MATHEMATICS OF OPERATIONS RESEARCH
Vol. 31, No. 3, August 2006, pp. 544–561
ISSN 0364-765X | EISSN 1526-5471 | 06 | 3103 | 0544

DOI 10.1287/moor.1060.0204

Conditional Risk Mappings

Andrzej Ruszczyński
Department of Management Science and Information Systems, Rutgers University,
Piscataway, New Jersey 08854, USA, rusz@business.rutgers.edu, http://www.rusz.rutgers.edu

Alexander Shapiro
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332,
ashapiro@isye.gatech.edu, http://www2.isye.gatech.edu/people/faculty/Alex_Shapiro/

We introduce an axiomatic definition of a conditional convex risk mapping and we derive its properties. In particular, we prove a representation theorem for conditional risk mappings in terms of conditional expectations. We also develop dynamic programming relations for multistage optimization problems involving conditional risk mappings.

Key words: risk; conjugate duality; stochastic optimization; dynamic programming; multistage stochastic programming
MSC2000 subject classification: Primary: 90C15, 90C48; secondary: 91B30, 90C46
OR/MS subject classification: Primary: stochastic programming; secondary: risk
History: Received February 21, 2004; revised April 27, 2005, and January 31, 2006.

1. Introduction. Models of risk, and optimization problems involving these models, attracted considerable attention in recent years. One direction of research associated with an axiomatic approach was initiated by Kijima and Ohnishi [8]. The influential paper of Artzner et al. [1] introduced the concept of *coherent risk measures*. Subsequently, this approach was developed by Föllmer and Schied [7], Cheridito et al. [5], Rockafellar et al. [16], Ruszczyński and Shapiro [17], and others. In Ruszczyński and Shapiro [17] a general duality framework has been developed that allows us to view earlier representation theorems for risk measures as special cases of the theory of conjugate duality in paired topological vector spaces. In the discussion below we follow the general setting and terminology of Ruszczyński and Shapiro [17].

We assume that Ω is a measurable space equipped with a sigma algebra $\mathcal{F}$ of subsets of Ω, and that an uncertain outcome is represented by a measurable function $X\colon \Omega \to \mathbb{R}$. We assume that the smaller the values of X, the better (for example, X represents uncertain costs). Of course, our constructions can be easily adapted to the reverse situation.

If we introduce a space $\mathcal{X}$ of measurable functions on Ω, we can talk of a *risk function* as a mapping $\rho\colon \mathcal{X} \to \mathbb{R}$ (we can also consider risk functions with values in the extended real line). In our earlier work (Ruszczyński and Shapiro [17]) we have refined and extended the analysis of Artzner et al. [1], Delbaen [6], Cheridito et al. [5], Föllmer and Schied [7], and Rockafellar et al. [16], and we have derived, from a handful of axioms, fairly general properties of risk functions. Most importantly, we have analyzed optimization problems involving risk functions. In such a problem the uncertain outcome X results from our decision z modeled as an element of some vector space $\mathcal{Z}$. Formally $X = F(z)$, where $F\colon \mathcal{Z} \to \mathcal{X}$. The associated optimization problem takes on the form:

$$\min_{z\in S} \rho(F(z)), \tag{1.1}$$

where S is a convex subset of $\mathcal{Z}$. In Ruszczyński and Shapiro [17] we have derived optimality conditions and duality theory for problems of form (1.1).

Our objective now is to analyze models of risk in a dynamic setting. Suppose that our information, decisions, and costs are associated with stages $t = 1, \dots, T$. After each stage t, a sigma subalgebra $\mathcal{F}_t$ of $\mathcal{F}$ models the information available. We assume that these sigma subalgebras form a filtration: $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T$, with $\mathcal{F}_T = \mathcal{F}$.

The cost incurred at stage t is represented by a function $X_t \in \mathcal{X}_t$, where $\mathcal{X}_t$ is a space of measurable functions on $(\Omega, \mathcal{F}_t)$. The total cost is thus

$$X = X_1 + X_2 + \cdots + X_T.$$

One way to model risk in problems involving such random outcomes would be to apply a certain risk function $\rho(\cdot)$ to the entire sum X. However, this would ignore the dynamic character of the problem in question, and the sequential nature of the decision-making process. For these reasons we aim at developing *conditional risk mappings* that represent future risk from the point of view of the information available at the current stage.

Our approach, as well as that of Riedel [13], is different from the method of Artzner et al. [2]. In Artzner et al. [2], an adapted sequence $\{X_t\}$, $t = 1, \dots, T$, is viewed as a measurable function on a new measurable space $(\Omega', \mathcal{F}')$, with $\Omega' := \Omega \times \{1, \ldots, T\}$, and with the sigma algebra $\mathcal{F}'$ generated by sets of form $B_t \times \{t\}$, for all $B_t \in \mathcal{F}_t$ and $t = 1, \ldots, T$. They then use the properties of coherent (scalar) risk measures on this new space to develop risk models in the dynamic setting. Our intention is to develop models suitable for sequential decision making, and eventually to extend dynamic programming equations to risk-averse problems.

The main issue here is our knowledge at the time when risk is evaluated. In the classical setting of multistage stochastic optimization, the main tool used to formulate the corresponding dynamic programming equations is the concept of *conditional expectation*. Given two sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2$ of subsets of Ω, with $\mathcal{F}_1$ representing our knowledge when the expectation is evaluated, and $\mathcal{F}_2$ representing all events under consideration, the conditional expectation can be defined as a *mapping* from a space of $\mathcal{F}_2$-measurable functions into a space of $\mathcal{F}_1$-measurable functions. Of course, the conditional expectation mapping is *linear*. The basic idea of our approach is to extend the concept of conditional expectation to an appropriate class of *convex* mappings.

Together with the sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2$, we consider two linear (vector) spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ of functions measurable with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$. A *conditional risk mapping* is defined in §2 as a convex, monotone, and translation equivariant mapping $\rho_{\mathcal{X}_2|\mathcal{X}_1}\colon \mathcal{X}_2 \to \mathcal{X}_1$. In §3 we extend our insights from Ruszczyński and Shapiro [17] to derive a duality representation theorem for conditional risk mappings. Section 4 is devoted to the analysis of relations of conditional risk mappings and conditional expectations. In §5 we consider a sequence of sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T$ and the corresponding linear spaces $\mathcal{X}_t$, $t = 1, \ldots, T$, of measurable functions, and we analyze compositions of risk mappings of the form $\rho_{\mathcal{X}_2|\mathcal{X}_1} \circ \cdots \circ \rho_{\mathcal{X}_{t-1}|\mathcal{X}_{t-2}} \circ \rho_{\mathcal{X}_T|\mathcal{X}_{T-1}}$. Two practically important examples of conditional risk mappings are thoroughly analyzed in §6. Finally, §7 addresses the issue of risk measures for sequences, and develops dynamic programming equations for associated optimization problems.

2. Axioms of conditional risk mappings. In order to construct dynamic models of risk we need to extend the concept of risk functions. We proceed as follows. Let $\mathcal{F}_1 \subset \mathcal{F}_2$ be sigma algebras of subsets of a set Ω, and $\mathcal{X}_1 \subset \mathcal{X}_2$ be linear spaces of real-valued functions $\phi(\omega)$, $\omega \in \Omega$, measurable with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively.

Definition 2.1. We say that a mapping $\rho\colon \mathcal{X}_2 \to \mathcal{X}_1$ is a *conditional risk mapping* if the following properties hold:

(A1) Convexity. *If* $\alpha \in [0, 1]$ *and* $X, Y \in \mathcal{X}_2$, *then*

$$\alpha\rho(X) + (1-\alpha)\rho(Y) \succeq \rho(\alpha X + (1-\alpha)Y);$$

(A2) Monotonicity. *If* $Y \succeq X$, *then* $\rho(Y) \succeq \rho(X)$;

(A3) Predictable Translation Equivariance. *If* $Y \in \mathcal{X}_1$ *and* $X \in \mathcal{X}_2$, *then*

$$\rho(X + Y) = \rho(X) + Y.$$

The inequalities in (A1) and (A2) are understood componentwise, i.e., $Y \succeq X$ means that $Y(\omega) \geq X(\omega)$ for every $\omega \in \Omega$. The above definition depends on the choice of the spaces $\mathcal{X}_1$ and $\mathcal{X}_2$. To emphasize this, we sometimes write $\rho_{\mathcal{X}_2|\mathcal{X}_1}$ for the conditional risk mapping. An example of a conditional risk mapping is the conditional expectation $\rho(X) := \mathbb{E}[X \mid \mathcal{F}_1]$, provided that $\mathbb{E}[X \mid \mathcal{F}_1]$ is an element of the space $\mathcal{X}_1$ for every $X \in \mathcal{X}_2$. We show in §4 that, in general, the concept of conditional risk mappings is closely related to the notion of conditional expectation. This motivates the use of the adjective *conditional* in the name of these mappings.

Assumptions (A1)–(A3) generalize the conditions introduced in Riedel [13] for dynamic risk measures in the case of a finite space Ω. We postulate convexity rather than positive homogeneity, and we allow for a general measurable space Ω.

For each $\omega \in \Omega$, we associate with ρ the function

$$\rho_\omega(X) := [\rho(X)](\omega), \quad X \in \mathcal{X}_2. \tag{2.1}$$

Assumptions (A1) and (A2) mean that, for every $\omega \in \Omega$, the function $\rho_\omega\colon \mathcal{X}_2 \to \mathbb{R}$ is convex and monotone, respectively. Moreover, Assumption (A3) implies that $\rho_\omega(X + a) = \rho_\omega(X) + a$ for every $X \in \mathcal{X}_2$ and every $a \in \mathbb{R}$, provided that the space $\mathcal{X}_1$ includes the constant functions (see the following Assumption (C′)). That is, $\rho_\omega(\cdot)$ satisfies the axioms of convex risk functions, as given in Föllmer and Schied [7] and analyzed in our earlier paper, Ruszczyński and Shapiro [17]. In particular, if the sigma algebra $\mathcal{F}_1$ is trivial, i.e., $\mathcal{F}_1 = \{\varnothing, \Omega\}$, then any function $X \in \mathcal{X}_1$ is constant over Ω, and hence the space $\mathcal{X}_1$ can be identified with $\mathbb{R}$. In that case, $\rho(\cdot)$ becomes real valued, and Assumptions (A1)–(A3) become the axioms of convex (real-valued) risk functions.

We assume that with each space $\mathscr{X}_i$, $i = 1, 2$, is associated a linear space $\mathscr{Y}_i$ of signed finite measures on $(\Omega, \mathscr{F}_i)$ such that $\mathscr{Y}_1 \subset \mathscr{Y}_2$, and[1] $\int_\Omega |X|\, d|\mu| < +\infty$ for every $X \in \mathscr{X}_i$ and $\mu \in \mathscr{Y}_i$. Then we can define the scalar product (bilinear form)

$$\langle \mu, X \rangle := \int_\Omega X(\omega)\, d\mu(\omega), \quad X \in \mathscr{X}_i, \ \ \mu \in \mathscr{Y}_i. \tag{2.2}$$

By $\mathscr{P}_{\mathscr{Y}_i}$ we denote the set of probability measures $\mu \in \mathscr{Y}_i$, i.e., $\mu \in \mathscr{P}_{\mathscr{Y}_i}$ if μ is nonnegative and $\mu(\Omega) = 1$. We assume that $\mathscr{X}_i$ and $\mathscr{Y}_i$ are *paired, locally convex* topological vector spaces. That is, $\mathscr{X}_i$ and $\mathscr{Y}_i$ are endowed with respective topologies that make them locally convex topological vector spaces. Moreover, these topologies are compatible with the scalar product (2.2), i.e., every continuous linear functional on $\mathscr{X}_i$ can be represented in the form $\langle \mu, \cdot \rangle$ for some $\mu \in \mathscr{Y}_i$, and every continuous linear functional on $\mathscr{Y}_i$ can be represented in the form $\langle \cdot, X \rangle$ for some $X \in \mathscr{X}_i$. In particular, we can endow each space $\mathscr{X}_i$ and $\mathscr{Y}_i$ with its weak topology induced by its paired space. This will make $\mathscr{X}_i$ and $\mathscr{Y}_i$ paired, locally convex topological vector spaces provided that for any $X \in \mathscr{X}_i \backslash \{0\}$ there exists $\mu \in \mathscr{Y}_i$ such that $\langle \mu, X \rangle \neq 0$, and for any $\mu \in \mathscr{Y}_i \backslash \{0\}$ there exists $X \in \mathscr{X}_i$ such that $\langle \mu, X \rangle \neq 0$.

A natural choice of $\mathscr{X}_i, i = 1, 2$, is the space of all bounded $\mathscr{F}_i$-measurable functions $X\colon \Omega \to \mathbb{R}$. In that case, we can take $\mathscr{Y}_i$ to be the space of all signed finite measures on $(\Omega, \mathscr{F}_i)$. Another possible choice is $\mathscr{X}_i := \mathscr{L}_p(\Omega, \mathscr{F}_i, P)$ for some positive (probability) measure P on $(\Omega, \mathscr{F}_2)$ and $p \in [1, +\infty]$. Note that because $\mathscr{F}_1 \subset \mathscr{F}_2$, P is also a positive measure on $(\Omega, \mathscr{F}_1)$, and hence $\mathscr{X}_1 \subset \mathscr{X}_2$. We can then take $\mathscr{Y}_i$ to be the linear space of measures ν, which are absolutely continuous with respect to P and whose density (Radon-Nikodym derivative) $h = d\nu/dP$ belongs to the space $\mathscr{L}_q(\Omega, \mathscr{F}_i, P)$, where $q \geq 1$ is such that $1/p + 1/q = 1$. In that case, we identify $\mathscr{Y}_i$ with $\mathscr{L}_q(\Omega, \mathscr{F}_i, P)$, and define the scalar product

$$\langle h, X \rangle := \int_\Omega X(\omega) h(\omega)\, dP(\omega), \quad X \in \mathscr{L}_p(\Omega, \mathscr{F}_i, P), \ \ h \in \mathscr{L}_q(\Omega, \mathscr{F}_i, P). \tag{2.3}$$

Note that an element $X \in \mathscr{L}_p(\Omega, \mathscr{F}_i, P)$ (an element $h \in \mathscr{L}_q(\Omega, \mathscr{F}_i, P)$) is a *class* of functions that are equal to each other for almost every (a.e.) $\omega \in \Omega$ with respect to the measure P. The space $\mathscr{X}_i := \mathscr{L}_p(\Omega, \mathscr{F}_i, P)$ is a Banach space and, for $p \in [1, +\infty)$, $\mathscr{Y}_i := \mathscr{L}_q(\Omega, \mathscr{F}_i, P)$ is its dual space of all linear-continuous functionals on $\mathscr{X}_i$. When dealing with Banach spaces we endow $\mathscr{X}_i$ and $\mathscr{Y}_i := \mathscr{X}_i^*$ with the strong (norm) and weak* topologies, respectively. If $\mathscr{X}_i$ is a reflexive Banach space, i.e., $\mathscr{X}_i^{**} = \mathscr{X}_i$, then $\mathscr{X}_i$ and $\mathscr{X}_i^*$, both endowed with strong topologies, form paired spaces.

We assume throughout the paper that the space $\mathscr{X}_2$ is sufficiently large so that the following assumption holds (recall that for $X \in \mathscr{X}_2$, the notation $X \succeq 0$ means that $X(\omega) \geq 0$ for all $\omega \in \Omega$).

Assumption (C). *If $\mu \in \mathscr{Y}_2$ is not nonnegative, then there exists $X \in \mathscr{X}_2$ such that $X \succeq 0$ and $\langle \mu, X \rangle < 0$.*

The above condition ensures that the cone of nonnegative-valued functions in $\mathscr{X}_2$ and the cone of nonnegative measures in $\mathscr{Y}_2$ are dual to each other. It is a mild technical requirement on the pairing of $\mathscr{X}_2$ and $\mathscr{Y}_2$. We are using it in the key Theorem 3.1.

A measure μ is not nonnegative if $\mu(A) < 0$ for some $\mathscr{A} \in \mathscr{F}_2$. Therefore, Assumption (C) holds, for example, if the space $\mathscr{X}_2$ contains all functions $\mathbb{1}_A(\cdot)$, $A \in \mathscr{F}_2$, where $\mathbb{1}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ for $\omega \notin A$. For technical reasons we assume that this property also holds for the space $\mathscr{X}_1$:

Assumption (C′). *For every $B \in \mathscr{F}_1$, the function $\mathbb{1}_B$ belongs to the space $\mathscr{X}_1$.*

We say that Y is an $\mathscr{F}_1$-*step function* if it can be represented in the form $Y = \sum_{k=1}^K \alpha_k \mathbb{1}_{B_k}$, where $B_k \in \mathscr{F}_1$ and $B_k \cap B_l = \varnothing$, if $k \neq l$. Clearly, if $\alpha_k \geq 0$, then the step function Y is nonnegative. By Assumption (C′), the space $\mathscr{X}_1$ contains all $\mathscr{F}_1$-step functions, and in particular all constant functions.

It is said that the conditional risk mapping ρ is *positively homogeneous* if

$$\rho(\alpha X) = \alpha \rho(X), \quad \text{for all } X \in \mathscr{X}_2 \text{ and } \alpha > 0. \tag{2.4}$$

In that case, $\rho(0) = 0$, and for any $Y \in \mathscr{X}_1$ we have

$$\rho(Y) = \rho(0 + Y) = \rho(0) + Y = Y,$$

and hence $\rho[\rho(X)] = \rho(X)$ for any $X \in \mathscr{X}_2$.

[1] For a signed measure μ we denote by $|\mu|$ the corresponding total variation measure, i.e., $|\mu| = \mu^+ + \mu^-$ where $\mu = \mu^+ - \mu^-$ is the Jordan decomposition of μ.

3. Conjugate duality of conditional risk mappings. We say that mapping $\rho\colon \mathscr{X}_2 \to \mathscr{X}_1$ is *lower semicontinuous* if, for every $\omega \in \Omega$, the corresponding function $\rho_\omega\colon \mathscr{X}_2 \to \mathbb{R}$ is lower semicontinuous (in the considered topology of $\mathscr{X}_2$). Note that the function $\rho_\omega(\cdot)$ is real valued here, and hence if $\mathscr{X}_2$ is a Banach space, equipped with its strong (norm) topology, and $\rho_\omega(\cdot)$ is convex and lower semicontinuous, then actually $\rho_\omega(\cdot)$ is continuous. With ρ_ω is associated its conjugate function

$$\rho^*_\omega(\mu) := \sup_{X\in\mathscr{X}_2} \{\langle \mu, X\rangle - \rho_\omega(X)\}. \tag{3.1}$$

Note that the conjugate function $\rho^*_\omega(\cdot)$ can take the value $+\infty$. Recall that the extended real-valued function $\rho^*_\omega(\cdot)$ is said to be *proper* if $\rho^*_\omega(\mu) > -\infty$ for any $\mu \in \mathscr{Y}_2$, and its domain

$$\operatorname{dom}(\rho^*_\omega) := \{\mu \in \mathscr{Y}_2\colon \rho^*_\omega(\mu) < +\infty\}$$

is nonempty. Because it is assumed that $\rho_\omega(\cdot)$ is finite valued, we always have $\rho^*_\omega(\cdot) > -\infty$. We also use the notation $\rho^*(\mu, \omega)$ for the function $\rho^*_\omega(\mu)$ in order to emphasize that it is a function of two variables, i.e., $\rho^*\colon \mathscr{Y}_2 \times \Omega \to \overline{\mathbb{R}}$. It has the following properties: For every $\omega \in \Omega$ the function $\rho^*(\cdot, \omega)$ is convex and lower semicontinuous, and for every $\mu \in \mathscr{Y}_2$ the function $\rho^*(\mu, \cdot)$ is $\mathscr{F}_1$-measurable.

We denote by $\mathscr{P}_{\mathscr{Y}_2}$ the set of all probability measures on $(\Omega, \mathscr{F}_2)$ that are in $\mathscr{Y}_2$. In particular, if $\mathscr{Y}_2 := \mathscr{L}_q(\Omega, \mathscr{F}_2, P)$, then (identifying measures with their densities)

$$\mathscr{P}_{\mathscr{Y}_2} = \left\{ h \in \mathscr{L}_q(\Omega, \mathscr{F}_2, P)\colon \int_\Omega h(\omega)\, dP(\omega) = 1 \text{ and } h(\omega) \ge 0 \text{ for a.e. } \omega \in \Omega \right\}.$$

With each $\omega \in \Omega$ we associate a set of probability measures $\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega) \subset \mathscr{P}_{\mathscr{Y}_2}$, defined as the set of all $\nu \in \mathscr{P}_{\mathscr{Y}_2}$ such that for every $B \in \mathscr{F}_1$

$$\nu(B) = \begin{cases} 1, & \text{if } \omega \in B, \\ 0, & \text{if } \omega \notin B. \end{cases} \tag{3.2}$$

Note that ω is fixed here and B varies in $\mathscr{F}_1$. Condition (3.2) means that for every ω and every $B \in \mathscr{F}_1$, we know whether B happened or not, and can be written equivalently as $\nu(B) = \mathbb{1}_B(\omega)$, $\forall\, \omega \in \Omega$. In particular, if $\mathscr{F}_1 = \{\varnothing, \Omega\}$, then $\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega) = \mathscr{P}_{\mathscr{Y}_2}$ for all $\omega \in \Omega$.

We can now formulate the basic duality result for conditional risk mappings.

Theorem 3.1. *Let $\rho = \rho_{\mathscr{X}_2|\mathscr{X}_1}$ be a lower-semicontinuous conditional risk mapping satisfying Assumptions* (A1)–(A3). *Then*

$$\rho_\omega(X) = \sup_{\mu\in\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)} \{\langle \mu, X\rangle - \rho^*(\mu, \omega)\}, \quad \omega \in \Omega, \quad X \in \mathscr{X}_2, \tag{3.3}$$

where $\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$ is the set of probability measures defined in (3.2), *and $\rho^*(\mu, \omega)$ is defined in* (3.1). *Conversely, suppose that a mapping $\rho\colon \mathscr{X}_2 \to \mathscr{X}_1$ can be represented in form* (3.3) *for some* (*proper*) *function $\rho^*\colon \mathscr{Y}_2 \times \Omega \to \overline{\mathbb{R}}$. Then, ρ is lower semicontinuous and satisfies Assumptions* (A1)–(A3).

Proof. If Assumptions (A1)–(A3) hold true, then ρ_ω is a convex risk function. As ρ_ω is lower semicontinuous, it follows from the Fenchel-Moreau theorem that

$$\rho_\omega(X) = \sup_{\mu\in\mathscr{P}_{\mathscr{Y}_2}} \{\langle \mu, X\rangle - \rho^*_\omega(\mu)\}, \quad X \in \mathscr{X}_2. \tag{3.4}$$

Conversely, if ρ_ω can be represented in form (3.3) for some ρ^*_ω, then ρ is lower semicontinuous and satisfies Assumptions (A1)–(A2). All these facts can be established by applying *verbatim* the proof of Theorem 2 in Ruszczyński and Shapiro [17] to the function ρ_ω. Therefore, the only issue that needs to be clarified is the restriction of $\operatorname{dom}(\rho^*_\omega)$ to $\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$.

Let $\omega \in \Omega$ be fixed and let $\mu_\omega \in \operatorname{dom}(\rho^*_\omega)$, and hence $\rho^*_\omega(\mu_\omega)$ is finite. It follows from (A3) that for every $Y \in \mathscr{X}_1$ we have

$$\begin{aligned}\rho^*_\omega(\mu_\omega) &= \sup_{X\in\mathscr{X}_2} \{\langle \mu_\omega, X+Y\rangle - \rho_\omega(X+Y)\} \\ &= \sup_{X\in\mathscr{X}_2} \{\langle \mu_\omega, X\rangle + \langle \mu_\omega, Y\rangle - \rho_\omega(X) - Y(\omega)\} \\ &= \rho^*_\omega(\mu_\omega) + \langle \mu_\omega, Y\rangle - Y(\omega).\end{aligned}$$

Therefore, $\langle \mu_\omega, Y\rangle = Y(\omega)$ for all $Y \in \mathscr{X}_1$. Setting $Y := \mathbb{1}_B$, where $B \in \mathscr{F}_1$, and noting that $\langle \mu_\omega, Y\rangle = \mathbb{E}_{\mu_\omega}[\mathbb{1}_B] = \mu_\omega(B)$, we conclude that $\mu_\omega(B) = \mathbb{1}_B(\omega)$ for all $\omega \in \Omega$. It follows that $\mu_\omega \in \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$, and hence $\operatorname{dom} \rho^*_\omega \subseteq \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$.

To prove the converse, we need only to verify Assumption (A3). Suppose that (3.3) holds true. Then every $\mu_\omega \in \operatorname{dom}(\rho^*_\omega)$ is an element of $\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$. Let $Y := \sum_{k=1}^K \alpha_k \mathbb{1}_{B_k}$, be an $\mathscr{F}_1$-step function. By Assumption (C′), we have $Y \in \mathscr{X}_1$. Then,

$$\langle \mu_\omega, Y\rangle = \sum_{k=1}^K \alpha_k \mu_\omega(B_k) = Y(\omega).$$

Passing to the limit, we obtain $\langle \mu_\omega, Y\rangle = Y(\omega)$ for every $\mathscr{F}_1$-measurable Y. Therefore, (3.3) implies that for every $Y \in \mathscr{X}_1$ and all $\omega \in \Omega$, we have

$$\begin{aligned}[\rho(X+Y)](\omega) &= \sup_{\mu \in \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)} \{\langle \mu, X+Y\rangle - \rho^*(\mu,\omega)\} \\ &= \sup_{\mu \in \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)} \{\langle \mu, X\rangle - \rho^*(\mu,\omega)\} + Y(\omega).\end{aligned}$$

This is identical to (A3). □

Let us provide a sufficient condition for the lower-semicontinuity assumption. Recall that the space $\mathscr{X}_2$ is said to be a *lattice* if for any $X_1, X_2 \in \mathscr{X}_2$ the element $X_1 \vee X_2$, defined as

$$[X_1 \vee X_2](\omega) := \max\{X_1(\omega), X_2(\omega)\}, \quad \omega \in \Omega,$$

belongs to $\mathscr{X}_2$. For every $X \in \mathscr{X}_2$ we can then define $|X| \in \mathscr{X}_2$ in a natural way, i.e., $|X|(\omega) = |X(\omega)|$, $\omega \in \Omega$. The space $\mathscr{X}_2$ is a *Banach lattice* if it is a Banach space and $|X_1| \le |X_2|$ implies $\|X_1\| \le \|X_2\|$. For example, every space $\mathscr{X}_2 := \mathscr{L}_p(\Omega, \mathscr{F}_2, P)$, $p \in [1, +\infty]$, is a Banach lattice. We can remark here that the lower semicontinuity of ρ follows from Assumptions (A1)–(A2) if $\mathscr{X}_2$ has the structure of a Banach lattice. Note that $[\rho(X)](\cdot)$ is a finite-valued function, and hence $\rho_\omega(X)$ is finite for all $\omega \in \Omega$ and all $X \in \mathscr{X}_2$. Direct application of Ruszczyński and Shapiro [17, Proposition 1] yields the following result.

Proposition 3.1. *Suppose that $\mathscr{X}_2$ is a Banach lattice and ρ: $\mathscr{X}_2 \to \mathscr{X}_1$ satisfies Assumptions* (A1) *and* (A2)*. Then $\rho_\omega(\cdot)$ is continuous for all $\omega \in \Omega$.*

Clearly, if ρ: $\mathscr{X}_2 \to \mathscr{X}_1$ is positively homogeneous, then the corresponding function ρ_ω is also positively homogeneous. Therefore, if $\rho = \rho_{\mathscr{X}_2|\mathscr{X}_1}$ is a positively homogeneous, lower-semicontinuous, conditional risk mapping, then $\rho^*(\cdot, \omega)$ is the indicator function of a closed convex set $\mathscr{A}(\omega) \subset \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$, and hence

$$\rho_\omega(X) = \sup_{\mu \in \mathscr{A}(\omega)} \langle \mu, X\rangle, \quad \omega \in \Omega, \ \ X \in \mathscr{X}_2. \tag{3.5}$$

We view $\omega \mapsto \mathscr{A}(\omega)$ as a multifunction from Ω into the set $\mathscr{P}_{\mathscr{Y}_2}$ of probability measures, on $(\Omega, \mathscr{F}_2)$, which are included in $\mathscr{Y}_2$. Formula (3.5) was first derived in Riedel [13, Theorem 1] for finite spaces Ω. Our results extend it to arbitrary measurable spaces.

The property of positive homogeneity can be strengthened substantially.

Proposition 3.2. *Let $\rho = \rho_{\mathscr{X}_2|\mathscr{X}_1}$ be a positively homogeneous, lower-semicontinuous conditional risk mapping. Suppose that for all $B \in \mathscr{F}_1$ and $X \in \mathscr{X}_2$, it holds that $\mathbb{1}_B X \in \mathscr{X}_2$. Then for every nonnegative $\mathscr{F}_1$-step function Y and every $X \in \mathscr{X}_2$, we have that $YX \in \mathscr{X}_2$ and*

$$\rho(YX) = Y\rho(X). \tag{3.6}$$

Proof. Consider a set $B \in \mathscr{F}_1$ and any $X \in \mathscr{X}_2$. It follows from (3.5) that

$$\rho_\omega(X) = \sup_{\mu \in \mathscr{A}(\omega)} \{\langle \mu, \mathbb{1}_B X\rangle + \langle \mu, \mathbb{1}_{\Omega\setminus B} X\rangle\}, \quad \omega \in \Omega. \tag{3.7}$$

If $\omega \in B$, then (3.2) implies that $\mu(\Omega \setminus B) = 0$ for all $\mu \in \mathscr{A}(\omega)$, and the second term at the right-hand side of (3.7) vanishes. Hence,

$$\rho_\omega(X) = \sup_{\mu \in \mathscr{A}(\omega)} \langle \mu, \mathbb{1}_B X\rangle = \rho_\omega(\mathbb{1}_B X). \tag{3.8}$$

By a similar argument, $\rho_\omega(\mathbb{1}_B X) = 0$ for all $\omega \notin B$. Thus,

$$\rho(\mathbb{1}_B X) = \mathbb{1}_B \rho(X). \tag{3.9}$$

Consider now a nonnegative $\mathscr{F}_1$-step function $Y := \sum_{k=1}^K \alpha_k \mathbb{1}_{B_k}$. Then $YX = \sum_{k=1}^K \alpha_k \mathbb{1}_{B_k} X \in \mathscr{X}_2$. It follows from (3.9) that for $\omega \in B_k$ the following chain of equations holds true:

$$\rho_\omega(YX) = \mathbb{1}_{B_k}(\omega)\rho_\omega(YX) = \rho_\omega(\mathbb{1}_{B_k} YX) = \rho_\omega(\alpha_k \mathbb{1}_{B_k} X).$$

Using positive homogeneity and (3.9) again, we obtain

$$\rho_\omega(YX) = \alpha_k \rho_\omega(\mathbb{1}_{B_k} X) = \alpha_k \rho_\omega(X), \quad \text{if } \omega \in B_k.$$

This means that

$$\rho_\omega(YX) = \sum_{k=1}^K \alpha_k \mathbb{1}_{B_k}(\omega)\rho_\omega(X) = Y(\omega)\rho_\omega(X), \quad \omega \in \Omega.$$

This completes the proof. □

Remark 3.1. In order to pass in (3.6) from step functions to general functions $Y \in \mathscr{X}_1$, we need some additional assumptions about the spaces involved. For example, consider $\mathscr{X}_i := \mathscr{L}_p(\Omega, \mathscr{F}_i, P)$, $i = 1, 2$, where P is a probability measure on $(\Omega, \mathscr{F}_2)$ and $p \in [1, +\infty)$. Then for every $\mathscr{F}_1$-step function Y and every $X \in \mathscr{X}_2$, we have $YX \in \mathscr{X}_2$. Moreover, the set of $\mathscr{F}_1$-step functions is dense in $\mathscr{X}_1$. Hence, by "passing to the limit operation" and using Proposition 3.1, we conclude that (3.6) is valid for all $Y \in \mathscr{X}_1$ and $X \in \mathscr{X}_2$, provided that $YX \in \mathscr{X}_2$.

4. Conditional expectation representation. In this section we discuss relations between conditional risk mappings and conditional expectations. In Theorem 3.1 we have established the representation (3.3) of a conditional risk mapping. Our objective now is to analyze in more detail the set of probability measures $\mathscr{A}(\omega) := \operatorname{dom}(\rho_\omega^*)$. Recall that $\mathscr{A}(\omega) \subset \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$, and that representation (3.3) reduces to (3.5) if the risk mapping is positively homogeneous. Definition (3.2) of the set $\mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$ means that its every element μ_ω is a certain probability measure on $\mathscr{F}_2$, which assigns value one or zero to sets in $\mathscr{F}_1$, depending whether ω is an element of the set or not. It is thus reasonable to ask if it is possible to represent these measures as conditional probability measures.

In order to gain an insight into this question, let us view $\mathscr{A}$ as a multifunction from Ω to $\mathscr{P}_{\mathscr{Y}_2}$. For the sake of illustration, we temporarily suppose that Ω is finite, say $\Omega = \{\omega_1, \ldots, \omega_N\}$, and that $\mathscr{F}_2$ contains all subsets of Ω. Consider a selection $\mu_\omega \in \mathscr{A}(\omega)$, i.e., $\mu_{\omega_i} \in \mathscr{A}(\omega_i)$, $i = 1, \ldots, N$. Because $\mathscr{A}(\omega) \subset \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$, we have, of course, that $\mu_\omega \in \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$. By the definition of the conditional risk mapping, the function $\omega \mapsto \rho_\omega(X)$ is $\mathscr{F}_1$-measurable. Therefore, the selection μ_ω is $\mathscr{F}_1$-measurable as well. The selection μ_ω is a conditional probability measure, of some measure ν_2 on $(\Omega, \mathscr{F}_2)$ with respect to the sigma algebra $\mathscr{F}_1$, if and only if the complete probability formula is valid:

$$\nu_2(B \cap S) = \sum_{\omega \in S} \mu_\omega(B)\nu_2(\omega) \quad \text{for all } S \in \mathscr{F}_1,\ B \in \mathscr{F}_2 \tag{4.1}$$

(see, e.g., Billingsley [3, p. 430]). By noting that the left-hand side of (4.1) is equal to $\sum_{\omega \in B\cap S} \nu_2(\omega)$, we can view (4.1) as a system of linear equations in unknowns $\nu_2(\omega_i)$, $i = 1, \ldots, N$.

The question is whether system (4.1) has a solution $\nu_2(\omega)$, which is a probability measure. The answer is rather simple. Consider a set $S \in \mathscr{F}_1$. We have that $\mu_\omega \in \mathscr{P}_{\mathscr{Y}_2|\mathscr{F}_1}(\omega)$, and hence it follows by (3.2) that if $\omega \notin S$, then $\mu_\omega(S) = 0$, and consequently $\mu_\omega(\tilde\omega) = 0$ for any $\tilde\omega \in S$. Let $S_1, \ldots, S_K \in \mathscr{F}_1$ be sets generating $\mathscr{F}_1$, i.e., $S_i \cap S_j = \varnothing$ for $i \neq j$, $\bigcup_{i=1}^K S_i = \Omega$, and if S is an $\mathscr{F}_1$-measurable subset of S_i, $i = 1, \ldots, N$, then either $S = S_i$ or $S = \varnothing$. For $\omega \in S_i$, $i = 1, \ldots, N$, the value of $\mu_\omega(\tilde\omega)$ is constant, because μ_ω is $\mathscr{F}_1$-measurable and the set S_i is indivisible. Let us denote this value by $\mu_{S_i}(\tilde\omega)$. Note that

$$\sum_{\tilde\omega \in S_i} \mu_{S_i}(\tilde\omega) = \mu_\omega(S_i) = 1, \quad i = 1, \ldots, N,$$

where the last equality holds by (3.2). Now let us take an arbitrary probability measure ν_1 on $(\Omega, \mathscr{F}_1)$ and define

$$\nu_2(\tilde\omega) := \nu_1(S_i)\mu_{S_i}(\tilde\omega), \quad \tilde\omega \in S_i,\ \ i = 1, \ldots, K. \tag{4.2}$$

Of course, the condition "$\tilde{\omega} \in S_i$" means that S_i is the smallest $\mathcal{F}_1$-measurable set containing $\tilde{\omega}$. Clearly, $\nu_2(\tilde{\omega}) \geq 0$ and

$$\sum_{\tilde{\omega}\in\Omega} \nu_2(\tilde{\omega}) = \sum_{i=1}^{K} \nu_1(S_i) \sum_{\tilde{\omega}\in S_i} \mu_{S_i}(\tilde{\omega}) = \sum_{i=1}^{K} \nu_1(S_i) = 1,$$

and hence ν_2 is a probability measure.

Let us verify that Equation (4.1) holds here. We have that for $\omega \in S_j$, it follows by (3.2) that $\mu_\omega(B) = \mu_\omega(B \cap S_j)$, and hence

$$\sum_{\omega\in S} \mu_\omega(B)\nu_2(\omega) = \sum_{i=1}^{K} \nu_1(S_i) \sum_{\omega\in S\cap S_i} \mu_{S_i}(\omega) \sum_{\tilde{\omega}\in B\cap S_i} \mu_{S_i}(\tilde{\omega}).$$

Now $\sum_{\omega\in S\cap S_i} \mu_{S_i}(\omega)$ is equal to one if $S_i \subset S$, and is zero otherwise. It follows that

$$\sum_{\omega\in S} \mu_\omega(B)\nu_2(\omega) = \sum_{i:\, S_i\subset S} \nu_1(S_i) \sum_{\tilde{\omega}\in B\cap S_i} \mu_{S_i}(\tilde{\omega}) = \sum_{\omega\in S\cap B} \nu_2(\omega).$$

We showed that for every selection $\mu_\omega \in \mathcal{A}(\omega)$ there exists a probability measure ν_2 on $(\Omega, \mathcal{F}_2)$ such that μ_ω is the conditional probability measure of ν_2. The measure ν_2 is given explicitly by formula (4.2). In particular, we have that $\nu_2(S) = \nu_1(S)$ for every $S \in \mathcal{F}_1$. This shows that not only can μ_ω be represented as a conditional probability of some measure ν_2, but that we can also fix the values of ν_2 on the sigma algebra $\mathcal{F}_1$ by taking an arbitrary probability measure ν_1 on $(\Omega, \mathcal{F}_1)$.

In order to extend the above analysis to a general setting, we proceed as follows.

Definition 4.1. We say that a multifunction $\mathcal{M}\colon \Omega \rightrightarrows \mathcal{P}_{\mathcal{Y}_2}$ is weakly* $\mathcal{F}_1$-measurable if for every $X \in \mathcal{X}_2$ the multifunction $\mathcal{M}_X\colon \Omega \rightrightarrows \mathbb{R}$, defined as

$$\mathcal{M}_X(\omega) := \{\langle \mu, X\rangle\colon \mu \in \mathcal{M}(\omega)\},$$

is $\mathcal{F}_1$-measurable. We say that a selection $\mu_\omega \in \mathcal{M}(\omega)$ is weakly* $\mathcal{F}_1$-measurable if for every $X \in \mathcal{X}_2$ the function $\omega \mapsto \langle \mu_\omega, X\rangle$ is $\mathcal{F}_1$-measurable.

The multifunction $\omega \mapsto \mathcal{A}(\omega)$, associated with representation (3.5), is weakly* $\mathcal{F}_1$-measurable. Indeed,

$$\mathcal{A}_X(\omega) = [-\rho_\omega(-X), \rho_\omega(X)],$$

and hence $\mathcal{F}_1$-measurability of $\mathcal{A}_X(\cdot)$ follows from the fact that $\rho(X) \in \mathcal{X}_1$, which ensures $\mathcal{F}_1$-measurability of $[\rho(X)](\cdot)$ and $[\rho(-X)](\cdot)$. In the sequel, whenever speaking about measurability of multifunctions and their selections, we shall mean weak* measurability.

By Theorem 3.1, for all $\omega \in \Omega$, every measure $\mu \in \mathcal{A}(\omega)$ satisfies condition (3.2). Therefore, if $\mu_\omega = \mu(\omega)$ is a selection of $\mathcal{A}(\omega)$, then

$$[\mu(\cdot)](S) = \mathbb{1}_S(\cdot), \quad \text{for all } S \in \mathcal{F}_1. \tag{4.3}$$

Moreover, if $\mu(\omega)$ is weakly* $\mathcal{F}_1$-measurable, then $[\mu(\cdot)](B)$ is $\mathcal{F}_1$-measurable, for every $B \in \mathcal{F}_2$.

Example 4.1. Consider $\rho(X) := \mathbb{E}[X \mid \mathcal{F}_1]$, $X \in \mathcal{X}_2$, where the conditional expectation is taken with respect to a probability measure P on $(\Omega, \mathcal{F}_2)$. It is assumed here that this conditional expectation is well defined for every $X \in \mathcal{X}_2$, and the space $\mathcal{X}_1$ is large enough such that it contains $\mathbb{E}[X \mid \mathcal{F}_1]$ for all $X \in \mathcal{X}_2$. Note that the function $\mathbb{E}[X \mid \mathcal{F}_1](\cdot)$ is defined up to a set of P-measure zero, i.e., two versions of $\mathbb{E}[X \mid \mathcal{F}_1](\cdot)$ can be different on a set of P-measure zero. The conditional expectation mapping ρ satisfies Assumptions (A1)–(A3) and is a linear mapping. Representation (3.5) holds, with $\mathcal{A}(\omega) = \{\mu(\omega)\}$ being a singleton and $\mu_\omega = \mu(\omega)$ being a probability measure on $(\Omega, \mathcal{F}_2)$. By the definition of the conditional expectation, $\mathbb{E}[X \mid \mathcal{F}_1]$ is $\mathcal{F}_1$-measurable, and hence μ_ω is weakly* $\mathcal{F}_1$-measurable. Considering $X = \mathbb{1}_A$ for $A \in \mathcal{F}_2$, we see that

$$\mu_\omega(A) = \mathbb{E}[\mathbb{1}_A \mid \mathcal{F}_1](\omega) = [P(A \mid \mathcal{F}_1)](\omega). \tag{4.4}$$

This means that $\mu(\cdot)$ is the conditional probability of P with respect to $\mathcal{F}_1$ (see, e.g., Billingsley [3, pp. 430–431]). Clearly, it satisfies (3.2).

Remark 4.1. The family of conditional risk mappings is closed under the operation of taking the maximum. That is, let $\rho^\nu = \rho^\nu_{\mathcal{X}_2|\mathcal{X}_1}$, $\nu \in \mathcal{I}$, be a family of conditional risk mappings satisfying Assumptions (A1)–(A3). Here, $\mathcal{I}$ is an arbitrary set. Suppose, further, that for every $X \in \mathcal{X}_2$ the function

$$[\rho(X)](\cdot) := \sup_{\nu\in\mathcal{I}} [\rho^\nu(X)](\cdot) \tag{4.5}$$

belongs to the space $\mathscr{X}_1$, and hence ρ maps $\mathscr{X}_2$ into $\mathscr{X}_1$. It is then straightforward to verify that the max-function ρ also satisfies Assumptions (A1)–(A3). Moreover, if ρ^ν, $\nu \in \mathscr{I}$, are lower semicontinuous, then ρ is also lower semicontinuous. In particular, let $\rho^\nu(X) := \mathbb{E}_\nu[X \mid \mathscr{F}_1]$, $\nu \in \mathscr{I}$, where $\mathscr{I}$ is a subset of the set of probability measures on $(\Omega, \mathscr{F}_2)$. Suppose that the corresponding max-function ρ is well defined, i.e., ρ maps $\mathscr{X}_2$ into $\mathscr{X}_1$. Then ρ is a lower-semicontinuous, positively homogeneous, conditional risk mapping. We show below that, under mild regularity conditions, the converse is also true, i.e., a positively homogeneous conditional risk mapping can be represented as the maximum of a family of conditional expectations.

Remark 4.2. Let $(\mathscr{I}, \mathscr{G}, Q)$ be a probability space and $\rho^\nu = \rho^\nu_{\mathscr{X}_2|\mathscr{X}_1}$, $\nu \in \mathscr{I}$, be a family of conditional risk mappings satisfying Assumptions (A1)–(A3). Suppose further that the integral mapping

$$[\rho(X)](\omega) := \int_{\mathscr{I}} \rho^\nu_\omega(X)\, dQ(\nu) \tag{4.6}$$

is well defined for every $X \in \mathscr{X}_2$ and $\omega \in \Omega$, and $\rho(X)$ is an element of $\mathscr{X}_1$. It is not difficult to see, then, that the integral mapping ρ: $\mathscr{X}_2 \to \mathscr{X}_1$ also satisfies Assumptions (A1)–(A3). Moreover, if each ρ^ν, $\nu \in \mathscr{I}$, is lower semicontinuous, then by Fatou's lemma, ρ is also lower semicontinuous.

Let $\mu_\omega \in \mathscr{A}(\omega)$ be a weakly* $\mathscr{F}_1$-measurable selection. It follows that for any set $B \in \mathscr{F}_2$, the function $\xi(\omega) := \mu_\omega(B)$ is $\mathscr{F}_1$-measurable. Consider a probability measure ν_1 on $(\Omega, \mathscr{F}_1)$. Then the function $\xi(\omega)$ can be viewed as a random variable on the probability space $(\Omega, \mathscr{F}_1, \nu_1)$. We can now define a set function ν_2 on $(\Omega, \mathscr{F}_2)$ as follows:

$$\nu_2(B) = \int_\Omega \mu_\omega(B)\, d\nu_1(\omega), \qquad B \in \mathscr{F}_2. \tag{4.7}$$

Proposition 4.1. *Let $\mu_\omega \in \mathscr{A}(\omega)$ be a weakly* $\mathscr{F}_1$-measurable selection. Then for every probability measure ν_1 on $(\Omega, \mathscr{F}_1)$ the set function (4.7) is a probability measure on $(\Omega, \mathscr{F}_2)$, which is equal to ν_1 on $\mathscr{F}_1$, and is such that μ_ω is the conditional probability of ν_2 with respect to $\mathscr{F}_1$.*

Proof. The set function ν_2 is a probability measure, as an average of probability measures μ_ω. For $B \in \mathscr{F}_1$, by virtue of (4.3), formula (4.7) yields:

$$\nu_2(B) = \nu_1(B), \qquad B \in \mathscr{F}_1.$$

It remains to prove that μ_ω is the conditional probability of ν_2 with respect to the sigma subalgebra $\mathscr{F}_1$, i.e., that $\mu_\omega(B) = [\nu_2(B \mid \mathscr{F}_1)](\omega)$ for any $B \in \mathscr{F}_2$ and a.e. ω. Let us consider sets $B \in \mathscr{F}_2$ and $S \in \mathscr{F}_1$. From (4.7) we obtain

$$\nu_2(B \cap S) = \int_\Omega \mu_\omega(B \cap S)\, d\nu_1(\omega). \tag{4.8}$$

We have $\mu_\omega(B) = \mu_\omega(B \cap S) + \mu_\omega(B \backslash S)$ and $\mu_\omega(B \backslash S) \le \mu_\omega(\Omega \backslash S)$. Because $\Omega \backslash S \in \mathscr{F}_1$, it follows from (4.3) that $\mu_\omega(\Omega \backslash S) = 0$ for all $\omega \in S$. Hence, $\mu_\omega(B \backslash S) = 0$, for all $\omega \in S$. Thus, $\mu_\omega(B) = \mu_w(B \cap S)$, and Equation (4.8) can be rewritten as follows:

$$\nu_2(B \cap S) = \int_S \mu_\omega(B)\, d\nu_2(\omega), \quad \text{for all } S \in \mathscr{F}_1 \text{ and } B \in \mathscr{F}_2. \tag{4.9}$$

This is equivalent to the statement that μ_ω is the conditional probability of ν_2 with respect to $\mathscr{F}_1$ (see, e.g., Billingsley [3, p. 430]). □

Recall that $\mathscr{X}_i$ and $\mathscr{Y}_i$ are assumed to be paired locally convex topological vector spaces. It is said that $\mathscr{X}_i$ is *separable* if it has a countable dense subset.

Lemma 4.1. *Suppose that the space $\mathscr{X}_2$ is separable and the representation (3.5) holds. Then there exists a countable family μ^i_ω, $i \in \mathbb{N}$, of weakly* $\mathscr{F}_1$-measurable selections of $\mathscr{A}(\omega)$ such that*

$$[\rho(X)](\omega) = \sup_{i \in \mathbb{N}} \langle \mu^i_\omega, X\rangle \tag{4.10}$$

for all $X \in \mathscr{X}_2$ and $\omega \in \Omega$.

Proof. Let $\varepsilon_k \downarrow 0$ be a sequence of positive numbers and $\{X_n\}_{n\in\mathbb{N}}$ be a dense subset of $\mathscr{X}_2$. For every k, $n \in \mathbb{N}$ consider the multifunction

$$M_{k,n}(\omega) := \{\nu \in \mathscr{A}(\omega) \colon \langle \nu, X_n\rangle \ge [\rho(X_n)](\omega) - \varepsilon_k\}.$$

This multifunction is weakly* $\mathcal{F}_1$-measurable and nonempty valued. Because $\mathcal{X}_2$ is separable, the multifunction $M_{k,n}(\cdot)$ admits a weakly* $\mathcal{F}_1$-measurable selection $\mu^{k,n}(\cdot)$ (see Kuratowski and Ryll-Nardzewski [9]). By the definition of $M_{k,n}$, we have that

$$\langle \mu^{k,n}(\omega), X_n\rangle \geq [\rho(X_n)](\omega) - \varepsilon_k$$

for all $k, n \in \mathbb{N}$ and $\omega \in \Omega$. Because $\rho_\omega(\cdot)$ is lower semicontinuous for every $\omega \in \Omega$, it follows that

$$\sup_{k,n} \langle \mu^{k,n}(\omega), X\rangle \geq [\rho(X)](\omega), \quad \omega \in \Omega.$$

Because of (3.5), we also have that

$$[\rho(X)](\omega) \geq \sup_{k,n} \langle \mu^{k,n}(\omega), X\rangle, \quad \omega \in \Omega.$$

Thus, representation (4.10) follows with $\mu^i_\omega := \mu^{k,n}(\omega)$, $i = (k,n) \in \mathbb{N} \times \mathbb{N}$. □

Remark 4.3. Under the assumptions of Lemma 4.1, we can also write the following representation

$$[\rho(X)](\omega) = \sup_{\mu(\cdot) \in \mathcal{A}(\cdot)} \langle \mu(\omega), X\rangle, \tag{4.11}$$

where the supremum is taken over all weakly* $\mathcal{F}_1$-measurable selections $\mu(\omega) \in \mathcal{A}(\omega)$.

We can now formulate the main result of this section.

Theorem 4.1. *Let $\rho = \rho_{\mathcal{X}_2|\mathcal{X}_1}$ be a positively homogeneous, lower-semicontinuous, conditional risk mapping. Suppose that the space $\mathcal{X}_2$ is separable. Then, for every probability measure ν on $(\Omega, \mathcal{F}_1)$, there exists a countable family $\nu^i \in \mathcal{P}_{\mathcal{Y}_2}$, $i \in \mathbb{N}$, of probability measures on $(\Omega, \mathcal{F}_2)$ that agree with ν on $\mathcal{F}_1$ and are such that*

$$\rho_\omega(\cdot) = \sup_{i \in \mathbb{N}} \mathbb{E}_{\nu^i}[\cdot \mid \mathcal{F}_1](\omega), \quad \omega \in \Omega. \tag{4.12}$$

Proof. By Theorem 3.1, representation (3.5) holds. The assertion then follows by Lemma 4.1 together with Proposition 4.1. □

Remark 4.4. If in Theorem 4.1 we remove the assumption that ρ is positively homogeneous, then by the above analysis, Theorem 3.1 implies the following extension of the representation (4.12):

$$\rho_\omega(\cdot) = \sup_{i \in \mathbb{N}} \{\mathbb{E}_{\nu^i}[\cdot \mid \mathcal{F}_1](\omega) - \gamma^i(\omega)\}, \quad \omega \in \Omega, \tag{4.13}$$

where

$$\gamma^i(\omega) := \sup_{X \in \mathcal{X}_2} \{\mathbb{E}_{\nu^i}[X \mid \mathcal{F}_1](\omega) - \rho_\omega(X)\}. \tag{4.14}$$

Remark 4.5. Assuming that the representation (4.12) holds, we have that for any $Y \in \mathcal{X}_1$, $Y \succeq 0$, and $X \in \mathcal{X}_2$ such that $YX \in \mathcal{X}_2$,

$$\rho_\omega(YX) = \sup_{i \in \mathbb{N}} \mathbb{E}_{\nu^i}[YX \mid \mathcal{F}_1](\omega) = Y(\omega) \sup_{i \in \mathbb{N}} \mathbb{E}_{\nu^i}[X \mid \mathcal{F}_1](\omega) = Y(\omega)\rho_\omega(X), \quad \omega \in \Omega.$$

That is, under the assumptions of Theorem 4.1, the result of Proposition 3.2 (i.e., Equation (3.6)) holds for a general function $Y \in \mathcal{X}_1$ (compare with Remark 3.1).

5. Iterated risk mappings. In order to formulate dynamic programming equations for multistage stochastic optimization problems involving risk measures, we need to consider compositions of several conditional risk mappings. In this section we give a preliminary analysis of this subject. Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ be sigma algebras, $\mathcal{X}_1 \subset \mathcal{X}_2 \subset \mathcal{X}_3$ be respective spaces of measurable functions, with dual spaces $\mathcal{Y}_1 \subset \mathcal{Y}_2 \subset \mathcal{Y}_3$, and let $\rho_{\mathcal{X}_3|\mathcal{X}_2}: \mathcal{X}_3 \to \mathcal{X}_2$ and $\rho_{\mathcal{X}_2|\mathcal{X}_1}: \mathcal{X}_2 \to \mathcal{X}_1$ be conditional risk mappings. (For any inclusion like $\mathcal{X}_2 \subset \mathcal{X}_3$, we assume that the topology of $\mathcal{X}_2$ is induced by the topology of $\mathcal{X}_3$.) Then it can be easily verified that the composite mapping $\rho_{\mathcal{X}_3|\mathcal{X}_1}: \mathcal{X}_3 \to \mathcal{X}_1$, defined by

$$\rho_{\mathcal{X}_3|\mathcal{X}_1} := \rho_{\mathcal{X}_2|\mathcal{X}_1} \circ \rho_{\mathcal{X}_3|\mathcal{X}_2}, \tag{5.1}$$

is also a conditional risk mapping.

Suppose that both conditional risk mappings at the right-hand side of (5.1) are positively homogeneous and lower semicontinuous. We then have

$$[\rho_{\mathscr{X}_3|\mathscr{X}_2}(X)](\tilde{\omega}) = \sup_{\mu_2 \in \mathscr{A}_2(\tilde{\omega})} \langle \mu_2, X\rangle, \quad X \in \mathscr{X}_3, \ \ \tilde{\omega} \in \Omega, \tag{5.2}$$

$$[\rho_{\mathscr{X}_2|\mathscr{X}_1}(Y)](\omega) = \sup_{\mu_1 \in \mathscr{A}_1(\omega)} \langle \mu_1, Y\rangle, \quad Y \in \mathscr{X}_2, \ \ \omega \in \Omega, \tag{5.3}$$

with the multifunctions $\mathscr{A}_2\colon \Omega \rightrightarrows \mathscr{P}_{\mathscr{Y}_3}$ and $\mathscr{A}_1\colon \Omega \rightrightarrows \mathscr{P}_{\mathscr{Y}_2}$ having closed convex values and weakly* measurable with respect to $\mathscr{F}_2$ and $\mathscr{F}_1$, correspondingly. In order to analyze composition (5.1), it is convenient to consider weakly* measurable selections $\mu_i(\cdot)$ of the multifunctions $\mathscr{A}_i(\cdot)$, $i = 1, 2$.

Proposition 5.1. *Suppose that the space $\mathscr{X}_3$ is separable, and $\rho_{\mathscr{X}_2|\mathscr{X}_1}$ and $\rho_{\mathscr{X}_3|\mathscr{X}_2}$ are positively homogeneous, lower semicontinuous, and satisfy Assumptions* (A1)–(A3). *Then the conditional risk mapping $\rho_{\mathscr{X}_3|\mathscr{X}_1}$ can be represented in the form*

$$[\rho_{\mathscr{X}_3|\mathscr{X}_1}(X)](\omega) = \sup_{\mu_1 \in \mathscr{A}_1(\omega)} \ \sup_{\mu_2(\cdot) \in \mathscr{A}_2(\cdot)} \int_\Omega \langle \mu_2(\tilde{\omega}), X\rangle \, d\mu_1(\tilde{\omega}), \tag{5.4}$$

*where the second "*sup*" operation on the right-hand side of* (5.4) *is taken with respect to weakly* $\mathscr{F}_2$*-measurable selections $\mu_2(\tilde{\omega}) \in \mathscr{A}_2(\tilde{\omega})$.*

Proof. By (5.2) and (5.3) we have that, for every $\omega \in \Omega$,

$$[\rho_{\mathscr{X}_3|\mathscr{X}_1}(X)](\omega) = \sup_{\mu_1 \in \mathscr{A}_1(\omega)} \int_\Omega \sup_{\mu_2 \in \mathscr{A}_2(\tilde{\omega})} \langle \mu_2, X\rangle \, d\mu_1(\tilde{\omega}).$$

By Lemma 4.1 (see Remark 4.3), we also have that

$$\sup_{\mu_2 \in \mathscr{A}_2(\tilde{\omega})} \langle \mu_2, X\rangle = \sup_{\mu_2(\cdot) \in \mathscr{A}_2(\cdot)} \langle \mu_2(\tilde{\omega}), X\rangle, \tag{5.5}$$

where the supremum in the right-hand side of (5.5) is taken over all weakly* $\mathscr{F}_2$-measurable selections $\mu_2(\tilde{\omega}) \in \mathscr{A}_2(\tilde{\omega})$. Consequently,

$$[\rho_{\mathscr{X}_3|\mathscr{X}_1}(X)](\omega) = \sup_{\mu_1 \in \mathscr{A}_1(\omega)} \int_\Omega \sup_{\mu_2(\cdot) \in \mathscr{A}_2(\cdot)} \langle \mu_2(\tilde{\omega}), X\rangle \, d\mu_1(\tilde{\omega}). \tag{5.6}$$

Similarly to the proof of Lemma 4.1, we can now interchange the integral and "sup" operators at the right-hand side of (5.6), and hence (5.4) follows. □

Remark 5.1. By Lemma 4.1 we have that it actually suffices to take the second supremum at the right-hand side of (5.4) with respect to a *countable* number of weakly* $\mathscr{F}_2$-measurable selections $\mu_2(\tilde{\omega}) \in \mathscr{A}_2(\tilde{\omega})$.

Representation (5.4) means that $\rho_{\mathscr{X}_3|\mathscr{X}_1}$ can be written in form (3.5) with the set $\mathscr{A}(\omega)$ is formed by *all* measures $\mu \in \mathscr{Y}_3$ representable in the form

$$\mu(S) = \int_\Omega [\mu_2(\tilde{\omega})](S) \, d\mu_1(\tilde{\omega}), \quad S \in \mathscr{F}_3, \tag{5.7}$$

where $\mu_2(\cdot) \in \mathscr{A}_2(\cdot)$ is a weakly* $\mathscr{F}_2$-measurable selection and $\mu_1 \in \mathscr{A}_1(\omega)$. We denote the multifunction $\mathscr{A}$ by $\mathscr{A}_1 \circ \mathscr{A}_2$.

Consider now a sequence of sigma algebras (a filtration) $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}_T$, with $\mathscr{F}_1 = \{\varnothing, \Omega\}$ and $\mathscr{F}_T = \mathscr{F}$. We define linear (locally convex topological vector) spaces $\mathscr{X}_1 \subset \cdots \subset \mathscr{X}_T$ of real-valued functions on Ω such that all functions in $\mathscr{X}_t$ are $\mathscr{F}_t$-measurable. We also introduce the corresponding paired spaces $\mathscr{Y}_1 \subset \cdots \subset \mathscr{Y}_T$ of measures, $t = 1, \ldots, T$. Let $\rho_{\mathscr{X}_t|\mathscr{X}_{t-1}}$, $t = 2, \ldots, T$, be conditional risk mappings. Note that because $\mathscr{F}_1 = \{\varnothing, \Omega\}$, the space $\mathscr{X}_1$ is formed by constant over Ω functions and can be identified with $\mathbb{R}$, and hence $\rho_{\mathscr{X}_2|\mathscr{X}_1}$ is an (unconditional) risk function.

With the above sequence of conditional risk mappings, we associate the following (unconditional) risk functions

$$\rho_t := \rho_{\mathscr{X}_2|\mathscr{X}_1} \circ \cdots \circ \rho_{\mathscr{X}_{t-1}|\mathscr{X}_{t-2}} \circ \rho_{\mathscr{X}_t|\mathscr{X}_{t-1}}, \quad t = 2, \ldots, T. \tag{5.8}$$

The recursive application of Proposition 5.1 renders the following result.

THEOREM 5.1. *Let $\rho_{\mathscr{X}_{t+1}|\mathscr{X}_t}$, $t = 1, \ldots, T-1$, be positively homogeneous, lower-semicontinuous, conditional risk mappings. Suppose that the spaces $\mathscr{X}_t$, $t = 2, \ldots, T$, are separable. Then for every $X \in \mathscr{X}_t$, $t = 2, \ldots, T$,*

$$\rho_t(X) = \sup_{\mu \in \mathscr{A}_1 \circ \cdots \circ \mathscr{A}_{t-1}} \langle \mu, X \rangle, \tag{5.9}$$

where each $\mathscr{A}_\tau \colon \Omega \rightrightarrows \mathscr{P}_{\mathscr{Y}_{\tau+1}}$ is weakly $\mathscr{F}_\tau$-measurable and such that*

$$[\rho_{\mathscr{X}_{\tau+1}|\mathscr{X}_\tau}(X)](\omega) = \sup_{\mu \in \mathscr{A}_\tau(\omega)} \langle \mu, X \rangle. \tag{5.10}$$

Note that we always have $(\mathscr{A}_1 \circ \mathscr{A}_2) \circ \mathscr{A}_3 = \mathscr{A}_1 \circ (\mathscr{A}_2 \circ \mathscr{A}_3)$, and therefore there is no ambiguity in the notation $\mathscr{A}_1 \circ \cdots \circ \mathscr{A}_{t-1}$.

REMARK 5.2. Although formula (5.7) suggests a way for calculating the composition $\mathscr{A}_1 \circ \cdots \circ \mathscr{A}_{t-1}$ in the max-representation (5.9), its practical application is difficult. It seems that this is not a drawback of that formula, but rather a nature of the considered problem. In the classical setting of multistage stochastic programming, the situation simplifies considerably if the underlying process satisfies the so-called *between-stages independence* condition. That is what we discuss next.

Let $\xi_1, \ldots, \xi_T$ be a sequence of random vectors, $\xi_t \in \mathbb{R}^{d_t}$, on a probability space $(\Omega, \mathscr{F}, P)$, representing evolution of random data at times $t = 1, \ldots, T$. Let $\mathscr{F}_t$ be the sigma subalgebra of $\mathscr{F}$ generated by random vector $\xi_{[t]} := (\xi_1, \ldots, \xi_t)$, $t = 1, \ldots, T$. Clearly, the inclusions $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}_T$ hold. For $p \in [1, +\infty)$, we assume now that each space $\mathscr{X}_t$ is formed by functions of $\xi_{[t]}$ with finite pth moment. That is, every $X \in \mathscr{X}_t$ can be represented in the form $X(\omega) = \tilde{X}(\xi_{[t]}(\omega))$ and $\int_\Omega |X|^p \, dP < +\infty$, i.e., $\mathscr{X}_t = \mathscr{L}_p(\Omega, \mathscr{F}_t, P)$. We then take $\mathscr{Y}_t := \mathscr{L}_q(\Omega, \mathscr{F}_t, P)$ and use the corresponding scalar product of the form (2.3). With a slight abuse of the notation we sometimes write an element X of $\mathscr{X}_t$ as $X(\xi_{[t]})$ and an element h of $\mathscr{Y}_t$ as $h(\xi_{[t]})$. In this framework, the set $\mathscr{A}_\tau(\omega)$ in the max-representation (5.10) is a function of $\xi_{[t]}$. It also makes sense here to talk about the "between-stages independence" condition in the sense that random vectors ξ_{t+1} and $\xi_{[t]}$ are independent for $t = 1, \ldots, T-1$. Under this condition, the dynamic programming equations, which will be discussed in §7, simplify considerably.

6. Examples of conditional risk mappings.

In this section we discuss some examples of conditional risk mappings that can be considered as natural extensions of the corresponding examples of (real-valued) risk functions (see Ruszczyński and Shapiro [17]). We use the framework and notation of §2, and take P to be a probability measure on $(\Omega, \mathscr{F}_2)$. Unless stated otherwise, all expectations and probability statements in this section are made with respect to P.

EXAMPLE 6.1. Let $\mathscr{X}_i := \mathscr{L}_p(\Omega, \mathscr{F}_i, P)$ and $\mathscr{Y}_i := \mathscr{L}_q(\Omega, \mathscr{F}_i, P)$, $i = 1, 2$, for some $p \in [1, +\infty)$. Consider

$$\rho(X) := \mathbb{E}[X \mid \mathscr{F}_1] + c\sigma_p(X \mid \mathscr{F}_1), \quad X \in \mathscr{X}_2, \tag{6.1}$$

where $c \geq 0$ and $\sigma_p(\cdot \mid \mathscr{F}_1)$ is the conditional upper semideviation:

$$\sigma_p(X \mid \mathscr{F}_1) := (\mathbb{E}[(X - \mathbb{E}[X \mid \mathscr{F}_1])_+^p \mid \mathscr{F}_1])^{1/p}. \tag{6.2}$$

If the sigma algebra $\mathscr{F}_1$ is trivial, then $\mathbb{E}[\cdot \mid \mathscr{F}_1] = \mathbb{E}[\cdot]$ and $\sigma_p(X \mid \mathscr{F}_1)$ becomes the upper semideviation of X of order p. Thus, ρ is the conditional counterpart of the mean–p^{th}-semideviation models of Ogryczak and Ruszczyński [10, 11].

Let us show that for $c \in [0, 1]$, the above mapping ρ satisfies Assumptions (A1)–(A3). Assumption (A3) can be verified directly. That is, if $Y \in \mathscr{X}_1$ and $X \in \mathscr{X}_2$, then

$$\begin{aligned}\rho(X+Y) &= \mathbb{E}[X+Y \mid \mathscr{F}_1] + c(\mathbb{E}[(X+Y-\mathbb{E}[X+Y \mid \mathscr{F}_1])_+^p \mid \mathscr{F}_1])^{1/p} \\ &= \mathbb{E}[X \mid \mathscr{F}_1] + Y + c(\mathbb{E}[(X - \mathbb{E}[X \mid \mathscr{F}_1])_+^p \mid \mathscr{F}_1])^{1/p} = \rho(X) + Y.\end{aligned}$$

In order to verify Assumptions (A1) and (A2), consider function ρ_ω defined in (2.1). For $\omega \in \Omega$ we can write

$$\mathbb{E}[\cdot \mid \mathscr{F}_1](\omega) = \mathbb{E}_{\mu_\omega}[\cdot], \tag{6.3}$$

where $\mu(\omega) = \mu_\omega$ is the conditional probability of P with respect to $\mathscr{F}_1$ (see Example 4.1). Therefore, for any $X \in \mathscr{X}_2$ and $\omega \in \Omega$,

$$\rho_\omega(X) = \mathbb{E}_{\mu_\omega}[X] + c\big(\mathbb{E}_{\mu_\omega}[(X - \mathbb{E}_{\mu_\omega}[X])_+^p]\big)^{1/p}. \tag{6.4}$$

We have that $\mu_\omega \in \mathcal{P}_{\mathcal{Y}_2|\mathcal{F}_1}(\omega)$ and its (conditional probability) density $f_\omega = d\mu_\omega/dP$ has the following properties: $f_\omega \in \mathcal{Y}_2$, $f_\omega \geq 0$, for any $A \in \mathcal{F}_2$, the function $\omega \mapsto \int_A f_\omega \, dP$ is $\mathcal{F}_1$-measurable and, moreover, for any $B \in \mathcal{F}_1$, the following equality holds

$$\int_B \int_A f_\omega(\tilde{\omega}) \, dP(\tilde{\omega}) \, dP(\omega) = P(A \cap B).$$

We see that for a fixed ω the function $\rho_\omega(X)$ is identical with the risk function analyzed in Ruszczyński and Shapiro [17, Example 2]; the conditional measure μ_ω plays the role of the probability measure. It follows from the analysis in Ruszczyński and Shapiro [17] that, for $c \in [0, 1]$, the function $\rho_\omega(\cdot)$ satisfies Assumptions (A1) and (A2). Moreover, the representation

$$\rho_\omega(X) = \sup_{\gamma \in \mathcal{A}^*} \int_\Omega \gamma X \, d\mu_\omega$$

holds with

$$\mathcal{A}^* = \left\{ \gamma = 1 + h - \int_\Omega h \, d\mu_\omega \colon \int_\Omega h^q \, d\mu_\omega \leq c^q, \; h \succeq 0 \right\}.$$

Because $d\mu_\omega = f_\omega dP$, we conclude that the representation (3.5) follows with

$$\mathcal{A}(\omega) = \{ g \in \mathcal{Y}_2 \colon g = f_\omega(1 + h - \mathbb{E}[f_\omega h]), \; h \in cB_q(\omega), \; h \succeq 0 \}, \tag{6.5}$$

where

$$B_q(\omega) := \{ h \in \mathcal{Y}_2 \colon \mathbb{E}[h^q f_\omega] \leq 1 \}.$$

Consider now the framework outlined in Remark 5.2. That is, there are two random vectors ξ_1, ξ_2; the sigma algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ are generated by ξ_1 and (ξ_1, ξ_2), respectively; an element $X \in \mathcal{X}_2$ is a function of (ξ_1, ξ_2); and the conditional expectations in (6.1) and (6.2) are taken with respect to the random vector ξ_1. We then have that $[\rho(X)](\xi_1)$ is a function of ξ_1. Now if ξ_1 and ξ_2 are *independent* and every $X \in \mathcal{X}_2$ is a function of ξ_2 only, then the corresponding conditional expectations are independent of ξ_1, and hence $[\rho(X)](\xi_1)$ is constant, and

$$\rho(X) = \mathbb{E}[X] + c \left(\mathbb{E}[(X - \mathbb{E}[X])_+^p] \right)^{1/p}.$$

In that case $\rho(X)$ can be viewed as the mean–pth-semideviation risk function.

Example 6.2. Let $\mathcal{X}_i := \mathcal{L}_1(\Omega, \mathcal{F}_i, P)$ and $\mathcal{Y}_i := \mathcal{L}_\infty(\Omega, \mathcal{F}_i, P)$, $i = 1, 2$. For constants $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, consider

$$\rho(X) := \mathbb{E}[X \mid \mathcal{F}_1] + \Phi(X \mid \mathcal{F}_1), \quad X \in \mathcal{X}_2, \tag{6.6}$$

where

$$[\Phi(X \mid \mathcal{F}_1)](\omega) := \inf_{Z \in \mathcal{X}_1} \mathbb{E}\{ \varepsilon_1 [Z - X]_+ + \varepsilon_2 [X - Z]_+ \mid \mathcal{F}_1 \}(\omega). \tag{6.7}$$

It is straightforward to verify that Assumption (A3) holds here. Indeed, for $X \in \mathcal{X}_2$ and $Y \in \mathcal{X}_1$, we have

$$[\Phi(X + Y \mid \mathcal{F}_1)](\omega) = \inf_{Z \in \mathcal{X}_1} \mathbb{E}\{ \varepsilon_1 [(Z - Y) - X]_+ + \varepsilon_2 [X - (Z - Y)]_+ \mid \mathcal{F}_1 \}(\omega).$$

By changing of variables $Z \mapsto Z - Y$, we obtain that $\Phi(X + Y \mid \mathcal{F}_1) = \Phi(X \mid \mathcal{F}_1)$, and hence Assumption (A3) follows.

Because of (6.3), we can write, as in the previous example, that

$$\rho_\omega(X) = \mathbb{E}_{\mu_\omega}[X] + \inf_{Z \in \mathcal{X}_1} \mathbb{E}_{\mu_\omega}\{ \varepsilon_1 [Z - X]_+ + \varepsilon_2 [X - Z]_+ \} \tag{6.8}$$

$$= \mathbb{E}[f_\omega X] + \inf_{Z \in \mathcal{X}_1} \mathbb{E}\{ \varepsilon_1 [f_\omega Z - f_\omega X]_+ + \varepsilon_2 [f_\omega X - f_\omega Z]_+ \}, \tag{6.9}$$

where $f_\omega = d\mu_\omega/dP$ is the conditional density. We can continue now in a way similar to the analysis of Example 3 in Ruszczyński and Shapiro [17]. We have that

$$\mathbb{E}\{ \varepsilon_1 [f_\omega Z - f_\omega X]_+ + \varepsilon_2 [f_\omega X - f_\omega Z]_+ \} = \sup_{h \in \mathcal{M}} \mathbb{E}[h(f_\omega X - f_\omega Z)], \tag{6.10}$$

where

$$\mathcal{M} := \{ h \in \mathcal{Y}_2 \colon -\varepsilon_1 \leq h(\omega) \leq \varepsilon_2, \text{ a.e. } \omega \in \Omega \}. \tag{6.11}$$

By substituting the right-hand side of (6.10) into (6.9), we obtain

$$\rho_\omega(X) = \mathbb{E}[f_\omega X] + \inf_{Z\in\mathcal{X}_1} \sup_{h\in\mathcal{M}} \mathbb{E}[h(f_\omega X - f_\omega Z)].$$

Because the set $\mathcal{M}$ is compact in the weak* topology of $\mathcal{Y}_2$, we can interchange the "inf" and "sup" operators in the right-hand side of the above equation. Also, we have that

$$\inf_{Z\in\mathcal{X}_1} \mathbb{E}[hf_\omega Z] = \inf_{Z\in\mathcal{X}_1} \mathbb{E}[Z\mathbb{E}[hf_\omega \mid \mathcal{F}_1]] = \begin{cases} 0, & \text{if } \mathbb{E}[hf_\omega \mid \mathcal{F}_1] = 0,\\ -\infty, & \text{otherwise.}\end{cases}$$

We obtain that

$$\rho_\omega(X) = \mathbb{E}[f_\omega X] + \sup\{\mathbb{E}[hf_\omega X]\colon h\in\mathcal{M},\ \mathbb{E}[hf_\omega \mid \mathcal{F}_1] = 0\}.$$

It follows that for $\varepsilon_1 \in (0,1]$ and $\varepsilon_2 > 0$, Assumptions (A1) and (A2) are satisfied, and representation (3.5) holds with

$$\mathcal{A}(\omega) = \{g\in\mathcal{Y}_2\colon g = (1+h)f_\omega,\ h\in\mathcal{M},\ \mathbb{E}[hf_\omega \mid \mathcal{F}_1] = 0\}, \tag{6.12}$$

where $\mathcal{M}$ is defined in (6.11).

It is also possible to derive the max-representation of conditional expectations, of the form (4.12), either from (6.12) or directly as follows. We have that

$$\varepsilon_1[Z-X]_+ + \varepsilon_2[X-Z]_= \sup_{t\in[-\varepsilon_1,\varepsilon_2]} t(X-Z).$$

Consequently, by interchanging the "sup" and integral operators (see the following Remark 7.1) and then the "inf" and "sup" operators, we obtain

$$\begin{aligned}[\Phi(X\mid\mathcal{F}_1)](\omega) &= \inf_{Z\in\mathcal{X}_1}\sup_{\tau\in\mathcal{M}} \mathbb{E}\{\tau(X-Z)\mid\mathcal{F}_1\}(\omega)\\ &= \sup_{\tau\in\mathcal{M}}\inf_{Z\in\mathcal{X}_1} \mathbb{E}\{\tau(X-Z)\mid\mathcal{F}_1\}(\omega).\end{aligned}$$

Now

$$\inf_{Z\in\mathcal{X}_1} \mathbb{E}[-\tau Z\mid\mathcal{F}_1](\omega) = \inf_{Z\in\mathcal{X}_1} Z(\omega)\mathbb{E}[-\tau\mid\mathcal{F}_1](\omega) = \begin{cases} 0, & \text{if } \mathbb{E}[-\tau\mid\mathcal{F}_1] = 0,\\ -\infty, & \text{otherwise.}\end{cases}$$

It follows that

$$[\Phi(X\mid\mathcal{F}_1)](\omega) = \sup\{\mathbb{E}[\tau X\mid\mathcal{F}_1](\omega)\colon \tau\in\mathcal{M},\ \mathbb{E}[\tau\mid\mathcal{F}_1] = 0\}, \tag{6.13}$$

and hence, for $\varepsilon_1 \in (0,1]$ and $\varepsilon_2 > 0$,

$$\rho_\omega(X) = \sup\{\mathbb{E}[\tau X\mid\mathcal{F}_1](\omega)\colon 1-\varepsilon_1 \le \tau(\omega) \le 1+\varepsilon_2,\ \text{a.e. } \omega\in\Omega,\ \mathbb{E}[\tau\mid\mathcal{F}_1] = 1\}. \tag{6.14}$$

The above Equation (6.14) can also be written as follows

$$\rho_\omega(X) = \sup_{\nu\in\mathcal{D}} \mathbb{E}_\nu[X\mid\mathcal{F}_1](\omega), \tag{6.15}$$

where

$$\mathcal{D} := \{\nu = \tau dP\colon \tau\in\mathcal{P}_{\mathcal{Y}_2},\ 1-\varepsilon_1 \le \tau(\omega) \le 1+\varepsilon_2,\ \text{a.e. } \omega\in\Omega\}. \tag{6.16}$$

Because

$$\varepsilon_1[Z-X]_+ + \varepsilon_2[X-Z]_+ = \varepsilon_1(Z + (1-p)^{-1}[X-Z]_+ - X),$$

where $p := \varepsilon_2/(\varepsilon_1+\varepsilon_2)$, we have that

$$\rho(X) = (1-\varepsilon_1)\mathbb{E}[X\mid\mathcal{F}_1] + \varepsilon_1 \mathrm{CV@R}_{\mathcal{X}_2\mid\mathcal{X}_1}[X], \tag{6.17}$$

where

$$\mathrm{CV@R}_{\mathcal{X}_2\mid\mathcal{X}_1}[X](\omega) := \inf_{Z\in\mathcal{X}_1} \mathbb{E}\{Z + (1-p)^{-1}[X-Z]_+ \mid \mathcal{F}_1\}(\omega). \tag{6.18}$$

Clearly, for $\varepsilon_1 = 1$ we have that $\rho(\cdot) = \mathrm{CV@R}_{\mathcal{X}_2\mid\mathcal{X}_1}[\cdot]$. By the above analysis, we obtain that for $p \in (0,1)$, $\mathrm{CV@R}_{\mathcal{X}_2\mid\mathcal{X}_1}[\cdot]$ is a positively homogeneous, continuous risk mapping. If $\mathcal{F}_1 = \{\varnothing, \Omega\}$, then $\mathrm{CV@R}_{\mathcal{X}_2\mid\mathcal{X}_1}[\cdot]$ becomes the Conditional Value at Risk function analyzed in Rockafellar and Uryasev [14], Rockafellar et al. [16], and Shapiro and Ahmed [18]. For a nontrivial $\mathcal{F}_1$, the measure $\mathrm{CV@R}_{\mathcal{X}_2\mid\mathcal{X}_1}$ was analyzed in Pflug and Ruszczyński [12].

7. Multistage risk optimization problems. In order to construct risk models for multistage decision problems, we first introduce recursive risk models for sequences.

As in §5, consider a sequence of sigma algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_T$, with $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and $\mathcal{F}_T = \mathcal{F}$, and let $\mathcal{X}_1 \subset \cdots \subset \mathcal{X}_T$ be a corresponding sequence of linear spaces of $\mathcal{F}_t$-measurable functions, $t = 1, \ldots, T$. Let $\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}: \mathcal{X}_t \to \mathcal{X}_{t-1}$ be conditional risk mappings. Denote $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_T$ and $X := (X_1, X_2, \ldots, X_T)$, where $X_t \in \mathcal{X}_t$, $t = 1, \ldots, T$, and define a function $\tilde{\rho}: \mathcal{X} \to \mathbb{R}$ as follows:

$$\tilde{\rho}(X) := X_1 + \rho_{\mathcal{X}_2|\mathcal{X}_1}[X_2 + \rho_{\mathcal{X}_3|\mathcal{X}_2}(X_3 + \cdots + \rho_{\mathcal{X}_{T-1}|\mathcal{X}_{T-2}}[X_{T-1} + \rho_{\mathcal{X}_T|\mathcal{X}_{T-1}}(X_T)])]. \tag{7.1}$$

Because $\mathcal{F}_1 = \{\emptyset, \Omega\}$, the space $\mathcal{X}_1$ can be identified with $\mathbb{R}$, and hence $\tilde{\rho}(X)$ is real valued. By Assumption (A3) we have

$$X_{T-1} + \rho_{\mathcal{X}_T|\mathcal{X}_{T-1}}(X_T) = \rho_{\mathcal{X}_T|\mathcal{X}_{T-1}}(X_{T-1} + X_T).$$

Applying this formula for $t = T, T-1, \ldots, 2$ we obtain the equation:

$$\tilde{\rho}(X) = \rho_T(X_1 + \cdots + X_T), \tag{7.2}$$

where, similarly to (5.8),

$$\rho_t := \rho_{\mathcal{X}_2|\mathcal{X}_1} \circ \cdots \circ \rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}, \quad t = 2, \ldots, T. \tag{7.3}$$

Because each conditional risk mapping $\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}$ satisfies (A1)–(A3), it follows that the function ρ_T satisfies (A1)–(A3) as well. Moreover, if conditional risk mappings $\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}$ are positively homogeneous, then ρ is positively homogeneous. Assuming further that the spaces $\mathcal{X}_t$ are separable and $\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}$ are lower semicontinuous, we obtain by Theorem 5.1 that the following representation holds true

$$\tilde{\rho}(X) = \sup_{\mu \in \mathfrak{A}} \mathbb{E}_\mu[X_1 + \cdots + X_T], \tag{7.4}$$

where the set $\mathfrak{A} := \mathfrak{A}_1 \circ \cdots \circ \mathfrak{A}_{T-1}$ is given by the composition of the multifunctions $\mathfrak{A}_\tau: \Omega \rightrightarrows \mathcal{P}_{\mathcal{Y}_{\tau+1}}$, $\tau = 1, \ldots, T-1$, defined in Equation (5.10) of Theorem 5.1.

It may be of interest to discuss the difference between our approach and a construction in Artzner et al. [2]. In Artzner et al. [2] an adapted sequence $\{X_t\}$, $t = 1, \ldots, T$, is viewed as a measurable function on a new measurable space $(\Omega', \mathcal{F}')$, with $\Omega' = \Omega \times \{1, \ldots, T\}$, and with the sigma algebra $\mathcal{F}'$ generated by sets of form $B_t \times \{t\}$, for all $B_t \in \mathcal{F}_t$ and $t = 1, \ldots, T$. Then representation (7.4), for some set $\mathfrak{A}$, can be derived from axioms of coherent risk measures of Artzner et al. [1]. In our setting, these axioms correspond to Assumptions (A1)–(A3) for the trivial sigma algebra $\mathcal{F}_1 = \{\Omega', \emptyset\}$, and to the positive homogeneity of the (unconditional) risk function $\rho(X)$. Our approach is via axioms of conditional risk mappings, which allows for a specific analysis of the structure of the set $\mathfrak{A}$. This connects the theory of dynamic risk measures with the concept of conditional probability, which is crucial for the development of dynamic programming equations.

In applications, we frequently deal with random outcomes $X_t \in \mathcal{X}_t$ resulting from decisions z_t in some stochastic system. In order to model this situation, we introduce linear spaces $\mathcal{Z}_t$ of $\mathcal{F}_t$-measurable functions[2] $Z_t: \Omega \to \mathbb{R}^{n_t}$ and consider functions $f_t: \mathbb{R}^{n_t} \times \Omega \to \mathbb{R}$, $t = 1, \ldots, T$. With functions f_t we associate mappings $F_t: \mathcal{Z}_t \to \mathcal{X}_t$ defined as follows

$$[F_t(Z_t)](\omega) := f_t(Z_t(\omega), \omega), \quad Z_t \in \mathcal{Z}_t, \quad \omega \in \Omega.$$

We assume that the functions $f_t(z_t, \omega)$ are *random lower semicontinuous*,[3] and that the mappings F_t are well defined, i.e., for every $Z_t \in \mathcal{Z}_t$, the function $f_t(Z_t(\cdot), \cdot)$ belongs to the space $\mathcal{X}_t$, $t = 1, \ldots, T$. We say that the mapping F_t is convex if $[F_t(\cdot)](\omega)$ is convex for all $\omega \in \Omega$. Then for every conditional risk mapping $\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}$ satisfying (A1)–(A3), the function $\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}(F_t(\cdot))$ is convex in the sense that the function $[\rho_{\mathcal{X}_t|\mathcal{X}_{t-1}}(F_t(\cdot))](\omega)$ is convex for every $\omega \in \Omega$. This follows by Assumptions (A1) and (A2) and can be shown in the same way as Ruszczyński and Shapiro [17, Proposition 2].

Let $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots \times \mathcal{Z}_T$, and let $F: \mathcal{Z} \to \mathcal{X}$ be defined as

$$F(Z) := (F_1(Z_1), \ldots, F_T(Z_T)).$$

[2] Note that because $\mathcal{F}_1$ is trivial, the space $\mathcal{Z}_1$ coincides with $\mathbb{R}^{n_1}$ and elements $Z_1 \in \mathcal{Z}_1$ are n_1-dimensional vectors.

[3] Random lower-semicontinuous functions are also called *normal integrands* (see Definition 14.27 in Rockafellar and Wets [15, p. 676]).

With the risk function $\tilde{\rho}$, defined in (7.1), and the mapping F, we can associate the function

$$\vartheta(Z) := \tilde{\rho}(F(Z)) = F_1(Z_1) + \rho_{\mathscr{X}_2|\mathscr{X}_1}\big[F_2(Z_2) + \rho_{\mathscr{X}_3|\mathscr{X}_2}(F_3(Z_3) + \cdots + \rho_{\mathscr{X}_{T-1}|\mathscr{X}_{T-2}}[F_{T-1}(Z_{T-1}) + \rho_{\mathscr{X}_T|\mathscr{X}_{T-1}}(F_T(Z_T))])\big].$$

As discussed above, by the recursive application of Ruszczyński and Shapiro [17, Proposition 2], it can be easily shown that $\vartheta(\cdot)$ is a convex function. Also, by using (7.2) and (7.4) we can write

$$\vartheta(Z) = \rho_T(F_1(Z_1) + F_2(Z_2) + \cdots + F_T(Z_T)) = \sup_{\mu \in \mathscr{A}} \int_\Omega [f_1(Z_1) + f_2(Z_2(\omega), \omega) \cdots + f_T(Z_T(\omega), \omega)]\, d\mu(\omega).$$

Suppose that we are given $\mathscr{F}_t$-measurable, closed-valued multifunctions

$$\mathscr{G}_t\colon \mathbb{R}^{n_{t-1}} \times \Omega \rightrightarrows \mathbb{R}^{n_t}, \quad t = 2, \ldots, T,$$

with $\mathscr{G}_1 \subset \mathbb{R}^{n_1}$ being a fixed (deterministic) set. We define the set

$$\mathfrak{S} := \{Z \in \mathscr{Z}\colon Z_t(\omega) \in \mathscr{G}_t(Z_{t-1}(\omega), \omega),\ \omega \in \Omega,\ t = 1, \ldots, T\},$$

and consider the problem

$$\min_{Z \in \mathfrak{S}} \vartheta(Z). \tag{7.5}$$

We refer to problem (7.5) as the *nested* formulation of a multistage optimization problem. We shall derive dynamic programming equations for this problem.

In order to accomplish that, we need some mild technical assumptions. We assume that the spaces $\mathscr{X}_t$ are *solid* in the sense that for every two elements $\underline{X}, \overline{X} \in \mathscr{X}_t$ and every $\mathscr{F}_t$-measurable function X_t satisfying $\underline{X}(\cdot) \le X_t(\cdot) \le \overline{X}(\cdot)$, the function X_t is an element of $\mathscr{X}_t$. For example, the spaces $\mathscr{L}_p(\Omega, \mathscr{F}_t, P)$, $p \in [1, +\infty]$ are solid. Furthermore, we assume that there exist elements $\underline{X}_t \in \mathscr{X}_t$ such that for all $Z \in \mathfrak{S}$ we have $F_t(Z_t) \succeq \underline{X}_t$, $t = 1, \ldots, T$.

Remark 7.1. We also need the following result about interchageability of the "min" and "integral" operators. Let $(\Omega, \mathscr{F})$ be a measurable space, $\mathscr{X}$ be a linear space of $\mathscr{F}$-measurable functions $X\colon \Omega \to \mathbb{R}$, and $\mathscr{M}$ be a linear space of $\mathscr{F}$-measurable functions $Z\colon \Omega \to \mathbb{R}^n$. It is said that the space $\mathscr{M}$ is *decomposable* if for every $Z \in \mathscr{M}$ and $B \in \mathscr{F}$, and every bounded and $\mathscr{F}$-measurable function $W\colon \Omega \to \mathbb{R}^n$, the space $\mathscr{M}$ also contains the function $V(\cdot) := \mathbb{1}_{\Omega \setminus B}(\cdot) Z(\cdot) + \mathbb{1}_B(\cdot) W(\cdot)$ (Castaing and Valadier [4, p. 197], Rockafellar and Wets [15, p. 676]). For example, the spaces $\mathscr{L}_p(\Omega, \mathscr{F}, P; \mathbb{R}^n)$, $p \in [1, +\infty)$, of $\mathscr{F}$-measurable functions $Z\colon \Omega \to \mathbb{R}^n$ such that $\int_\Omega \|Z\|^p\, dP < +\infty$, are decomposable.

Let the space $\mathscr{M}$ be decomposable and $h\colon \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ be a random lower-semicontinuous function. Then for every probability measure μ on $(\Omega, \mathscr{F})$ the following interchangeability formula holds:

$$\int_\Omega \inf_{z \in \mathbb{R}^n} h(z, \omega)\, d\mu(\omega) = \inf_{Z \in \mathscr{M}} \int_\Omega h(Z(\omega), \omega)\, d\mu(\omega) \tag{7.6}$$

(Rockafellar and Wets [15, Theorem 14.60]). Now let $\rho\colon \mathscr{X} \to \overline{\mathbb{R}}$ be a risk function. By using monotonicity of ρ, it is possible to extend the interchangeability formula (7.6) to risk functions as follows (cf., Ruszczyński and Shapiro [17, Theorem 4]):

$$\rho\Big(\inf_{z \in \mathbb{R}^n} h(z, \cdot)\Big) = \inf_{Z \in \mathscr{M}} \rho(H_Z), \tag{7.7}$$

where $H_Z(\omega) := h(Z(\omega), \omega)$.

Let us go back to the multistage problem (7.5). We assume that the spaces $\mathscr{Z}_t$, $t = 1, \ldots, T$, are decomposable. Problem (7.5) can be written in a more explicit form as follows:

$$\min_{Z_1 \in \mathscr{G}_1} \min_{Z_2(\cdot) \in \mathscr{G}_2(Z_1, \cdot)} \cdots \min_{Z_T(\cdot) \in \mathscr{G}_T(Z_{T-1}(\cdot), \cdot)} \rho_T[F_1(Z_1) + F_2(Z_2) + \cdots + F_T(Z_T)]. \tag{7.8}$$

Consider the minimization with respect to Z_T in the above problem. Because the function ρ_T is a risk function, and in particular is monotone in the sense of (A2), and Z_T is required to be only $\mathscr{F}_T$-measurable, the interchangeability formula (7.7) allows us to carry out this minimization inside the argument of ρ_T. We obtain the following equivalent formulation of (7.8):

$$\min_{Z_1 \in \mathscr{G}_1} \min_{Z_2(\cdot) \in \mathscr{G}_2(Z_1, \cdot)} \cdots \min_{Z_{T-1}(\cdot) \in \mathscr{G}_{T-1}(Z_{T-2}(\cdot), \cdot)} \rho_T\Big[F_1(Z_1) + F_2(Z_2) + \cdots + F_{T-1}(Z_{T-1}) + \inf_{z_T \in \mathscr{G}_T(Z_{T-1}(\cdot), \cdot)} f_T(z_T, \cdot)\Big].$$

Owing to $\mathcal{F}_t$-measurability of $\mathcal{G}_T$ and random lower semicontinuity of $f_T(\cdot,\cdot)$, the pointwise infimum $\inf_{z_T\in\mathcal{G}_T(Z_{T-1}(\omega),\,\omega)} f_T(z_T,\omega)$ is $\mathcal{F}_T$-measurable (e.g., Rockafellar and Wets [15, Theorem 14.37]). By the assumption that $\mathcal{X}_T$ is solid, this infimum (as a function of ω) is an element of $\mathcal{X}_T$.

Using the fact that

$$\rho_t := \rho_{t-1}\circ\rho_{\mathcal{X}_t\mid\mathcal{X}_{t-1}}, \quad t=2,\ldots,T,$$

we can rewrite the last problem as follows:

$$\min_{Z_1\in\mathcal{G}_1}\min_{Z_2(\cdot)\in\mathcal{G}_2(Z_1,\cdot)}\cdots\min_{Z_{T-1}(\cdot)\in\mathcal{G}_{T-1}(Z_{T-2}(\cdot),\,\cdot)}\rho_{T-1}\Big[F_1(Z_1)+F_2(Z_2)+\cdots+F_{T-1}(Z_{T-1}) + \rho_{\mathcal{X}_T\mid\mathcal{X}_{T-1}}\Big(\inf_{z_T\in\mathcal{G}_T(Z_{T-1}(\cdot),\,\cdot)} f_T(z_T,\cdot)\Big)\Big]. \tag{7.9}$$

Our argument can be now repeated for $T-1, T-2,\ldots,1$. In order to simplify the notation, we define the following function

$$Q_T(z_{T-1},\omega) := [\rho_{\mathcal{X}_T\mid\mathcal{X}_{T-1}}(V_T(z_{T-1}))](\omega), \tag{7.10}$$

where

$$[V_T(z_{T-1})](\omega) := \inf_{z_T\in\mathcal{G}_T(z_{T-1},\,\omega)} f_T(z_T,\omega). \tag{7.11}$$

Repeating our analysis for $t=T-1,\ldots,2$, we move the minimization with respect to Z_t inside the argument of ρ_t. We define

$$Q_t(z_{t-1},\omega) := [\rho_{\mathcal{X}_t\mid\mathcal{X}_{t-1}}(V_t(z_{t-1}))](\omega), \tag{7.12}$$

where

$$[V_t(z_{t-1})](\omega) := \inf_{z_t\in\mathcal{G}_t(z_{t-1},\,\omega)}\{f_t(z_t,\omega)+Q_{t+1}(z_t,\omega)\}. \tag{7.13}$$

Of course, Equations (7.12) and (7.13) for t can be combined into one equation:

$$[V_t(z_{t-1})](\omega) = \inf_{z_t\in\mathcal{G}_t(z_{t-1},\,\omega)}\{f_t(z_t,\omega)+[\rho_{\mathcal{X}_{t+1}\mid\mathcal{X}_t}(V_{t+1}(z_t))](\omega)\}. \tag{7.14}$$

Finally, at the first stage we solve the problem

$$\min_{z_1\in\mathcal{G}_1} Q_2(z_1), \tag{7.15}$$

where $Q_2(z_1):=\rho_{\mathcal{X}_2\mid\mathcal{X}_1}(V_2(z_1))$. Note again that the set $\mathcal{G}_1$ and function $Q_2(z_1)$ are deterministic, i.e., independent of ω.

We can interpret functions $Q_t(z_{t-1},\omega)$ as *cost-to-go functions* and Equations (7.12)–(7.13), or equivalently (7.14), as *dynamic programming equations* for the multistage risk optimization problem (7.5).

Suppose now that the conditional risk mappings $\rho_{\mathcal{X}_t\mid\mathcal{X}_{t-1}}$ are lower semicontinuous and positively homogeneous. Then it follows from (3.5) that there exist convex closed sets $\mathcal{A}_t(\omega)\subset\mathcal{P}_{\mathcal{Y}_t\mid\mathcal{F}_{t-1}}(\omega)$ such that

$$[\rho_{\mathcal{X}_t\mid\mathcal{X}_{t-1}}(X_t)](\omega) = \sup_{\mu_t\in\mathcal{A}_t(\omega)}\mathbb{E}_{\mu_t}[X_t], \quad \omega\in\Omega,\ \ X_t\in\mathcal{X}_t. \tag{7.16}$$

Substitution of (7.16) into (7.14) yields the following form of the dynamic programming equations:

$$[V_t(z_{t-1})](\omega) = \inf_{z_t\in\mathcal{G}_t(z_{t-1},\,\omega)}\Big\{f_t(z_t,\omega)+\sup_{\mu_t\in\mathcal{A}_t(\omega)}\mathbb{E}_{\mu_t}[V_{t+1}(z_t)].\Big\}. \tag{7.17}$$

The above dynamic programming equations provide a framework for extending the theory of multistage stochastic optimization problems to risk functions. The only difference is the existence of the additional "sup" operation with respect to a set of conditional probabilities. This makes the dynamic programming equations more difficult than in the expected value case, but the problem is much more difficult.

Remark 7.2. Suppose that the conditional risk mappings $\rho_{\mathcal{X}_t\mid\mathcal{X}_{t-1}}$ are lower semicontinuous and positively homogeneous, and that every space $\mathcal{X}_t$, $t=1,\ldots,T$, is separable. We have then by Theorem 4.1 that there exist (countable) families $\mathcal{D}_t\subset\mathcal{P}_{\mathcal{Y}_t}$, $t=1,\ldots,T$, of probability measures such that

$$[\rho_{\mathcal{X}_t\mid\mathcal{X}_{t-1}}(X_t)](\omega) = \sup_{\nu\in\mathcal{D}_t}\mathbb{E}_\nu[X_t\mid\mathcal{F}_{t-1}](\omega). \tag{7.18}$$

Note that Equation (7.18) still holds if the set $\mathcal{D}_t$ is replaced by $\mathcal{D}_t^* := \text{cl}[\text{conv}(\mathcal{D}_t)]$, where the topological closure of the convex hull of $\mathcal{D}_t$ is taken in the paired topology of the space $\mathcal{Y}_t$. Because the set $\mathcal{P}_{\mathcal{Y}_t}$ is convex and closed in $\mathcal{Y}_t$, we have that $\mathcal{D}_t^* \subset \mathcal{P}_{\mathcal{Y}_t}$. Substitution of (7.18) into (7.14), with $\mathcal{D}_t$ replaced by $\mathcal{D}_t^*$, yields the following form of the dynamic programming equations:

$$[V_t(z_{t-1})](\omega) = \inf_{z_t \in \mathcal{G}_t(z_{t-1},\,\omega)} \left\{ f_t(z_t, \omega) + \sup_{\nu \in \mathcal{D}_t^*} \mathbb{E}_\nu [V_{t+1}(z_t) \mid \mathcal{F}_{t-1}](\omega) \right\} \tag{7.19}$$

Suppose, further, that the problem is convex, i.e., the functions $f_t(\cdot, \omega)$, $[V_t(\cdot)](\omega)$, and sets $\mathcal{G}_t(z_{t-1}, \omega)$ are convex for all $\omega \in \Omega$ and z_{t-1}. Then, under various regularity conditions, the min-max problem in the right-hand side of (7.19) has a saddle point $(\bar{z}_t, \bar{\nu}_t) \in \mathcal{G}_t(z_{t-1}, \omega) \times \mathcal{D}_t^*$. It then follows that an optimal solution of problem (7.5) satisfies the following system of dynamic equations

$$[V_t(z_{t-1})](\omega) = \inf_{z_t \in \mathcal{G}_t(z_{t-1},\,\omega)} \left\{ f_t(z_t, \omega) + \mathbb{E}_{\bar{\nu}_t} [V_{t+1}(z_t) \mid \mathcal{F}_{t-1}](\omega) \right\}, \tag{7.20}$$

where $\bar{\nu}_t$, $t = 1, \ldots, T$, can be viewed as worst-case distributions. Moreover, for a given probability measure $\bar{\nu}_{t-1}$ on $(\Omega, \mathcal{F}_{t-1})$, we can construct $\mathcal{D}_t$ in such a way that every measure $\nu_t \in \mathcal{D}_t$, and hence every $\nu_t \in \mathcal{D}_t^*$, coincides with $\bar{\nu}_{t-1}$ on $\mathcal{F}_{t-1}$. In that way we can construct the worst-case distributions $\bar{\nu}_t$ in a consistent way, i.e., each $\bar{\nu}_t$ coincides with $\bar{\nu}_{t-1}$ on $\mathcal{F}_{t-1}$.

The dynamic programming equations simplify considerably if we assume the between-stages independence condition. Following the framework of Remark 5.2, let $\xi_1, \ldots, \xi_T$ be a sequence of random vectors representing evolution of the data. Suppose that each $X_t \in \mathcal{X}_t$ is a function of ξ_t, and the objective functions $f_t(z_t, \xi_t)$ and multifunctions $\mathcal{G}_t(z_{t-1}, \omega)$ are actually functions of ξ_t. With a slight abuse of notation we denote them by $\mathcal{G}_t(z_{t-1}, \xi_t)$ (it is also possible to consider dependence on all previous data $\xi_{[t]} = (\xi_1, \ldots, \xi_t)$). This implies that the cost-to-go function $Q_t(z_{t-1}, \xi_{t-1})$ is a function of z_{t-1} and ξ_{t-1}, and the value function $V_t(z_{t-1}, \xi_t) = [V_t(z_{t-1})](\xi_t)$ is a function of z_{t-1} and ξ_t (we change the notation in a corresponding way). Suppose that the between-stages independence condition holds, i.e., ξ_t and ξ_{t-1} are independent for $t = 2, \ldots, T$. Suppose, further, that $[\rho_{\mathcal{X}_t \mid \mathcal{X}_{t-1}}(X_t)](\cdot)$ is constant for any $X_t \in \mathcal{X}_t$, $t = 2, \ldots, T$ (its values are constant functions on Ω). Under the between-stages independence condition, this holds for the conditional mappings discussed in Examples 4.1–6.2. Then $\rho_{\mathcal{X}_t \mid \mathcal{X}_{t-1}}$ maps every element of $\mathcal{X}_t$ into a constant, and hence $\rho_{\mathcal{X}_t \mid \mathcal{X}_{t-1}} = \rho_t$, where ρ_t is the composite mapping defined in (7.3). Furthermore, the cost-to-go functions

$$Q_t(z_{t-1}) = \rho_t(V_t(z_{t-1})) \tag{7.21}$$

are deterministic (independent of ξ_{t-1}) and Equations (7.14) take on the form

$$V_t(z_{t-1}, \xi_t) = \inf_{z_t \in \mathcal{G}_t(z_{t-1},\,\xi_t)} \{ f_t(z_t, \xi_t) + \rho_{t+1}(V_{t+1}(z_t)) \}. \tag{7.22}$$

For illustration, let $\rho_{\mathcal{X}_t \mid \mathcal{X}_{t-1}}$ be the mean–absolute-semideviation risk mapping of Example 6.1 (with $p = 1$). Then the corresponding cost-to-go function, defined in (7.12), can be written as

$$Q_t(z_{t-1}, \xi_{t-1}) = \mathbb{E}[V_t(z_{t-1}, \xi_t) \mid \xi_{t-1}] + c\mathbb{E}[(V_t(z_{t-1}, \xi_t) - \mathbb{E}[V_t(z_{t-1}, \xi_t) \mid \xi_{t-1}])_+ \mid \xi_{t-1}].$$

If, moreover, ξ_t and ξ_{t-1} are independent, then

$$Q_t(z_{t-1}) = \mathbb{E}[V_t(z_{t-1}, \xi_t)] + c\mathbb{E}[(V_t(z_{t-1}, \xi_t) - \mathbb{E}[V_t(z_{t-1}, \xi_t)])_+].$$

Then the dynamic programming Equations (7.22) become more transparent: At each stage we minimize the sum of the current cost, $f_t(z_t, \xi_t)$ and the (static) mean-semideviation risk function of the next value function $V_{t+1}(z_t)$.

Example 7.1. Consider the financial planning model governed by the equations

$$\sum_{i=1}^n z_{it} = W_t \quad \text{and} \quad \sum_{i=1}^n \xi_{i,\,t+1} z_{it} = W_{t+1}, \qquad t = 0, \ldots, T-1,$$

where all $z_{it} \geq 0$. Here W_t denotes the wealth at stage t, and z_{it} are the positions in assets $i = 1, \ldots, n$ at stages $t = 0, \ldots, T-1$. The random multipliers $\xi_{i,\,t+1}$ represent the changes in the investment values between stages t and $t+1$. They are assumed to be nonnegative.

Suppose that the between-stages independence condition holds for the random process $\xi_1, \ldots, \xi_T$, and at every period t we want to maximize $\rho_t[W_t]$, where ρ_t is a positively homogeneous risk function. For example, it may be the mean-semideviation or CV@R function. At the last stage, the value function $V_T(W_{T-1})$ is the optimal value of the problem

$$\begin{aligned} \max_{W_T, z_{T-1}} \quad & \rho_T(W_T) \\ \text{s.t.} \quad & W_T = \sum_{i=1}^n \xi_{iT} z_{i,T-1}, \\ & \sum_{i=1}^n z_{i,T-1} = W_{T-1}, \\ & z_{i,T-1} \geq 0, \quad i = 1, \ldots, n. \end{aligned} \tag{7.23}$$

The wealth at the preceding stage, W_{T-1}, is the parameter of this problem. Because ρ_T is positively homogeneous, we see that the optimal value is simply proportional to the wealth:

$$V_T(W_{T-1}) = W_{T-1} V_T(1),$$

where $V_T(1)$ is the (nonnegative) optimal value of (7.23) for $W_{T-1} = 1$. We can use this fact at stage $T-2$. Because ρ_{T-1} is positively homogeneous and its argument, $V_T(W_{T-1})$, is linear, by a similar argument we conclude that

$$V_{T-1}(W_{T-2}) = V_T(1) V_{T-1}(1) W_{T-2},$$

where $V_{T-1}(1)$ is the optimal value of the problem obtained from (7.23) by replacing T with $T-1$. Continuing in this way, we conclude that the optimal solution of the first-stage problem is obtained by solving a problem of the form (7.23) with $T = 1$. That is, under the assumption of between-stages independence, the optimal policy is *myopic* and employs single-stage risk models.

Acknowledgments. The research of the first author was supported by the NSF awards DMS-0303545 and DMI-0354678. The research of the second author was supported by the NSF award DMS-0510324. The authors are indebted to Darinka Dentcheva for helpful discussions regarding weakly measurable selections of multifunctions.

References

[1] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–228.
[2] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath, H. Ku. 2003. Coherent multiperiod risk measurement. Manuscript, ETH Zürich, Zurich, Switzerland.
[3] Billingsley, P. 1995. *Probability and Measure*. Wiley, New York.
[4] Castaing, C., M. Valadier. 1977. *Convex Analysis and Measurable Multifunctions*. Springer-Verlag, Berlin, Germany.
[5] Cheridito, P., F. Delbaen, M. Kupper. 2004. Coherent and convex risk measures for bounded càdlàg processes. *Stochastic Processes Appl.* **112**(1) 1–22.
[6] Delbaen, F. 2002. Coherent risk measures on general probability spaces. *Essays in Honour of Dieter Sondermann*. Springer-Verlag, Berlin, Germany, 1–3.
[7] Föllmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance Stochastics* **6** 429–447.
[8] Kijima, M., M. Ohnishi. 1993. Mean-risk analysis of risk aversion and wealth effects on optimal portfolios with multiple investment opportunities. *Ann. Oper. Res.* **45** 147–163.
[9] Kuratowski, K., C. Ryll-Nardzewski. 1965. A general theorem on selectors. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **13** 397–403.
[10] Ogryczak, W., A. Ruszczyński. 1999. From stochastic dominance to mean-risk models: Semideviations as risk measures. *Eur. J. Oper. Res.* **116** 33–50.
[11] Ogryczak, W., A. Ruszczyński. 2001. On consistency of stochastic dominance and mean–semideviation models. *Math. Programming* **89** 217–232.
[12] Pflug, G., A. Ruszczyński. 2004. A risk measure for income processes. G. Szegö, ed. *Risk Measures for the 21st Century*. John Wiley & Sons, Chichester, UK.
[13] Riedel, F. 2004. Dynamic coherent risk measures. *Stochastic Processes Appl.* **112** 185–200.
[14] Rockafellar, R. T., S. P. Uryasev. 2000. Optimization of conditional value-at-risk. *J. Risk* **2** 21–41.
[15] Rockafellar, R. T., R. J-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, Berlin, Germany.
[16] Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2000b. Generalized deviations in risk analysis. *Stochastic Finance* **10** 51–74.
[17] Ruszczyński, A., A. Shapiro. Optimization of convex risk functions. *Math. Oper. Res.* **31**(3) 433–452.
[18] Shapiro, A., S. Ahmed. 2004. On a class of minimax stochastic programs. *SIAM J. Optim.* **14** 1237–1249.