot)$ is measurable and w-bounded. Then problem* (6.1) *has an optimal solution and its optimal value $v_1(x)$ is the solution of the following dynamic programming equations:*

$$\tag{6.3} v_t(x) = \min_{u \in U(x)} \ \sigma\big(c(x, u, \cdot) + v_{t+1}(\cdot), x, Q(x, u) \big), \quad x \in \mathscr{X}, \quad t = T, \ldots, 1.$$

Moreover, an optimal Markov policy $\hat{\Pi} = \{\hat{\pi}_1, \ldots, \hat{\pi}_T\}$ exists and satisfies the equations

$$\tag{6.4} \hat{\pi}_t(x) \in \operatorname*{argmin}_{u \in U(x)} \ \sigma\big(c(x, u, \cdot) + v_{t+1}(\cdot), x, Q(x, u) \big), \quad x \in \mathscr{X}, \quad t = T, \ldots, 1.$$

Conversely, every solution of (6.3)–(6.4) *defines an optimal Markov policy $\hat{\Pi}$.*

Proof. Our proof is based on the ideas of the proof of Ruszczyński [41, Thm. 2], but with refinements rectifying some technical inaccuracies.[2]

Consider two policies $\Pi = \{\pi_1, \ldots, \pi_{T-1}, \pi_T\}$ and $\Pi' = \{\pi_1, \ldots, \pi_{T-1}, \pi_T'\}$, differing in the last decision rule. The corresponding state and control sequences,

$$\{x_1, u_1, \ldots, x_T, u_T, x_{T+1}\} \quad \text{and} \quad \{x_1, u_1, \ldots, x_T, u_T', x_{T+1}'\},$$

---

[2]In [41, Thm. 2] we missed the measurability condition on $U(\cdot)$ and the assumptions of *joint* continuity (lower semicontinuity) of the kernel and the cost functions.

differ in the last control and the final state. Since the risk measures are Markov and share the same transition risk mapping σ, we have

$$\rho_T^\Pi\big(c(x_T, u_T, x_{T+1}) + v_{T+1}(x_{T+1})\big)$$
$$= \sigma\big(c(x_T, \pi_T(x_T), \cdot) + v_{T+1}(\cdot), x_T, Q(x_T, \pi_T(x_T))\big),$$
$$\rho_T^{\Pi'}\big(c(x_T, u_T', x_{T+1}') + v_{T+1}(x_{T+1}')\big)$$
$$= \sigma\big(c(x_T, \pi_T'(x_T), \cdot) + v_{T+1}(\cdot), x_T, Q(x_T, \pi_T'(x_T))\big).$$

Moreover, the distribution of x_T, which depends only on $\pi_1, \ldots, \pi_{T-1}$, is the same in both cases. If

$$\sigma\big(c(x_T, \pi_T(x_T), \cdot) + v_{T+1}(\cdot), x_T, Q(x_T, \pi_T(x_T))\big)$$
$$\leq \sigma\big(c(x_T, \pi_T'(x_T), \cdot) + v_{T+1}(\cdot), x_T, Q(x_T, \pi_T'(x_T))\big) \quad \text{a.s.,}$$

then using the monotonicity condition (A2) for $t = T-1, \ldots, 1$ we obtain

$$\rho_1^\Pi\Big(c(x_1, u_1, x_2) + \cdots + \rho_T^\Pi\big(c(x_T, u_T, x_{T+1}) + v_{T+1}(x_{T+1})\big) \cdots\Big)$$
$$\leq \rho_1^{\Pi'}\Big(c(x_1, u_1, x_2) + \cdots + \rho_T^{\Pi'}\big(c(x_T, u_T', x_{T+1}') + v_{T+1}(x_{T+1}')\big) \cdots\Big).$$

Therefore, we can move the optimization with respect to π_T inside, and rewrite problem (6.1) as follows:

$$\inf_{\pi_1, \ldots, \pi_T} \left\{ \rho_1^\Pi\Big(c(x_1, \pi_1(x_1), x_2) + \cdots + \rho_T^\Pi\big(c(x_T, \pi_T(x_T), x_{T+1}) + v_{T+1}(x_{T+1})\big) \cdots\Big)\right\}$$
$$= \inf_{\pi_1, \ldots, \pi_{T-1}} \left\{ \rho_1^\Pi\Big(c(x_1, \pi_1(x_1), x_2) + \cdots \right.$$
$$\left. + \inf_{\pi_T} \rho_T^\Pi\big(c(x_T, \pi_T(x_T), x_{T+1}) + v_{T+1}(x_{T+1})\big) \cdots\Big)\right\}.$$

Owing to the Markov structure of the conditional risk measure ρ_T, the innermost optimization problem can be rewritten as follows:

$$(6.5) \qquad \inf_{\pi_T} \sigma\big(c(x_T, \pi_T(x_T), \cdot) + v_{T+1}(\cdot), x_T, Q(x_T, \pi_T(x_T))\big)$$
$$= \inf_{u \in U(x_T)} \sigma\big(c(x_T, u, \cdot) + v_{T+1}(\cdot), x_T, Q(x_T, u)\big).$$

The problem becomes equivalent to (6.3) for $t = T$, and its solution is given by (6.4) for $t = T$. By Proposition 5.2, a measurable selector $\hat{\pi}_T(\cdot)$ exists, such that $\hat{\pi}_T(x_T)$ is the minimizer in (6.5) for any x_T. Finally, the optimal value in (6.5), which we denote by $v_T(x_T)$, is measurable and w-bounded. It follows from the above considerations that for every fixed x,

$$v_T(x) = \min_{\pi_T} \sigma\big(c(x, \pi_T(x), \cdot) + v_{T+1}(\cdot), x, Q(x, \pi_T(x))\big)$$
$$= \sigma\big(c(x, \hat{\pi}_T(x), \cdot) + v_{T+1}(\cdot), x, Q(x, \hat{\pi}_T(x))\big)$$

is the optimal value of the problem starting at time T from $x_T = x$. It is achieved by the control $\hat{\pi}_T(x)$.

After that, the horizon $T+1$ is decreased to T, and the final cost becomes $v_T(x_T)$. Proceeding in this way for $T, T-1, \ldots, 1$ we obtain the assertion of the theorem. $\square$

It follows from our proof that the functions $v_t(\cdot)$ calculated in (6.3) are the optimal values of tail subproblems formulated for a fixed $x_t = x$ as follows:

$$v_t(x) = \min_{\pi_t,\ldots,\pi_T} \rho_t^\Pi\Big(c(x_t, \pi_t(x_t), x_{t+1}) + \rho_{t+1}^\Pi\big(c(x_{t+1}, \pi_{t+1}(x_{t+1}), x_{t+2}) + \cdots$$
$$+ \rho_T^\Pi\big(c(x_T, \pi_T(x_T), x_{T+1}) + v_{T+1}(x_{T+1})\big)\cdots\big)\Big).$$

We call them *value functions*, as in risk-neutral dynamic programming. It is obvious that we may have nonstationary costs, transition kernels, and transition risk mappings in this case. Also, the assumption that the process is transient is not needed.

Equations (6.3)–(6.4) provide a computational recipe for solving finite horizon problems.

7. Evaluation of stationary Markov policies in infinite horizon problems. Consider a stationary policy $\Pi = \{\pi, \pi, \ldots\}$ and define the cost until absorption as follows:

$$J_\infty(\Pi, x_1) = \lim_{T \to \infty} J_T(\Pi, x_1), \tag{7.1}$$

where each $J_T(\Pi, x_1)$ is defined by the formula

$$
\begin{aligned}
J_T(\Pi, x_1) &= \rho_{1,T+1}^\Pi\big(0, c(x_1, \pi(x_1), x_2), c(x_2, \pi(x_2), x_3), \ldots, c(x_T, \pi(x_T), x_{T+1})\big) \\
&= \rho_1^\Pi\Big(c(x_1, \pi(x_1), x_2) + \rho_2^\Pi\Big(c(x_2, \pi(x_2), x_3) + \cdots \\
&\quad + \rho_T^\Pi\big(c(x_T, \pi(x_T), x_{T+1})\big)\cdots\Big)\Big)
\end{aligned}
\tag{7.2}
$$

with Markov conditional risk measures ρ_t^Π, $t = 1, \ldots, T$, sharing the same transition risk mapping $\sigma(\cdot, \cdot, \cdot)$. We assume all conditions of Theorem 6.1.

The first question to answer is when this cost is finite. This question is nontrivial, because even for uniformly bounded costs $Z_t = c(x_{t-1}, \pi(x_{t-1}), x_t)$, $t = 2, 3, \ldots$, and for a transient finite-state Markov chain, the limit in (7.1) may be infinite, as the following example demonstrates.

Example 7.1. Consider a transient Markov chain with two states and with the following transition probabilities: $Q_{11} = Q_{12} = \frac{1}{2}$, $Q_{22} = 1$. Only one control is possible in each state, the cost of each transition from state 1 is equal to 1, and the cost of the transition from 2 to 2 is 0. Clearly, the time until absorption is a geometric random variable with parameter $\frac{1}{2}$. Let $x_1 = 1$. If the limit (7.1) is finite, then (skipping the dependence on Π) we have

$$J_\infty(1) = \lim_{T \to \infty} J_T(1) = \lim_{T \to \infty} \rho_1\big(1 + J_{T-1}(x_2)\big) = \rho_1\big(1 + J_\infty(x_2)\big).$$

In the last equation we used the continuity of $\rho_1(\cdot)$. Clearly, $J_\infty(2) = 0$.

Suppose that we are using the average value at risk from Example 3.2 with $0 < \alpha \le \frac{1}{2}$ to define $\rho_1(\cdot)$. Using standard identities for the average value at risk (see, e.g., [47, Thm. 6.2]), we obtain

$$
\begin{aligned}
J_\infty(1) &= \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha}\mathbb{E}\big[\big(1 + J_\infty(x_2) - \eta\big)_+\big] \right\} \\
&= 1 + \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha}\mathbb{E}\big[\big(J_\infty(x_2) - \eta\big)_+\big] \right\} = 1 + \frac{1}{\alpha}\int_{1-\alpha}^1 F^{-1}(\beta)\, d\beta,
\end{aligned}
\tag{7.3}
$$

where $F(\cdot)$ is the distribution function of $J_\infty(x_2)$. As all β-quantiles of $J_\infty(x_2)$ for $\beta \geq \frac{1}{2}$ are equal to $J_\infty(1)$, the last equation yields $J_\infty(1) = 1 + J_\infty(1)$, a contradiction. It follows that a composition of average values at risk has no finite limit, if $0 < \alpha \leq \frac{1}{2}$.

On the other hand, if $\frac{1}{2} < \alpha < 1$, then

$$F^{-1}(\beta) = \begin{cases} J_\infty(2) = 0 & \text{if } 1 - \alpha \leq \beta < \frac{1}{2}, \\ J_\infty(1) & \text{if } \frac{1}{2} \leq \beta \leq 1. \end{cases}$$

Formula (7.3) then yields $J_\infty(1) = 1 + \frac{1}{2\alpha} J_\infty(1)$. This equation has a solution $J_\infty(1) = 2\alpha/(2\alpha - 1)$.

If we use the mean-semideviation model of Example 3.1, we obtain

$$J_\infty(1) = \mathbb{E}\big[1 + J_\infty(x_2)\big] + \kappa \mathbb{E}\Big[\Big(1 + J_\infty(x_2) - \mathbb{E}\big[1 + J_\infty(x_2)\big]\Big)_+\Big]$$

$$= 1 + \frac{1}{2} J_\infty(1) + \kappa \frac{1}{2}\Big(J_\infty(1) - \frac{1}{2} J_\infty(1)\Big) = 1 + \frac{2 + \kappa}{4} J_\infty(1).$$

Thus $J_\infty(1) = 4/(2 - \kappa)$, which is finite for all $\kappa \in [0, 1]$, that is, for all values of κ for which the model defines a coherent measure of risk.

It follows that deeper properties of the measures of risk and their interplay with the transition kernel need to be investigated to answer the question about finiteness of the dynamic measure of risk in this case.

Recall that with every transition risk mapping $\sigma(\cdot, \cdot, \cdot)$, every controlled kernel Q, and every decision rule π, a multikernel $\mathfrak{M}^\pi$ is associated, as defined in (4.7). Similar to the expected value case, it is convenient to consider the effective state space $\widetilde{\mathscr{X}} = \mathscr{X} \setminus \{x_A\}$, and the *effective substochastic multikernel* $\widetilde{\mathfrak{M}}^\pi$ whose arguments are restricted to $\widetilde{\mathscr{X}}$ and whose values are convex sets of nonnegative measures on $\widetilde{\mathscr{X}}$ defined by the identity: $\widetilde{\mathfrak{M}}^\pi(B|x) \equiv \mathfrak{M}^\pi(B|x)$ for all $B \in \mathscr{B}(\widetilde{\mathscr{X}})$ and $x \in \widetilde{\mathscr{X}}$.

A function $v \in \mathscr{V}$ with $v(x_A) = 0$ can be identified with a function $\tilde{v}$ on $\widetilde{\mathscr{X}}$; we shall write $\|\tilde{v}\|$ for the norm $\|v\|$ in $\mathscr{V}$; we shall also write $\tilde{v} \in \mathscr{V}$ to indicate that the corresponding extension v is an element of $\mathscr{V}$. Recall that the norm $\|\cdot\|_w$ associated with a weight function w is defined as follows:

$$\|v\|_w = \sup_{x \in \widetilde{\mathscr{X}}} \frac{|v(x)|}{w(x)}.$$

The corresponding operator norm $\|A\|_w$ of a substochastic kernel A is defined as follows:

$$\|A\|_w = \sup_{x \in \widetilde{\mathscr{X}}} \frac{1}{w(x)} \int_{\widetilde{\mathscr{X}}} w(y) \, A(dy|x).$$

Hernández-Lerma and Lasserre [19] extensively discuss the role of weighted norms in dynamic programming models.

DEFINITION 7.1. *We call the Markov model with a transition risk mapping $\sigma(\cdot, \cdot, \cdot)$ and with a stationary Markov policy $\{\pi, \pi, \dots\}$ risk transient if a weight function $w : \widetilde{\mathscr{X}} \to [1, \infty)$, $w \in \mathscr{V}$, and a constant K exist such that*

$$(7.4) \qquad \big\|M\big\|_w \leq K \quad \text{for all} \quad M \prec \sum_{j=1}^{T} \big(\widetilde{\mathfrak{M}}^\pi\big)^j \quad \text{and all} \quad T \geq 0.$$

If the estimate (7.4) *is uniform for all Markov policies, the model is called* uniformly risk transient.

In the special case of a risk-neutral model, owing to (4.9), Definition 7.1 reduces to the condition that

$$\tag{7.5} \left\| \sum_{j=1}^{\infty} \left(\widetilde{Q}^{\pi} \right)^j \right\|_w \leq K,$$

which has been analyzed by Pliska [35] and [19, section 9.6].

Example 7.2. Consider the simple transient chain of Example 7.1 with the average value at risk from Examples 3.2 and 4.2, where $0 < \alpha \leq 1$. From (4.5) we obtain

$$\mathscr{A}(i, m) = \left\{ (\mu_1, \mu_2) : 0 \leq \mu_j \leq \frac{m_j}{\alpha}, \ j = 1, 2; \ \mu_1 + \mu_2 = 1 \right\}.$$

As only one control is possible, formula (4.7) simplifies to

$$\mathfrak{M}(i) = \left\{ (\mu_1, \mu_2) : 0 \leq \mu_j \leq \frac{Q_{ij}}{\alpha}, \ j = 1, 2; \ \mu_1 + \mu_2 = 1 \right\}, \quad i = 1, 2.$$

The effective state space is just $\widetilde{\mathscr{X}} = \{1\}$, and we conclude that the effective multi-kernel is the interval

$$\widetilde{\mathfrak{M}} = \left[0, \min\left(1, \frac{1}{2\alpha} \right) \right].$$

For $0 < \alpha \leq \frac{1}{2}$ we can select $\widetilde{M} = 1 \in \widetilde{\mathfrak{M}}$ to show that $1 \in \left(\widetilde{\mathfrak{M}} \right)^j$ for all j, and thus condition (7.4) is not satisfied. On the other hand, if $\frac{1}{2} < \alpha \leq 1$, then for every $\widetilde{M} \in \widetilde{\mathfrak{M}}$ we have $0 \leq \widetilde{M} < 1$, and condition (7.4) is satisfied.

Consider now the mean-semideviation model of Examples 3.1 and 4.1. From (4.4) we obtain

$$\mathscr{A}(i, m) = \left\{ (\mu_1, \mu_2) : \mu_j = m_j \left(1 + h_j - (h_1 m_1 + h_2 m_2) \right), \ 0 \leq h_j \leq \kappa, \ j = 1, 2 \right\},$$

$$\mathfrak{M}(i) = \left\{ (\mu_1, \mu_2) : \mu_j = Q_{ij} \left(1 + h_j - (h_1 Q_{i1} + h_2 Q_{i2}) \right), \ 0 \leq h_j \leq \kappa, \ j = 1, 2 \right\},$$

$$i = 1, 2.$$

Calculating the lowest and the largest possible values of μ_1 we conclude that

$$\widetilde{\mathfrak{M}} = \left[\frac{1}{2} \left(1 - \frac{\kappa}{2} \right), \frac{1}{2} \left(1 + \frac{\kappa}{2} \right) \right].$$

For every $\kappa \in [0, 1]$, Definition 7.1 is satisfied.

We can now provide sufficient conditions for the finiteness of the limit (7.1).

THEOREM 7.2. *Suppose a stationary policy* $\Pi = \{\pi, \pi, \dots\}$ *is applied to the controlled Markov model with a transition risk mapping* $\sigma(\cdot, \cdot, \cdot)$. *If the model satisfies conditions* (G0)–(G3) *and is risk transient for the policy* Π, *then the limit*

$$\tag{7.6} J_{\infty}(\Pi, \cdot) = \lim_{T \to \infty} J_T(\Pi, \cdot)$$

exists in $\mathscr{V}$ *and is w-bounded. If the model is additionally uniformly risk transient, then* $\|J_{\infty}(\Pi, \cdot)\|_w$ *is uniformly bounded for all* Π *and the limit function* $(\pi, x) \mapsto J_{\infty}(\Pi, x)$ *is lower semicontinuous.*

Proof. By Conditions (A1)–(A4), each conditional risk measure $\rho_{1,T}(\cdot)$ is convex and positively homogeneous, and thus subadditive. For any $1 < T_1 < T_2$ we obtain the following estimate of (7.2):

(7.7)
$$
\begin{aligned}
J_{T_2-1}(\Pi, x_1) &= \rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_2}) \\
&\leq \rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_1}, 0, \ldots, 0) + \rho_{1,T_2}^{\Pi}(0, \ldots, 0, Z_{T_1+1}, \ldots, Z_{T_2}) \\
&= \rho_{1,T_1}^{\Pi}(0, Z_2, \ldots, Z_{T_1}) + \rho_{1,T_2}^{\Pi}(0, \ldots, 0, Z_{T_1+1}, \ldots, Z_{T_2}) \\
&= J_{T_1-1}(\Pi, x_1) + \rho_{1,T_2}^{\Pi}(0, \ldots, 0, Z_{T_1+1}, \ldots, Z_{T_2}).
\end{aligned}
$$

As the cost function is w-bounded, $Z_{j+1} \leq C(\bar{w}(x_j) + \bar{w}(x_{j+1}))$, where $\bar{w}(x) = w(x)$ if $x \in \widetilde{\mathscr{X}}$, and $\bar{w}(x_{\mathrm{A}}) = 0$. Owing to the monotonicity and positive homogeneity of the conditional risk mappings,

(7.8)
$$
\begin{aligned}
\rho_{1,T_2}^{\Pi}(0, \ldots, 0, Z_{T_1+1}, \ldots, Z_{T_2}) &\leq 2C\rho_{1,T_2}^{\Pi}(0, \ldots, 0, \bar{w}(x_{T_1+1}), \ldots, \bar{w}(x_{T_2})) \\
&= 2C\rho_1^{\Pi}\Big(\rho_2^{\Pi}\Big(\cdots \rho_{T_1}^{\Pi}\Big(\bar{w}(x_{T_1+1}) + \rho_{T_1+1}^{\Pi}\big(\bar{w}(x_{T_1+2}) + \cdots \\
&\qquad\qquad + \rho_{T_2-1}^{\Pi}\big(\bar{w}(x_{T_2})\big)\cdots\big)\big)\cdots\Big)\Big).
\end{aligned}
$$

If $x_{T_2-1} \neq x_{\mathrm{A}}$, applying (4.8) to the innermost expression, we obtain

$$
\rho_{T_2-1}^{\Pi}\big(\bar{w}(x_{T_2})\big) = \max_{m \in \widetilde{\mathfrak{M}}^{\pi}(x_{T_2-1})} \int_{\widetilde{\mathscr{X}}} w(y)\, m(dy).
$$

It is a function of x_{T_2-1}, which we denote as $v_{T_2-1}(x_{T_2-1})$. Consider the function

(7.9)
$$
v_{T_2-1}(x) = \max_{m \in \widetilde{\mathfrak{M}}^{\pi}(x)} \int_{\widetilde{\mathscr{X}}} w(y)\, m(dy), \quad x \in \widetilde{X}.
$$

Owing to the weak* compactness of the values of the multikernel $\widetilde{\mathfrak{M}}^{\pi}$, the maximizers in (7.9) exist and can be chosen to depend in a measurable way on x. Thus, they form a measurable selector $\widetilde{M}_{T_2-1}$ of $\widetilde{\mathfrak{M}}^{\pi}$. Therefore,

(7.10)
$$
v_{T_2-1} = \widetilde{M}_{T_2-1} w, \quad \widetilde{M}_{T_2-1} \lessdot \widetilde{\mathfrak{M}}^{\pi}.
$$

One step earlier, we obtain
(7.11)
$$
\begin{aligned}
\rho_{T_2-2}^{\Pi}\Big(\bar{w}(x_{T_2-1}) + \rho_{T_2-1}^{\Pi}\big(\bar{w}(x_{T_2})\big)\Big) &= \rho_{T_2-2}^{\Pi}\big(\bar{w}(x_{T_2-1}) + v_{T_2-1}(x_{T_2-1})\big) \\
&= \max_{m \in \widetilde{\mathfrak{M}}^{\pi}(x_{T_2-2})} \int_{\mathscr{X}} \big[w(y) + v_{T_2-1}(y)\big]\, m(dy).
\end{aligned}
$$

Again, the maximizers $\widetilde{M}_{T_2-2}(x_{T_2-2})$ in (7.11) exist, and they can be chosen in a measurable way. Denoting the optimal value by $v_{T_2-2}(x_{T_2-2})$, we obtain a relation similar to (7.10):

(7.12)
$$
\begin{aligned}
v_{T_2-2} &= \widetilde{M}_{T_2-2}\big(w + v_{T_2-1}\big) \\
&= \Big(\widetilde{M}_{T_2-2} + \widetilde{M}_{T_2-2}\widetilde{M}_{T_2-1}\Big) w, \quad \widetilde{M}_{T_2-2} \lessdot \widetilde{\mathfrak{M}}^{\pi}, \quad \widetilde{M}_{T_2-1} \lessdot \widetilde{\mathfrak{M}}^{\pi}.
\end{aligned}
$$

Proceeding in this way, we can calculate the function

$$v_{T_1}(x_{T_1}) = \rho_{T_1}^{\Pi}\Big(\bar{w}(x_{T_1+1}) + \rho_{T_1+1}^{\Pi}\big(\bar{w}(x_{T_1+2}) + \rho_{T_1+2}^{\Pi}\big(\bar{w}(x_{T_1+3}) + \cdots$$
$$+ \rho_{T_2-1}^{\Pi}\big(\bar{w}(x_{T_2})\big)\cdots\big)\big)\Big)$$

on $\widetilde{\mathscr{X}}$ as follows:

$$v_{T_1} = \Big(\widetilde{M}_{T_1} + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} + \cdots + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1}\cdots\widetilde{M}_{T_2-1}\Big)\,w$$

with $\widetilde{M}_j \prec \widetilde{\mathfrak{M}}^{\pi}$, $j = T_1, \ldots, T_2 - 1$. In the formula above, we restrict the domains of the functions to $\widetilde{X}$; at x_A their values are zero. Finally, defining

$$v_1(x_1) = \rho_1^{\Pi}\Big(\rho_2^{\Pi}\Big(\cdots\rho_{T_1}^{\Pi}\Big(\bar{w}(x_{T_1+1}) + \rho_{T_1+1}^{\Pi}\big(\bar{w}(x_{T_1+1}) + \cdots$$
$$+ \rho_{T_2-1}^{\Pi}\big(\bar{w}(x_{T_2})\big)\cdots\big)\Big)\cdots\Big)\Big),$$

we obtain the representation

$$(7.13)\quad v_1 = \widetilde{M}_1\widetilde{M}_2\cdots\widetilde{M}_{T_1-1}\Big(\widetilde{M}_{T_1} + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} + \cdots + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1}\ldots\widetilde{M}_{T_2-1}\Big)\,w$$

with $\widetilde{M}_j \prec \widetilde{\mathfrak{M}}^{\pi}$, $j = 1, \ldots, T_2 - 1$. This combined with (7.7)–(7.8) yields an estimate

$$(7.14)\quad J_{T_2-1}(\Pi, \cdot) - J_{T_1-1}(\Pi, \cdot)$$
$$\leq 2C\widetilde{M}_1\widetilde{M}_2\cdots\widetilde{M}_{T_1-1}\Big(\widetilde{M}_{T_1} + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} + \cdots + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1}\cdots\widetilde{M}_{T_2-1}\Big)\,w.$$

Consider now the sequence of costs $Z_1, \ldots, Z_{T_1}, -Z_{T_1+1}, \ldots, -Z_{T_2}$, in which we flip the sign of the costs $Z_{t+1} = c(x_t, u_t, x_{t+1})$ for $t \geq T_1$. From subadditivity, in a similar way (7.7), we obtain

$$(7.15)\quad \rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_1}, -Z_{T_1+1}, \ldots, -Z_{T_2})$$
$$\leq \rho_{1,T_1}^{\Pi}(0, Z_2, \ldots, Z_{T_1}) + \rho_{1,T_2}^{\Pi}(0, \ldots, 0, -Z_{T_1+1}, \ldots, -Z_{T_2}).$$

By convexity of $\rho_{1,T_2}(\cdot)$,

$$2\rho_{1,T_1}^{\Pi}(0, Z_2, \ldots, Z_{T_1}) \leq \rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_1}, Z_{T_1+1}, \ldots, Z_{T_2})$$
$$+ \rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_1}, -Z_{T_1+1}, \ldots, -Z_{T_2}).$$

Substituting the estimate (7.15), we deduce that

$$\rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_2}) \geq \rho_{1,T_1}^{\Pi}(0, Z_2, \ldots, Z_{T_1}) - \rho_{1,T_2}^{\Pi}(0, \ldots, 0, -Z_{T_1+1}, \ldots, -Z_{T_2}).$$

As the $|Z_{t+1}|$ are bounded by $C(\bar{w}(x_t) + \bar{w}(x_{t+1}))$, the estimate (7.8) applies to the last element on the right-hand side. We obtain

$$J_{T_2-1}(\Pi, x_1) - J_{T_1-1}(\Pi, x_1)$$
$$= \rho_{1,T_2}^{\Pi}(0, Z_2, \ldots, Z_{T_2}) - \rho_{1,T_1}^{\Pi}(0, Z_2, \ldots, Z_{T_1})$$
$$\geq -2C\rho_1^{\Pi}\Big(\rho_2^{\Pi}\Big(\cdots\rho_{T_1}^{\Pi}\Big(\bar{w}(x_{T_1+1}) + \rho_{T_1+1}^{\Pi}\big(\bar{w}(x_{T_1+1}) + \cdots$$
$$+ \rho_{T_2-1}^{\Pi}\big(\bar{w}(x_{T_2})\big)\cdots\big)\Big)\cdots\Big)\Big)$$
$$= -2Cv_1(x_1),$$

where $v_1(\cdot)$ has representation (7.13). This combined with (7.14) yields

$$\left| J_{T_2-1}(\Pi, x_1) - J_{T_1-1}(\Pi, x_1) \right| \leq 2C|v_1(x_1)|, \quad x_1 \in \widetilde{\mathscr{X}}.$$

This pointwise estimate implies the relations between the norms

$$\left\| J_{T_2-1}(\Pi, \cdot) - J_{T_1-1}(\Pi, \cdot) \right\|_w \leq 2C\left\| v_1 \right\|_w.$$

In view of representation (7.13), we obtain the estimate

$$\left\| J_{T_2-1}(\Pi, \cdot) - J_{T_1-1}(\Pi, \cdot) \right\|_w$$
$$\leq 2C \left\| \widetilde{M}_1 \widetilde{M}_2 \cdots \widetilde{M}_{T_1-1} \left(\widetilde{M}_{T_1} + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} + \cdots + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} \cdots \widetilde{M}_{T_2-1} \right) w \right\|_w.$$

By Definition 7.1, $\|\widetilde{M}_{T_1} + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} + \cdots + \widetilde{M}_{T_1}\widetilde{M}_{T_1+1} \cdots \widetilde{M}_{T_2-1}\|_w \leq K$. Since $\|w\|_w = 1$, we infer that

$$(7.16) \qquad \left\| J_{T_2-1}(\Pi, \cdot) - J_{T_1-1}(\Pi, \cdot) \right\|_w \leq 2CK \left\| \widetilde{M}_1 \widetilde{M}_2 \cdots \widetilde{M}_{T_1-1} \right\|_w.$$

Observe that $\widetilde{M}_1 \widetilde{M}_2 \cdots \widetilde{M}_{T_1-1} \prec (\widetilde{\mathfrak{M}}^\pi)^{T_1-1}$. It follows from Definition 7.1 that for any sequence of selectors $A_j \prec (\widetilde{\mathfrak{M}}^\pi)^j$ we have $\| \sum_{j=1}^\infty A_j \|_w \leq K$. Therefore, $\|A_j\|_w \to 0$ as $j \to \infty$. Consequently, the right-hand side of (7.16) converges to 0 when $T_1, T_2 \to \infty$, $T_1 < T_2$. Hence, the sequence of functions $J_T(\Pi, \cdot)$, $T = 1, 2, \ldots$, is convergent to some w-bounded limit $J_\infty(\Pi, \cdot) \in \mathscr{V}$. The convergence is w-uniform, that is,

$$\lim_{T \to \infty} \sup_{x \in \widetilde{\mathscr{X}}} \frac{\left| J_T(\Pi, x) - J_\infty(\Pi, x) \right|}{w(x)} = 0.$$

If the model is uniformly risk transient, then the estimate (7.16) is the same for all Markov policies Π, and thus $\|J_\infty(\Pi, \cdot)\|_w$ is uniformly bounded. Moreover,

$$\lim_{T \to \infty} \sup_{\substack{x \in \widetilde{\mathscr{X}} \\ \Pi \in \Pi^{\mathrm{DM}}}} \frac{\left| J_T(\Pi, x) - J_\infty(\Pi, x) \right|}{w(x)} = 0,$$

where Π^{DM} is the set of all stationary deterministic Markov policies. As each of the functions $(\pi, x) \mapsto J_T(\Pi, x)$ is lower semicontinuous, so is the limit function $(\pi, x) \mapsto J_\infty(\Pi, x)$. $\quad\square$

Remark 7.1. It is clear from the proof of Theorem 7.2 that

$$(7.17) \qquad J_\infty(\Pi, x_1) = \lim_{T \to \infty} \rho_{1,T}^\Pi \big(0, Z_2, \ldots, Z_T + f(x_T) \big)$$

for any w-bounded measurable function $f : \mathscr{X} \to \mathbb{R}$, because $c(x_{T-1}, u_t, x_T) + f(x_T)$ is still w-bounded.

This analysis allows us to derive policy evaluation equations for the infinite horizon problem in the case of a fixed Markov policy.

THEOREM 7.3. *Suppose a controlled Markov model with a transition risk mapping $\sigma(\cdot, \cdot, \cdot)$ is risk transient for the stationary Markov policy $\Pi = \{\pi, \pi, \ldots\}$ with some weight function $w(\cdot)$. If condition (G3) is satisfied, then a w-bounded function $v \in \mathscr{V}$ satisfies the equations*

$$(7.18) \qquad v(x) = \sigma\big(c(x, \pi(x), \cdot) + v(\cdot), x, Q(x, \pi(x))\big), \quad x \in \widetilde{\mathscr{X}},$$
$$(7.19) \qquad v(x_{\mathrm{A}}) = 0$$

if and only if $v(x) = J_\infty(\Pi, x)$ for all $x \in \mathscr{X}$.

Proof. Suppose a w-bounded function $v \in \mathscr{V}$ satisfies (7.18)–(7.19). By (G3), the function $c(x, \pi(x), \cdot) \in \mathscr{V}$, and thus the right-hand side of (7.18) is well-defined. Iterating (7.18), we obtain for all $x_1 \in \mathscr{X}$ the following equation:

$$v(x_1) = \rho_1^\Pi \Big(c(x_1, \pi(x_1), x_2) + \rho_2^\Pi \Big(c(x_2, \pi(x_2), x_3) + \cdots$$
$$+ \rho_T^\Pi \big(c(x_T, \pi(x_T), x_{T+1}) + v(x_{T+1}) \big) \cdots \Big) \Big).$$

Denote $Z_t = c(x_{t-1}, \pi(x_{t-1}), x_t)$. Using subadditivity and monotonicity of the conditional risk measures we deduce that

$$
\begin{aligned}
v(x_1) &= \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} + v(x_{T+1}) \big) \\
&\leq \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} \big) + \rho_{1,T+1}^\Pi \big(0, 0, \ldots, v(x_{T+1}) \big) \\
&\leq \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} \big) + \rho_{1,T+1}^\Pi \big(0, 0, \ldots, |v(x_{T+1})| \big).
\end{aligned}
$$
(7.20)

By convexity of $\rho_{1,T+1}^\Pi(\cdot)$,
(7.21)
$$
\begin{aligned}
2\rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} \big) & \\
&\leq \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} + v(x_{T+1}) \big) + \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} - v(x_{T+1}) \big) \\
&= v(x_1) + \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} - v(x_{T+1}) \big).
\end{aligned}
$$

In a similar way to (7.20),

$$
\begin{aligned}
\rho_{1,T+1}^\Pi & \big(0, Z_2, \ldots, Z_{T+1} - v(x_{T+1}) \big) \\
&\leq \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} \big) + \rho_{1,T+1}^\Pi \big(0, 0, \ldots, -v(x_{T+1}) \big) \\
&\leq \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} \big) + \rho_{1,T+1}^\Pi \big(0, 0, \ldots, |v(x_{T+1})| \big).
\end{aligned}
$$

Substituting into (7.21) we obtain

$$v(x_1) \geq \rho_{1,T+1}^\Pi \big(0, Z_2, \ldots, Z_{T+1} \big) - \rho_{1,T+1}^\Pi \big(0, 0, \ldots, |v(x_{T+1})| \big).$$

Combining this estimate with (7.20), we conclude that

(7.22) $$\big| v(x_1) - J_T(\Pi, x_1) \big| \leq \rho_{1,T+1}^\Pi \big(0, 0, \ldots, |v(x_{T+1})| \big).$$

Consider the function

$$d_{1,T}(x_1) = \rho_{1,T+1}^\Pi \big(0, 0, \ldots, |v(x_{T+1})| \big).$$

Proceeding exactly as in the proof of Theorem 7.2, we obtain a representation similar to (7.13):

$$d_{1,T} = \widetilde{M}_1 \cdots \widetilde{M}_T |v|$$

with $\widetilde{M}_j \prec \widetilde{\mathfrak{M}}^\pi$, $j = 1, \ldots, T$. Thus, $d_{1,T} = A_T |v|$ with $A_T \prec (\widetilde{\mathfrak{M}}^\pi)^T$. By Definition 7.1, for any sequence of selectors $A_t \prec (\widetilde{\mathfrak{M}}^\pi)^t$, $t = 1, 2 \ldots$, we have $\big\| \sum_{t=1}^\infty A_t \big\|_w \leq K$. Therefore, $\big\| A_t \big\|_w \to 0$ and $\big\| d_{1,t} \big\|_w \to 0$, as $t \to \infty$. Using this in (7.22) we conclude that $v(\cdot) \equiv J_\infty(\Pi, \cdot)$, as postulated.

To prove the converse implication we can use the fact that all conditional risk measures $\rho_t^\Pi(\cdot)$ share the same transition risk mapping $\sigma(\cdot, \cdot, \cdot)$ to rewrite (7.2) as follows:

$$J_T(\Pi, x_1) = \sigma\big(c(x_1, \pi(x_1), \cdot) + J_{T-1}(\Pi, \cdot), x_1, Q(x_1, \pi(x_1))\big).$$

The function $\sigma(\cdot, x_1, \mu)$, as a finite-valued coherent measure of risk on a Banach lattice $\mathscr{V}$, is continuous (see [43, Prop. 3.1]). By Theorem 7.2, the sequence $\big\{J_T(\Pi, \cdot)\big\}$ is convergent to $J_\infty(\Pi, \cdot)$ in the space $\mathscr{V}$, and $J_\infty(\Pi, \cdot)$ is w-bounded. Therefore,

$$\lim_{T \to \infty} J_T(\Pi, x_1) = \sigma\big(c(x_1, \pi(x_1), \cdot) + \lim_{T \to \infty} J_{T-1}(\Pi, \cdot), x_1, Q(x_1, \pi(x_1))\big).$$

This is identical to (7.18) with $v(\cdot) \equiv J_\infty(\Pi, \cdot)$. Equation (7.19) is obvious. $\quad\square$

8. Dynamic programming equations for infinite-horizon problems. We shall now focus on the optimal value function

$$(8.1) \qquad J^*(x) = \inf_{\Pi \in \Pi^{\mathrm{DM}}} J_\infty(\Pi, x), \quad x \in \mathscr{X},$$

where Π^{DM} is the set of all stationary deterministic Markov policies. To simplify notation, we define the operators $\mathfrak{D} : \mathscr{V} \to \mathscr{V}$ and $\mathfrak{D}_\pi : \mathscr{V} \to \mathscr{V}$ as follows:

$$(8.2) \qquad [\mathfrak{D}v](x) = \min_{u \in U(x)} \sigma\big(c(x, u, \cdot) + v(\cdot), x, Q(x, u)\big), \quad x \in \mathscr{X},$$

$$(8.3) \qquad [\mathfrak{D}_\pi v](x) = \sigma\big(c(x, \pi(x), \cdot) + v^k(\cdot), x, Q(x, \pi(x))\big), \quad x \in \mathscr{X},$$

where $\pi \lessdot U$. Owing to the monotonicity of $\sigma(\cdot, x, \mu)$, both operators are nondecreasing. By construction, $\mathfrak{D}v \leq \mathfrak{D}_\pi v$ for all $v \in \mathscr{V}$ and all $\pi \lessdot U$.

THEOREM 8.1. *Assume that conditions* (G0)–(G4) *are satisfied and that the model is uniformly risk transient. Then a measurable w-bounded function $v : \mathscr{X} \to \mathbb{R}$ satisfies the equations*

$$(8.4) \qquad v(x) = \inf_{u \in U(x)} \sigma\big(c(x, u, \cdot) + v(\cdot), x, Q(x, u)\big), \quad x \in \mathscr{X},$$

$$(8.5) \qquad v(x_{\mathrm{A}}) = 0$$

if and only if $v(x) = J^*(x)$ for all $x \in \mathscr{X}$. Moreover, a measurable minimizer $\pi^*(x)$, $x \in \mathscr{X}$, on the right-hand side of (8.4) *exists and defines an optimal deterministic Markov policy $\Pi^* = \{\pi^*, \pi^*, \dots\}$.*

Proof. Consider a sequence of Markov deterministic policies $\Pi^k = \{\pi^k, \pi^k, \dots\}$, $k = 1, 2, \dots$, constructed in the following way. We choose any $\pi^1 \lessdot U$. Its value $v^1(\cdot) = J_\infty(\Pi^1, \cdot)$ is then given by (7.18)–(7.19). For $k = 1, 2, \dots$ we determine $\pi^{k+1}(\cdot)$ as the measurable solution of the problem

$$(8.6) \qquad \min_{u \in U(x)} \sigma\big(c(x, u, \cdot) + v^k(\cdot), x, Q(x, u)\big), \quad x \in \mathscr{X},$$

which exists by Proposition 5.2. The corresponding value of the policy $\Pi^{k+1} = \{\pi^{k+1}, \pi^{k+1}, \dots\}$ is the function $v^{k+1}(\cdot) = J_\infty(\Pi^{k+1}, \cdot)$, and the iteration continues. By construction, the sequences $\{\pi^k\}$ and $\{v^k\}$ satisfy the relations

$$(8.7) \qquad \mathfrak{D}_{\pi^{k+1}} v^k = \mathfrak{D}v^k \leq \mathfrak{D}_{\pi^k} v^k = v^k.$$

Applying the operator $\mathfrak{D}_{\pi^{k+1}}$ to this relation, we deduce from its monotonicity that

$$(8.8) \qquad [\mathfrak{D}_{\pi^{k+1}}]^{T-1}v^k \leq \mathfrak{D}_{\pi^{k+1}}v^k = \mathfrak{D}v^k \leq v^k, \quad T = 2, 3, \ldots.$$

Relation (8.8) can be equivalently written as

$$\rho_{1,T}^{\Pi^{k+1}}\left(0, Z_2, \ldots, Z_T + v^k(x_T)\right) \leq [\mathfrak{D}v^k](x_1) \leq v^k(x_1),$$

where $Z_t = c(x_{t-1}, u_{t-1}, x_t)$, $t = 2, 3, \ldots, T$, is the cost sequence resulting from the policy Π^{k+1}. Passing to the limit with $T \to \infty$, from Remark 7.1 we conclude that the sequence $\{v^k\}$ is nonincreasing:

$$(8.9) \qquad v^{k+1}(x) = J_\infty(\Pi^{k+1}, x) \leq [\mathfrak{D}v^k](x) \leq v^k(x), \quad x \in \mathscr{X}, \quad k = 0, 1, 2, \ldots.$$

Since $v^k \geq J^*$, the sequence $\{v^k\}$ is monotonically convergent to some limit $v^\infty \geq J^*$. By the Lebesgue theorem, it is also convergent in the space $\mathscr{V}$. As the function $\sigma(\cdot, x, \mu)$ is a coherent measure of risk, it follows from [43, Prop. 3.1] that it is continuous on $\mathscr{V}$ and thus

$$(8.10) \quad \sigma\left(c(x, u, \cdot) + v^k(\cdot), x, Q(x, u)\right) \downarrow \sigma\left(c(x, u, \cdot) + v^\infty(\cdot), x, Q(x, u)\right),$$
$$\text{as} \quad k \to \infty \quad \forall\, u \in U(x).$$

The left inequality in (8.9) also implies that

$$(8.11) \qquad v^{k+1}(x) \leq \sigma\left(c(x, u, \cdot) + v^k(\cdot), x, Q(x, u)\right) \quad \forall\, u \in U(x).$$

Passing to the limit with $k \to \infty$ on both sides of (8.11) and using (8.10), we conclude that

$$v^\infty(x) \leq \sigma\left(c(x, u, \cdot) + v^\infty(\cdot), x, Q(x, u)\right) \quad \forall\, u \in U(x).$$

Since this is true for all $x \in \mathscr{X}$ and all $u \in U(x)$, it follows that

$$(8.12) \qquad v^\infty(x) \leq [\mathfrak{D}v^\infty](x) = \min_{u \in U(x)} \sigma\left(c(x, u, \cdot) + v^\infty(\cdot), x, Q(x, u)\right), \quad x \in \mathscr{X}.$$

Iterating this inequality, we conclude that for every feasible decision rule π and every $T = 1, 2, \ldots,$

$$v^\infty \leq [\mathfrak{D}]^T v^\infty \leq [\mathfrak{D}_\pi]^T v^\infty.$$

Owing to Remark 7.1, the right-hand side of this inequality is convergent to $J_\infty(\Pi, \cdot)$ when $T \to \infty$. Since Π was arbitrary, $v^\infty \leq J^*$, and thus $v^\infty = J^*$. It remains to show that v^∞ satisfies (8.4). It follows from the monotonicity of the operator $\mathfrak{D}$ and relation (8.7) that

$$\mathfrak{D}v^\infty \leq \mathfrak{D}v^k \leq v^k.$$

Passing to the limit with $k \to \infty$ we conclude that $\mathfrak{D}v^\infty \leq v^\infty$. This combined with (8.12) yields $v^\infty = \mathfrak{D}v^\infty$, which is (8.4). Denoting by $\pi^*(x)$ the (measurable) minimizer on the right-hand side of (8.12), we also see that $v^\infty = \mathfrak{D}_{\pi^*}v^\infty$. By Theorem 7.3, $v^\infty(\cdot) = J_\infty(\Pi^*, \cdot)$.

To prove the converse implication, suppose $v \in \mathscr{V}$ satisfies (8.4)–(8.5) and $\|v\|_w < \infty$. By Proposition 5.2, a measurable minimizer $\hat{\pi}(x)$ on the right-hand side of (8.4) exists. We obtain the equation

$$v(x) = \sigma\big(c(x, \hat{\pi}(x), \cdot) + v(\cdot), x, Q(x, \hat{\pi}(x))\big), \quad x \in \mathscr{X}.$$

Due to Theorem 7.3,

$$(8.13) \qquad\qquad v(x) = J_\infty(\hat{\Pi}, x) \geq J^*(x), \quad x \in \mathscr{X},$$

where $\hat{\Pi} = \{\hat{\pi}, \hat{\pi}, \dots\}$. On the other hand, it follows from (8.4) that for any stationary deterministic Markov policy $\Pi = \{\pi, \pi, \dots\}$ we have

$$(8.14) \qquad v(x) \leq \sigma\big(c(x, \pi(x), \cdot) + v(\cdot), x, Q(x, \pi(x))\big), \quad x \in \mathscr{X}.$$

The transition risk mapping σ is nondecreasing with respect to the first argument. Therefore, iterating inequality (8.14) we obtain the following inequality:

$$v(x_1) \leq \rho_{1,T}^{\Pi}\big(0, Z_2, \dots, Z_T + v(x_T)\big).$$

Passing to the limit with $T \to \infty$ and applying Remark 7.1, we obtain for all stationary deterministic Markov policies $\Pi = \{\pi, \pi, \dots\}$ the inequality $v(\cdot) \leq J_\infty(\Pi, \cdot)$. The last estimate together with (8.13) implies that $v(\cdot) \equiv J^*(\cdot)$ and that the stationary policy $\hat{\Pi}$ is optimal. $\quad\square$

Remark 8.1. The construction of the sequence of stationary Markov policies $\{\Pi^k\}$ and their corresponding values $\{v^k\}$ employed in the first part of the proof of Theorem 8.1 can be interpreted as a risk-averse version of the policy iteration method. For more information on the risk-averse policy iteration method, the reader is referred to Çavuş and Ruszczyński [10], where methods for risk-averse models with finite state and control spaces are considered.

We can now address the case of nonstationary deterministic policies. For a deterministic policy $\Lambda = \{\lambda_1, \lambda_2, \dots\}$ we define

$$J_\infty(\Lambda, x_1) = \liminf_{T \to \infty} J_T(\Lambda, x_1) \qquad \text{and} \qquad \hat{J}(x_1) = \inf_\Lambda J_\infty(\Lambda, x_1).$$

In the theorem below, we make an additional technical assumption that the function $\hat{J}(\cdot)$ is measurable. It is obviously satisfied for a finite or countable state space. In general, its verification, even in the expected value case, requires additional assumptions (see, e.g., [19, section 9.3]). We discuss some sufficient conditions after the theorem.

THEOREM 8.2. *Assume that conditions* (G0)–(G4) *are satisfied and that the model is uniformly risk transient. Additionally, assume that the function $\hat{J}(\cdot)$ is measurable and a constant C exists such that $J_\infty(\Lambda, x) \geq -Cw(x)$ for all $x \in \mathscr{X}$ and for all policies Λ. Then a w-bounded function $v \in \mathscr{V}$ satisfies the equations* (8.4)–(8.5) *if and only if $v(x) = \hat{J}(x)$ for all $x \in \mathscr{X}$. Moreover, a measurable minimizer $\pi^*(x)$, $x \in \mathscr{X}$, on the right-hand side of* (8.4) *exists and defines an optimal deterministic policy $\Pi^* = \{\pi^*, \pi^*, \dots\}$.*

Proof. As for stationary Markov policies Π we have $\|J_\infty(\Pi, \cdot)\|_w < \infty$, in view of the additional assumption we have $\|\hat{J}\|_w < \infty$. Denote $\Lambda^2 = \{\lambda_2, \lambda_3, \dots\}$. Due to

the monotonicity and continuity of $\rho_1^\Pi(\cdot)$ we obtain the chain of relations

$$
\begin{aligned}
\hat{J}(x_1) &= \inf_{\lambda_1,\lambda_2,\dots} \liminf_{T\to\infty} \rho_1^{\lambda_1}\big(c(x_1,\lambda_1(x_1),x_2) + J_{T-1}(\Lambda^2,x_2)\big) \\
&\geq \inf_{\lambda_1,\lambda_2,\dots} \liminf_{T\to\infty} \rho_1^{\lambda_1}\big(c(x_1,\lambda_1(x_1),x_2) + \inf_{\tau\geq T-1} J_\tau(\Lambda^2,x_2)\big) \\
&= \inf_{\lambda_1,\lambda_2,\dots} \lim_{T\to\infty} \rho_1^{\lambda_1}\big(c(x_1,\lambda_1(x_1),x_2) + \inf_{\tau\geq T-1} J_\tau(\Lambda^2,x_2)\big) \\
&= \inf_{\lambda_1,\lambda_2,\dots} \rho_1^{\lambda_1}\Big(c(x_1,\lambda_1(x_1),x_2) + \liminf_{T\to\infty} J_{T-1}(\Lambda^2,x_2)\Big) \\
&= \inf_{\lambda_1,\lambda_2,\dots} \rho_1^{\lambda_1}\big(c(x_1,\lambda_1(x_1),x_2) + J_\infty(\Lambda^2,x_2)\big).
\end{aligned}
$$

Owing to the monotonicity of $\rho_1(\cdot)$, we can move the minimization with respect to Λ^2 inside:

$$
\hat{J}(x_1) \geq \inf_{\lambda_1} \rho_1^{\lambda_1}\Big(c(x_1,\lambda_1(x_1),x_2)+\inf_{\Lambda^2} J_\infty(\Lambda^2,x_2)\Big) = \inf_{\lambda_1} \rho_1^{\lambda_1}\big(c(x_1,\lambda_1(x_1),x_2)+\hat{J}(x_2)\big).
$$

Observe that we assumed $\hat{J}(\cdot)$ to be measurable, and thus the right-hand side of the last inequality is well-defined. Thus $\hat{J}(\cdot)$ satisfies the inequality

$$
(8.15) \qquad \hat{J}(x) \geq \inf_{u\in U(x)} \sigma\big(c(x,u,\cdot) + \hat{J}(\cdot), x, Q(x,u)\big), \quad x \in \mathscr{X}.
$$

We can now repeat the argument from the proof of Theorem 8.1. Denote by $\hat{\lambda}(x)$ the minimizer in (8.15), which exists by Proposition 5.2. Iterating inequality (8.15), and passing to the limit we conclude that

$$
\hat{J}(x) \geq J_\infty(\hat{\Lambda}, x), \quad x \in \mathscr{X},
$$

where $\hat{\Lambda} = \{\hat{\lambda}, \hat{\lambda}, \dots\}$ is a stationary Markov policy. Therefore, optimization with respect to stationary Markov policies is sufficient, and the result follows from Theorem 8.1. $\quad\square$

To prove that our additional technical assumption that $\hat{J}(\cdot)$ is measurable is true for nonnegative costs $c(\cdot,\cdot,\cdot)$, we shall represent $\hat{J}(\cdot)$ as a limit of functions $v^k(\cdot)$ defined by the *value iteration method*:

$$
(8.16) \qquad\qquad v^{k+1} = \mathfrak{D}v^k, \quad k = 0,1,2,\dots,
$$

where $v^0 = 0$. For details on the risk-averse value iteration method, the reader is referred to Çavuş and Ruszczyński [10], where methods for risk-averse models with finite state and control spaces are considered.

PROPOSITION 8.3. *Assume that conditions* (G0)–(G4) *are satisfied, the mapping* $\sigma(\cdot,x,\cdot)$ *is continuous,* $c(\cdot,\cdot,\cdot)$ *is nonnegative, and the model is uniformly risk transient. Then the sequence* $\{v^k\}$ *defined by the value iteration method satisfies the relation*

$$
\hat{J}(x) = \lim_{k\to\infty} v^k(x), \quad x \in \mathscr{X}.
$$

Moreover, $\hat{J} \in \mathscr{V}$.

Proof. Owing to the nonnegativity of the costs and to the monotonicity of $\sigma(\cdot,x,m)$, the sequence $\{v^k\}$ satisfies the inequalities

$$
(8.17) \qquad\qquad v^k \leq v^{k+1} \leq \hat{J} \leq J^*, \quad k = 0,1,2,\dots,
$$

where J^* is the best value of stationary Markov policies. Hence, this sequence has a pointwise limit v^∞, which is also (by the Lebesgue theorem) a limit in $\mathscr{V}$. As $\sigma(\cdot, x, \cdot)$ is continuous, due to [4, Thm. 1.4.16], the operator $\mathfrak{D}$ is continuous. It follows that

$$v^\infty = \lim_{k\to\infty} v^{k+1} = \lim_{k\to\infty} \mathfrak{D}v^k = \mathfrak{D}v^\infty.$$

By Theorem 8.1 and relations (8.17), $v^\infty = \hat{J} = J^*$. $\square$

9. Randomized decision rules. Although we focus on deterministic policies in this paper, it may be of use to provide preliminary discussion of models with randomized policies. If randomized policies are allowed, we need to revise our definition of the transition risk mapping, to account for the fact that for a given state $x_t = x$, both u_t and x_{t+1} are random. In this case, it is convenient to consider functions on the product space $\mathscr{U} \times \mathscr{X}$ equipped with its product Borel σ-algebra $\mathscr{B}$.

Suppose the current state is x and we use a randomized decision rule π. This control, together with the transition kernel Q defines a probability measure $[\pi Q]_x$ on the product space $\mathscr{U} \times \mathscr{X}$ as follows:

$$(9.1) \quad [\pi \circ Q]_x(B_u \times B_y) = \int_{B_u} Q(B_y | x, u)\, \pi(du | x), \quad B_u \in \mathscr{B}(U), \quad B_y \in \mathscr{B}(\mathscr{X}).$$

The measure is extended to other sets in $\mathscr{B}$ in the usual way.

The cost incurred at the current stage is given by the function $c(x, \cdot, \cdot)$ on the product space $\mathscr{U} \times \mathscr{X}$.

The construction below parallels our earlier presentation in section 3, but with notational complication resulting from the need to deal with the product space $\mathscr{U} \times \mathscr{X}$. We define $\overline{\mathscr{V}} = \mathscr{L}_p(\mathscr{U} \times \mathscr{X}, \mathscr{B}, P_0)$, where $p \in [1, \infty)$, where $\overline{P}_0$ is some reference probability measure on $\mathscr{U} \times \mathscr{X}$. The dual space $\overline{\mathscr{V}}'$ is the space of signed measures m on $(\mathscr{U} \times \mathscr{X}, \mathscr{B})$, which are absolutely continuous with respect to $\overline{P}_0$, with densities (Radon–Nikodym derivatives) lying in the space $\mathscr{L}_q(\mathscr{U} \times \mathscr{X}, \mathscr{B}, P_0)$, where $1/p + 1/q = 1$. We consider the set of probability measures in $\overline{\mathscr{V}}'$:

$$\mathscr{M} = \left\{ m \in \overline{\mathscr{V}}' : m(\mathscr{U} \times \mathscr{X}) = 1,\ m \geq 0 \right\}.$$

We also assume that the spaces $\overline{\mathscr{V}}$ and $\overline{\mathscr{V}}'$ are endowed with topologies that make them paired topological vector spaces with the bilinear form

$$\langle \varphi, m \rangle = \int_{\mathscr{U} \times \mathscr{X}} \varphi(u, y)\, m(du \times dy), \quad \varphi \in \overline{\mathscr{V}}, \quad m \in \overline{\mathscr{V}}'.$$

We can now define a *transition risk mapping* as a measurable function $\sigma : \mathscr{V} \times \mathscr{X} \times \mathscr{M} \to \mathbb{R}$, such that for every $x \in \mathscr{X}$ and every $m \in \mathscr{M}$, the function $\varphi \mapsto \sigma(\varphi, x, m)$ is a coherent measure of risk on $\mathscr{V}$.

Consider a controlled Markov process $\{x_t\}$ with some randomized Markov policy $\Pi = \{\pi_1, \pi_2, \dots\}$. For a fixed time t and a measurable w-bounded function $g : \mathscr{X} \times \mathscr{U} \times \mathscr{X} \to \mathbb{R}$, the value of $Z_{t+1} = g(x_t, u_t, x_{t+1})$, where u_t is distributed according to $\pi_t(x_t)$, is a random variable.

We define now a *Markov one-step conditional risk measure* $\rho_t(\cdot)$ by requiring that

$$\rho_t\big(g(x_t, u_t, x_{t+1})\big) = \sigma\big(g(x_t, \cdot, \cdot), x_t, [\pi_t \circ Q]_{x_t}\big) \quad \text{a.s.}$$

This is analogous to (3.4).

With these modifications, we can repeat the analysis presented in sections 3–8, with formal adjustments accounting for the use of randomized policies. For example, the dynamic programming equations (8.4)–(8.5) would take on the form:

$$(9.2) \qquad v(x) = \inf_{\lambda(x) \in \mathscr{P}(U(x))} \sigma\big(c(x, \cdot, \cdot) + v(\cdot), x, [\lambda \circ Q]_x\big), \quad x \in \mathscr{X},$$

$$(9.3) \qquad v(x_{\mathrm{A}}) = 0.$$

Justification of these equations, though, is straightforward only in the case of *finite control sets* $U(x)$. Otherwise, serious technical difficulties arise, associated with the nonlinearity of the mapping in (9.2). With the present state of our knowledge, though, we may already make an observation that a randomized policy may be strictly better than a deterministic policy.

Observe that the mapping $\lambda(x) \mapsto \sigma\big(c(x, \cdot, \cdot) + v(\cdot), x, [\lambda \circ Q]_x\big)$, which plays the key role in the dynamic programming equation (9.2), is nonlinear, in general, as opposed to the expected value model, where

$$\sigma\big(c(x, \cdot, \cdot) + v(\cdot), x, [\lambda \circ Q]_x\big) = \int_{U(x)} \int_{\mathscr{X}} \big(c(x, u, y) + v(y)\big)\, Q(dy|x, u)\, \lambda(du|x).$$

In the expected value case, it is sufficient to consider only the extreme points of the set $\mathscr{P}\big(U(x)\big)$, which are the measures assigning unit mass to one of the controls $u \in U(x)$:

$$\inf_{\lambda \in \mathscr{P}(U(x))} \int_{U(x)} \int_{\mathscr{X}} \big(c(x, u, y) + v(y)\big)\, Q(dy|x, u)\, \lambda(du|x)$$
$$= \inf_{u \in U(x)} \int_{\mathscr{X}} \big(c(x, u, y) + v(y)\big)\, Q(dy|x, u).$$

In the risk-averse case, this simplification is not justified and a randomized policy may be strictly better than any deterministic policy.

A question arises whether it is possible to identify cases in which deterministic policies are sufficient. It turns out that we can prove this for the conditional average value at risk from Example 3.2, which in our setting takes on the following form:

$$(9.4) \quad \sigma(\varphi, x, \mu) = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{\alpha} \int_{U(x) \times \mathscr{X}} \big(\varphi(u, y) - \eta\big)_+ \mu(du \times dy) \right\}, \quad \alpha \in (0, 1).$$

LEMMA 9.1. *If the transition risk mapping has the form* (9.4) *then the dynamic programming equations* (9.2) *have a solution in deterministic decision rules.*

Proof. Interchanging the integration and the infimum in (9.4), we obtain a lower bound

$$\sigma(\varphi, x, [\lambda \circ Q]_x) = \inf_{\eta \in \mathbb{R}} \int_{U(x)} \int_{\mathscr{X}} \left\{ \eta + \frac{1}{\alpha}\big(\varphi(u, y) - \eta\big)_+ \right\}\, Q(dy|x, u)\, \lambda(du|x)$$
$$\geq \int_{U(x)} \inf_{\eta \in \mathbb{R}} \int_{\mathscr{X}} \left\{ \eta + \frac{1}{\alpha}\big(\varphi(u, y) - \eta\big)_+ \right\}\, Q(dy|x, u)\, \lambda(du|x).$$

The above inequality becomes an equation for every Dirac measure λ. On the right-hand side of (9.2) we have

$$\inf_{\lambda(x) \in \mathscr{P}(U(x))} \sigma\big(c(x, \cdot, \cdot) + v(\cdot), x, [\lambda \circ Q]_x\big)$$
$$\geq \inf_{\lambda \in \mathscr{P}(U(x))} \int_{U(x)} \inf_{\eta \in \mathbb{R}} \int_{\mathscr{X}} \left\{ \eta + \frac{1}{\alpha}\big(\varphi(u, y) - \eta\big)_+ \right\}\, Q(dy|x, u)\, \lambda(du|x).$$

As the right-hand side achieves its minimum over $\lambda \in \mathscr{P}(U(x))$ at a Dirac measure concentrated at an extreme point of $U(x)$, and both sides coincide in this case, the minimum of the left-hand side is also achieved at such a measure. Consequently, for transition risk mappings of form (9.4) deterministic Markov policies are optimal. $\square$

In general, randomized policies may be better, as the example in section 10.2 illustrates.

10. Illustrative examples. We illustrate our models and results on two examples.

10.1. Asset selling. Let us at first consider the classical example of asset selling originating from Karlin [22]. Offers S_t arriving in time periods $t = 1, 2, \dots$ are independent integer-valued integrable random variables, distributed according to measure P. At each time we may accept the highest offer received so far, or we may wait, in which case a waiting cost $c_0 > 0$ is incurred. Denoting the random stopping time by τ we see that the total "cost" equals $Z = c_0(\tau - 1) - \max_{1 \le t \le \tau} S_t$. The problem is an example of an *optimal stopping problem*, a structure of considerable theoretical and practical relevance (see, e.g., Çınlar [11], Dynkin and Yushkevich [13, 14], and Puterman [36]).

Formally, we introduce the state space $\mathscr{X} = \{x_{\mathrm{A}}\} \cup \{0, 1, 2, \dots\}$, where x_{A} is the absorbing state reached after the transaction, and the other states represent the highest offer received so far. The control space is $\mathscr{U} = \{0, 1\}$, with 0 representing "wait" and 1 representing "sell." The state evolves according to the equation

$$x_{t+1} = \begin{cases} \max(x_t, S_{t+1}) & \text{if } u_t = 0, \\ x_{\mathrm{A}} & \text{if } u_t = 1. \end{cases}$$

Denoting by $F_S(\cdot)$ the distribution function of S, we can write the controlled transition kernel Q as follows:

$$Q(y|x, 0) = \begin{cases} P(y) & \text{if } y > x, \\ F_S(x) & \text{if } y = x, \\ 0 & \text{if } y < x, \end{cases} \qquad Q(y|x, 1) = \begin{cases} 1 & \text{if } y = x_{\mathrm{A}}, \\ 0 & \text{if } y \neq x_{\mathrm{A}}. \end{cases}$$

The cost function is

$$c(x, u, y) = \begin{cases} c_0 & \text{if } u = 0, \\ -x & \text{if } u = 1. \end{cases}$$

The expected value version of this problem has a known solution: accept the first offer greater than or equal to the solution $\hat{x}$ of the equation

$$(10.1) \qquad c_0 = \sum_{s=0}^{\infty} (s - \hat{x})_+ \, P(s).$$

We shall solve the risk-averse version of the problem. We choose $P_0(y) = P(y)/2$ for $y \in \mathbb{N}$, and $P_0(x_{\mathrm{A}}) = 1/2$. The space $\mathscr{V}$ is the space of P_0-integrable functions $v : \mathscr{X} \to \mathbb{R}$. The space $\mathscr{M}$ is the space of probability measures μ on $\mathscr{X}$, for which

$$\sup \left\{ \frac{\mu(y)}{P_0(y)} : P_0(y) > 0, \ y \in \mathscr{X} \right\} < \infty.$$

Observe that all measures $Q(\cdot, x, u)$ are elements of $\mathscr{M}$.

Suppose $\sigma : \mathscr{V} \times \mathscr{X} \times \mathscr{M} \to \mathbb{R}$ is a transition risk mapping. Owing to (A3), and to the fact that $v(x_{\mathrm{A}}) = 0$, we have

$$\sigma\big(c(x, u, \cdot) + v(\cdot), x, Q(x, u)\big) = \begin{cases} c_0 + \sigma\big(v(\cdot), x, Q(x, 0)\big) & \text{if } u = 0, \\ -x & \text{if } u = 1. \end{cases}$$

Equation (8.4) takes on the form

$$(10.2) \qquad v(x) = \min\Big\{ -x, c_0 + \sigma\big(v(\cdot), x, Q(x, 0)\big) \Big\}, \quad x = 0, 1, 2, \dots.$$

Suppose σ is law invariant (Definition 3.2). As the distribution of $v(\cdot)$ with respect to the measure $Q(x, 0)$ on $\mathbb{N}$ is the same as the distribution of $v\big(\max(x, S)\big)$ under the measure P of S, we obtain

$$\sigma\big(v(\cdot), x, Q(x, 0)\big) = \sigma\big(v\big(\max(x, \cdot)\big), x, P\big).$$

Suppose our attitude to risk does not depend on the current state, that is, $\sigma(\cdot, \cdot, \cdot)$ does not depend on its second argument. Using (4.2), we may rewrite the last equation as follows:

$$\sigma\big(v(\cdot), x, Q(x, 0)\big) = \max_{\mu \in \mathscr{A}} \sum_{s=0}^{\infty} v\big(\max(x, s)\big)\, \mu(s).$$

The convex closed set of probability measures $\mathscr{A}$ is fixed, because $\sigma(\cdot, \cdot, \cdot)$ does not depend on its second argument, and the third argument, P, is now fixed. Equation (10.2) takes on the form

$$(10.3) \qquad v(x) = \min\left\{ -x, c_0 + \max_{\mu \in \mathscr{A}} \sum_{s=0}^{\infty} v\big(\max(x, s)\big)\, \mu(s) \right\}, \quad x = 0, 1, 2 \dots.$$

Observe that $v(x) \leq -x$ and thus $v\big(\max(x, s)\big) \leq -\max(x, s)$. The last displayed inequality implies that

$$v(x) \leq \min\left\{ -x, c_0 + \max_{\mu \in \mathscr{A}} \sum_{s=0}^{\infty} \big[-\max(x, s) \big]\, \mu(s) \right\}$$

$$= \min\left\{ -x, c_0 - \min_{\mu \in \mathscr{A}} \sum_{s=0}^{\infty} \max(x, s)\, \mu(s) \right\}, \quad x = 0, 1, 2, \dots.$$

If the offer at level x is accepted, then $v(x) = -x$. After simplifications, we obtain the inequality

$$\min_{\mu \in \mathscr{A}} \sum_{s=0}^{\infty} (s - x)_+\, \mu(s) \leq c_0.$$

This suggests the solution: *accept any offer that is greater or equal to the solution x^* of the equation*

$$(10.4) \qquad \min_{\mu \in \mathscr{A}} \sum_{s=0}^{\infty} (s - x)_+\, \mu(s) = c_0;$$

if $x < x^*$, then wait. The corresponding value function is $v^*(x) = -\max(x, x^*)$. Equation (10.3) can be verified by direct substitution.

Let us check whether the solution of (10.4) exists and is unique. To guarantee that a solution of (10.4) exists, it is sufficient to assume that $\mathbb{E}_\mu[S] = \sum_{s=0}^\infty s\mu(s) > c_0$ for all $\mu \in \mathscr{A}$. Since the left-hand side of (10.4), as a function of x, is continuous, nonincreasing, and converges to 0, when $x \to \infty$, it must attain the value c_0 at some point x^*. Define μ^* as the measure in $\mathscr{A}$ for which the minimum in (10.4) with $x = x^*$ is attained. It exists, because $\mathscr{A}$ is weakly* compact. We thus have $\sum_{s=0}^\infty (s - x^*)_+ \, \mu^*(s) = c_0$. Since $c_0 > 0$, we see that $\mu^*\{s : s > x^*\} > 0$. Therefore, for every $x > x^*$ we obtain

$$\min_{\mu \in \mathscr{A}} \sum_{s=0}^\infty (s - x)_+ \, \mu(s)$$

$$\leq \sum_{s=0}^\infty (s - x)_+ \, \mu^*(s) = \sum_{s>x} (s - x) \, \mu^*(s) \leq \sum_{s>x^*} (s - x) \, \mu^*(s)$$

$$= \sum_{s>x^*} (s - x^*) \, \mu^*(s) - (x - x^*) \, \mu^*\{s : s > x^*\} < \sum_{s>x^*} (s - x^*) \, \mu^*(s) = c_0.$$

Consequently, x^* is the largest solution of (10.4). No other solution may exist, because they cannot both be the largest.

The solution (10.4) of the risk-averse problem is closely related to the solution (10.1) of the expected value problem. The only difference is that we have to account for the least favorable distribution of the offers. If $P \in \mathscr{A}$ (which is the typical situation), then the critical level $x^* \leq \hat{x}$.

10.2. Organ transplantation. We illustrate our results on a risk-averse and simplified version of an organ transplantation problem discussed by Alagoz et al. [1]. We consider the discrete-time absorbing Markov chain depicted in Figure 1. State S, which is the initial state, represents a patient in need of an organ transplantation. State L represents life after a successful transplantation. State D (absorbing state)
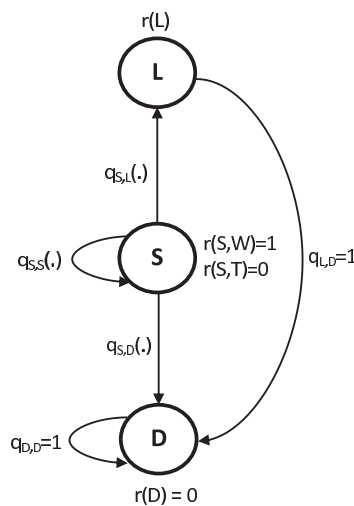

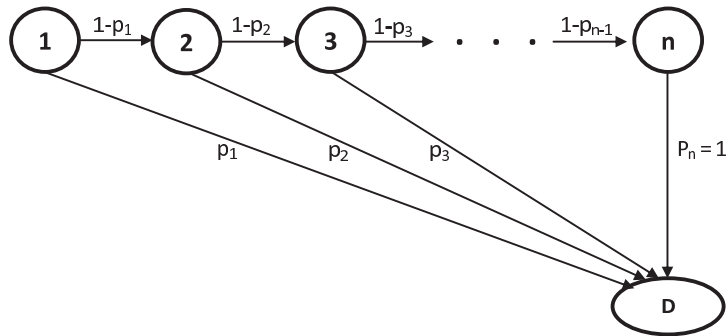FIG. 1. *The organ transplantation model.*

Fig. 2. *The survival model.*

represents death. Two control values are possible in state S: W (for "Wait"), in which case transition to state D or back to state S may occur, and T (for "Transplant"), which results in a transition to states L or D. The probability of death is lower for W than for T, but successful transplantation may result in a longer life, as explained below. In the other two states only one (formal) control value is possible: "Continue."

The rewards collected at each time step are months of life. In state S a reward equal to 1 is collected, if the control is W; otherwise, the immediate reward is 0. In state L the reward $r(L)$ is collected, representing the sure equivalent of the random length of life after transplantation. In state D the reward is 0.

Generally, in a cost minimization problem, the value of a dynamic measure of risk (3.1) is the "fair" sure charge one would be willing to incur, instead of a random sequence of costs. In our case, which will be a maximization problem, we shall work with the negatives of the months of life as our "costs." The value of the measure of risk, therefore, can be interpreted as the negative of a sure life length which we consider to be equivalent to the random life duration faced by the patient.

Let us start by describing the way the deterministic equivalent length of life $r(L)$ at state L is calculated. The state L is in fact an aggregation of n states in a survival model representing months of life after transplantation, as depicted in Figure 2.

In state $i = 1, \ldots, n$, the patient dies with probability p_i and survives with probability $1 - p_i$. The probability $p_n = 1$. The reward collected at each state $i = 1, \ldots, n$ is equal to 1. In order to follow the notation of our paper, we define the cost $c(\cdot) = -r(\cdot)$. For illustration, we apply the mean–semideviation model of Example 3.1 with $\kappa = 1$.

The transition risk mapping has the form

$$(10.5) \qquad \sigma(\varphi, i, \nu) = \underbrace{\mathbb{E}_\nu[\varphi]}_{\text{expected value}} + \kappa \underbrace{\mathbb{E}_\nu\left[\left(\varphi - \mathbb{E}_\nu[\varphi]\right)_+\right]}_{\text{semideviation}}.$$

Owing to the monotonicity property (A2), $\sigma(\varphi, i, \nu) \leq 0$, whenever $\varphi(\cdot) \leq 0$.

In (10.5), the measure ν is the transition kernel at the current state i, and the function $\varphi(\cdot)$ is the cost incurred at the current state and control plus the value function at the next state. At each state $i = 1, \ldots, n-1$ two transitions are possible: to D with probability p_i and $\varphi = -1$, and to $i + 1$ with probability $1 - p_i$ and $\varphi = -1 + v_{i+1}(i+1)$. At state $i = n$ the transition to D occurs with probability 1, and $\varphi = -1$. Therefore, $v_n(n) = -1$.

The survival problem is a finite horizon problem, and thus we apply (6.3). As there is no control to choose, the minimization operation in (6.3) is eliminated. The

TABLE 1
Transition probabilities from state S.

Control	S	L	D
W	0.99882	0	0.00118
T	0	0.90782	0.09218

equation has the form:

$$v_i(i) = \sigma(\varphi, i, Q_i), \quad i = 1, \ldots, n - 1,$$

with φ and Q_i as explained above. By induction, $v_i(i) \leq 0$, for $i = n - 1, n - 2, \ldots, 1$.

Let us calculate the mean and semideviation components of (10.5) at all states $i = 1, \ldots, n - 1$:

$$\mathbb{E}_{Q_i}[\varphi] = -p_i + (1 - p_i)\big(-1 + v_{i+1}(i+1)\big) = -1 + (1 - p_i)v_{i+1}(i+1),$$

$$\begin{aligned}
\mathbb{E}_{Q_i}\big[\big(\varphi - \mathbb{E}_{Q_i}[\varphi]\big)_+\big] &= \mathbb{E}_{Q_i}\big[\big(\varphi + 1 - (1 - p_i)v_{i+1}(i+1)\big)_+\big] \\
&= p_i\big(-1 + 1 - (1 - p_i)v_{i+1}(i+1)\big)_+ \\
&\quad + (1 - p_i)\big(-1 + v_{i+1}(i+1) + 1 - (1 - p_i)v_{i+1}(i+1)\big)_+ \\
&= p_i\big(-(1 - p_i)v_{i+1}(i+1)\big)_+ + (1 - p_i)\big(p_i v_{i+1}(i+1)\big)_+ \\
&= -p_i(1 - p_i)v_{i+1}(i+1).
\end{aligned}$$

In the last equation we used the fact that $v_{i+1}(i+1) \leq 0$. For $i = 1, \ldots, n - 1$, the dynamic programming equations (6.3) take on the form

$$v_i(i) = \underbrace{-1 + (1 - p_i)v_{i+1}(i+1)}_{\text{expected value}} - \kappa\,\underbrace{p_i(1 - p_i)v_{i+1}(i+1)}_{\text{semideviation}}, \quad i = n - 1, n - 2, \ldots, 1.$$

The value $v(1)$ is the negative of the risk-adjusted length of life with a new organ. For $\kappa = 0$ the above formulas give the negative of the expected length of life with a new organ.

In our calculations we used the transition data provided in Table 1. They have been chosen for purely illustrative purposes and do not correspond to any real medical situation.

For the survival model, we used the distribution function, $F(x)$, of lifetime of the American population from Jasiulewicz [20]. It is a mixture of Weibull, lognormal, and Gompertz distributions:

$$\begin{aligned}
F(x) = {} & w_1\left(1 - \exp\left(-\left(\frac{x}{\delta}\right)^\beta\right)\right) + w_2\Phi\left(\frac{\log x - m}{\sigma}\right) \\
& + w_3\left(1 - \exp\left(-\frac{b}{\alpha}(e^{\alpha x} - 1)\right)\right), \quad x \geq 0.
\end{aligned}$$

The values of the parameters and weights, provided by Jasiulewicz [20], are given in Table 2.

Then, we calculated the probability of dying at age k (in months) as follows:

$$p_k = \frac{F(k/12 + 1/24) - F(k/12 - 1/24)}{1 - F(k/12 - 1/24)}, \quad k = 1, 2, \ldots.$$

The maximum lifetime of the patient was taken to be 1200 months, and that the patient after transplantation has survival probabilities starting from $k = 300$. Therefore, $n = 900$ in the survival model used for calculating $r(\mathrm{L})$.

TABLE 2
Values of parameters for $F(x)$.

Distribution	Parameters	Weights
Weibull	$\delta = 0.297$, $\beta = 0.225$	$w_1 = 0.0170$
Lognormal	$m = 3.11$, $\sigma = 0.218$	$w_2 = 0.0092$
Gompertz	$b = 0.0000812$, $\alpha = 0.0844$	$w_3 = 0.9737$

Let $\lambda = (\lambda_{\mathrm{W}}, \lambda_{\mathrm{T}})$ be the randomized policy in the state S and let

$$\Lambda = \big\{ \lambda \in \mathbb{R}^2 : \lambda_{\mathrm{W}} + \lambda_{\mathrm{T}} = 1, \ \lambda \geq 0 \big\}.$$

The dynamic programming equation (8.4) at S takes on the form

$$v(\mathrm{S}) =$$
$$\min_{\lambda \in \Lambda} \Bigg\{ \underbrace{\lambda_{\mathrm{W}} \big[q_{\mathrm{S,S}}(\mathrm{W})\big(v(\mathrm{S}) - 1\big)\big) + q_{\mathrm{S,D}}(\mathrm{W})\big(v(\mathrm{D}) - 1\big) \big] + \lambda_{\mathrm{T}} \big[q_{\mathrm{S,L}}(\mathrm{T})v(\mathrm{L}) + q_{\mathrm{S,D}}(\mathrm{T})v(\mathrm{D}) \big]}_{\text{expected value } \mu}$$

$$+ \kappa \Big(\underbrace{\lambda_{\mathrm{W}} \big[q_{\mathrm{S,S}}(\mathrm{W})\big(v(\mathrm{S}) - 1 - \mu\big)_+ + q_{\mathrm{S,D}}(\mathrm{W})\big(v(\mathrm{D}) - 1 - \mu\big)_+ \big]}_{\text{semideviation} \dots}$$

$$+ \underbrace{\lambda_{\mathrm{T}} \big[q_{\mathrm{S,L}}(\mathrm{T})\big(v(\mathrm{L}) - \mu\big)_+ + q_{\mathrm{S,D}}(\mathrm{T})\big(v(\mathrm{D}) - \mu\big)_+ \big]}_{\dots \text{semideviation}} \Big) \Bigg\}.$$

In the semideviation parts, we wrote μ for the expectation of the value function in the next state, which is given by the first underbraced expression, and which is also dependent on λ. Of course, the above expression can be simplified, by using the fact that $v(\mathrm{L}) < v(\mathrm{S}) < v(\mathrm{D}) = 0$, but we prefer to leave it in the above form to illustrate the way it has been developed.

We compared two optimal control models for this problem. The first one was the expected value model ($\kappa = 0$), which corresponds to the expected reward $r(\mathrm{L}) = 610.46$ in the survival model. Standard dynamic programming equations were solved, and the optimal decision in state S turned out to be W.

The second model was the risk-averse model using the mean–semideviation transition risk mapping with $\kappa = 1$. This changed the reward at state L to 515.35. We considered two versions of this model. In the first version, we restricted the feasible policies to be deterministic. In this case, the optimal action in state S was T. In the second version, we allowed randomized policies, as in our general model. Then the optimal policy in state S was W with probability $\lambda_{\mathrm{W}} = 0.9873$ and T with probability $\lambda_{\mathrm{T}} = 0.0127$.

How can we interpret these results? The optimal randomized policy results in a random waiting time before transplanting the organ. This is due to the fact that immediate transplantation entails a significant probability of death, and a less risky policy is to "dilute" this probability in a long waiting time. This cannot be derived from an expected value model, no matter what the data, because deterministic policies are optimal in such a model: either transplant immediately or never.

This paradoxical conclusion follows from the assumption of a Markov structure of the model, and is hard to accept in a medical setting. Our aim here is to indicate intriguing properties of risk-averse models, rather than to recommend a particular course of action in real-life situations. Undoubtedly, more research is needed to fully understand all issues arising in this context.

11. Conclusion. In this paper, we used dynamic measures of risk to model a risk-averse version of the undiscounted transient Markov control process under the assumption of infinite state space and compact control sets. We showed that transient models may still have infinite risk. Using the concept of a multikernel, we introduced the class of risk-transient models. We derived dynamic programming equations for the case of deterministic feasible policies. Also, a preliminary discussion of the extension of this study to randomized policies was provided. We illustrated our theory with two examples: asset selling and organ transplantation. We showed that the solution of the risk-neutral asset selling problem has a counterpart in the risk-averse setting. In the organ transplantation example, we showed the usefulness of randomized policies.

REFERENCES

[1] O. ALAGOZ, L. M. MAILLART, A. J. SCHAEFER, AND M. S. ROBERTS, *The optimal timing of living-donor liver transplantation*, Management Sci., 50 (2004), pp. 1420–1430.

[2] P. ARTZNER, F. DELBAEN, J. M. EBER, AND D. HEATH, *Coherent measures of risk*, Math. Finance, 9 (1969), pp. 203–228.

[3] P. ARTZNER, F. DELBAEN, J.-M. EBER, D. HEATH, AND H. KU, *Coherent multiperiod risk adjusted values and Bellmans principle*, Ann. Oper. Res., 152 (2007), pp. 5–22.

[4] J.-P. AUBIN AND H. FRANKOWSKA, *Set-Valued Analysis*, Birkhäuser, Boston, 1990.

[5] N. BÄUERLE AND U. RIEDER, *Markov Decision Processes with Applications to Finance*, Springer, Berlin, 2011.

[6] D. BERTSEKAS AND S. E. SHREVE, *Stochastic Optimal Control. The Discrete Time Case*, Academic Press, New York, 1978.

[7] D. P. BERTSEKAS AND J. N. TSITSIKLIS, *An analysis of stochastic shortest-path problems*, Math. Oper. Res., 16 (1991), pp. 580–595.

[8] P. CHERIDITO, F. DELBAEN, AND M. KUPPER, *Dynamic monetary risk measures for bounded discrete-time processes*, Electron. J. Probab., 11 (2006), pp. 57–106.

[9] S. H. CHEW AND J. L. HO, *Hope: An empirical study of attitude toward the timing of uncertainty resolution*, J. Risk Uncertain., 8 (1994), pp. 267–288.

[10] Ö. ÇAVUŞ AND A. RUSZCZYŃSKI, *Computational methods for risk-averse undiscounted transient Markov models*, Oper. Res., 62 (2014), pp. 401–417.

[11] E. ÇINLAR, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Ciffs, NJ, 1975.

[12] E. V. DENARDO AND U. G. ROTHBLUM, *Optimal stopping, exponential utility, and linear programming*, Math. Program. 16 (1979), pp. 228–244.

[13] E. B. DYNKIN AND A.A. YUSHKEVICH, *Markov Processes: Theory and Problems*, Plenum, New York, 1969.

[14] E. B. DYNKIN AND A.A. YUSHKEVICH, *Controlled Markov Processes*, Springer-Verlag, New York, 1979.

[15] E. A. FEINBERG AND A. SHWARTZ, *Handbook of Markov Decision Processes: Methods and Applications*, Kluwer, Dordrecht, 2002.

[16] M. FRITELLI AND G. SCANDOLO, *Risk measures and capital requirements for processes*, Math. Finance, 16 (2006), pp. 589–612.

[17] J. I. GONZÁLEZ-TREJO, O. HERNÁNDEZ-LERMA, AND L. F. HOYOS-REYES, *Minimax control of discrete-time stochastic systems*, SIAM J. Control Optim., 41 (2002), pp. 1626–1659.

[18] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Discrete-Time Markov Control Processes. Basic Optimality Criteria*, Springer, New York, 1996.

[19] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.

[20] H. JASIULEWICZ, *Application of mixture models to approximation of age-at-death distribution*, Math. Econom., 19 (1997), pp. 237–241.

[21] A. JAŚKIEWICZ AND A. S. NOWAK, *Stochastic games with unbounded payoffs: Applications to robust control in economics*, Dynam. Games Appl., 1 (2011), pp. 253–279.

[22] S. KARLIN, *Stochastic models and optimal policies for selling an asset*, in Studies in Applied Probability and Management Science, K. J. Arrow, S. Karlin, and S. Scarf, eds., Stanford University Press, Palo Alto, CA, 1962, pp. 148–158.

[23] S. KLÖPPEL AND M. SCHWEIZER, *Dynamic indifference valuation via convex risk measures*, Math. Finance, 17 (2007), pp. 599–627.

[24] K. KURATOWSKI AND C. RYLL-NARDZEWSKI, *A general theorem on selectors*, Bull. Acad. Polonaise Sci., Ser. Math. Astron. Phys., 13 (1965), pp. 397–411.

[25] M. Kurt and J. P. Kharoufeh, *Monotone optimal replacement policies for a Markovian deteriorating system in a controllable environment*, Oper. Res. Lett., 38 (2010), pp. 273–279.

[26] K. Lin and S. I. Marcus, *Dynamic programming with non-convex risk-sensitive measures*, in IEEE American Control Conference, IEEE, 2013, pp. 6778–6783.

[27] Y. Nie and X. Wu, *Shortest path problem considering on-time arrival probability*, Transportation Res. B, 43 (2009), pp. 597–613.

[28] A. Nilim and L. El Ghaoui, *Robust control of Markov decision processes with uncertain transition matrices*, Oper. Res., 53 (2005), pp. 780–798.

[29] Y. Ohtsubo, *Minimizing risk models in stochastic shortest path problems*, Math. Methods Oper. Res., 57 (2003), pp. 79–88.

[30] W. Ogryczak and A. Ruszczyński, *From stochastic dominance to mean-risk models: Semideviations as risk measures*, European J. Oper. Res., 116 (1999), pp. 33–50.

[31] W. Ogryczak and A. Ruszczyński, *On consistency of stochastic dominance and mean-semideviation models*, Math. Program., 89 (2001), pp. 217–232.

[32] W. Ogryczak and A. Ruszczynski, *Dual stochastic dominance and related mean-risk models*, SIAM J. Optim., 13 (2002), pp. 60–78.

[33] S. D. Patek, *On terminating Markov decision processes with a risk averse objective function*, Automatica J. IFAC, 37 (2001), pp. 1379–1386.

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

[35] S. R. Pliska, *On the transient case for Markov decision chains with general state spaces*, in Dynamic Programming and Its Applications, M. L. Puterman, ed., Academic Press, New York, 1979, pp. 335–349.

[36] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, New York, 1994.

[37] F. Riedel, *Dynamic coherent risk measures*, Stochastic Process. Appl., 112 (2004), pp. 185–200.

[38] R. T. Rockafellar, *Conjugate Duality and Optimization*, Regional Conf. Ser. Appl. Math. 16, SIAM, Philadelphia, 1974.

[39] R. T. Rockafellar and S. P. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Banking Finance, 26 (2002), pp. 1443–1471.

[40] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[41] A. Ruszczyński, *Risk-averse dynamic programming for Markov decision processes*, Math. Program. Ser. B, 125 (2010), pp. 235–261.

[42] A. Ruszczyński and A. Shapiro, *Optimization of risk measures*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer, London, 2005.

[43] A. Ruszczyński and A. Shapiro, *Optimization of convex risk functions*, Math. Oper. Res., 31 (2006), pp. 433–452.

[44] A. Ruszczyński and A. Shapiro, *Conditional risk mappings*, Math. Oper. Res., 31 (2006), pp. 544–561.

[45] G. Scandolo, *Risk Measures in a Dynamic Setting*, Ph.D. thesis, Università degli Studi di Milano, Milan, 2003.

[46] M. Schäl, *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Probab. Theory Related Fields, 32 (1975), pp. 179–196.

[47] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, MPS/SIAM Ser. Optim. 9, SIAM, Philadelphia, 2009.

[48] Y. Shen, W. Stannat, and K. Obermayer, *Risk-sensitive Markov control processes*, SIAM J. Control Optim., 51 (2013), pp. 3652–3672.

[49] M. M. C. So and L. C. Thomas, *Modelling the profitability of credit cards by Markov decision processes*, European J. Oper. Res., 212 (2011), pp. 123–130.

[50] A. F. Veinott, *Discrete dynamic programming with sensitive discount optimality criteria*, Ann. Math. Statist., 40 (1969), pp. 1635–1660.

[51] J. Wessels, *Markov programming by successive approximations with respect to weighted supremum norms*, J. Math. Anal. Appl., 58 (1977), pp. 326–335.

[52] D. J. White, *A survey of applications of Markov decision processes*, J. Oper. Res. Soc., 44 (1993), pp. 1073–1096.