Conditional Distributionally Robust Functionals

October 9, 2022

Alexander Shapiro
School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia 30332-0205,
e-mail: ashapiro@isye.gatech.edu
Research of this author was partly supported by AFOSR grant FA9550-22-1-0244.

Alois Pichler
University of Technology, Chemnitz,
Faculty of Mathematics
e-mail: alois.pichler@math.tu-chemnitz.de
Funded by Deutsche Forschungsgemeinschaft
(DFG, German Research Foundation)
Project-ID 416228727 – SFB 1410

Abstract

Many decisions, in particular decisions in a managerial context, are subject to uncertainty. Risk measures cope with uncertainty by involving more than one candidate probability. The corresponding risk averse decision takes all potential, candidate probabilities into account and is robust with respect to all potential probabilities. This paper considers conditional robust decision-making, where decisions are subject to additional, prior knowledge or information. The literature discusses various definitions to characterize the corresponding conditional risk measure, which determines further the decision. The aim of this paper is to compare two different approaches to construction of conditional functionals employed in multistage distributionally robust optimization. As an application we discuss conditional counterparts of a distance between probability measures.

Key words: distributional robustness conditional risk measures strict monotonicity Wasserstein distance rectangularity

1 Introduction

Numerous real-world problems in sciences and economics need to build on inaccurate measurements, on uncertain, deviating observations or on uncertainty, which is inherent to the problem at hand. In this situation, summary functionals are used to aggregate and quantify the varying outcomes. The standard example of such summary functional is the expectation.

The expectation balances favorable and non-favorable outcomes. Risk-averse decision makers, in contrast, distinguish satisfactory and unpleasing outcomes. They thus prefer alternative schemes to aggregate the potential outcomes. In financial economics, for example, the Value-at-Risk is one of these functionals, which is well-established and accepted in assessing the random behavior of market returns. Many of these risk assessments allow for an additional interpretation by presuming that the underlying probabilistic model is not known itself. The terms *ambiguity*, *distributional robustness* and *Knightian uncertainty* are associated with this type of uncertainty; the latter term was coined by Knight [10], cf. also Ellsberg [5].

In sequential decision-making (i.e., decision-making in multiple stages or multistage optimization), the assessment of risk involves the history, i.e., the realization already observed. This assessment of risk naturally leads to the composition of risk measures and many related questions such as time and dynamic consistency.[1] In this situation, differing assessments of the risk, conditional on the observed past, are possible. There are two somewhat natural ways for extending risk measures - distributionally robust approaches to the sequential (multistage) setting which were suggested in the literature. One is based on a properly defined maximum of conditional expectations (e.g., Föllmer and Schied [7, p. 306], [21]), while the other employs an approach to construction of conditional coherent risk measures (cf. Riedel [18], Ruszczyński and Shapiro [19]). It was pointed out in Pichler and Shapiro [16] that these two approaches are different from each other.

Distributional Robust Optimization (DRO) is a paradigm, where the distribution governing the data is subject to uncertainty itself. Its origins can be traced to Scarf [20]. For a recent discussion of DRO we can refer, e.g., to Wiesemann et al. [26] and Hanasusanto et al. [8]. DRO often assumes existence of an explicit reference probability measure. Such a measure is, a priori, not present in many discrete settings, as is the case with empirical measures. Convergence is a further aspect, which is studied for distributinoally robust optimization problems, cf., Sun and Xu [24], Liu et al. [11]. We refer to the reviews Keith and Ahner [9], Rahimian and Mehrotra [17] for a summary on these perspectives.

This paper is aimed at discussion of the two approaches to construction of conditional distributionally robust functionals and connects them to various concepts of conditional distances between probability measures. Main contributions of this paper can be summarized as follows. We investigate a connection between the conditional expectation counterpart of the distributionally robust functionals and the related conditional risk measures. An initial analysis of this issue was already performed in Pichler and Shapiro [16], where it was discussed mainly from the DRO point of view. Second we relate this analysis to conditional counterparts of a distance between probability measures. In particular, we show how such conditional distances can be used for construction of the corresponding conditional functionals. This can be viewed as an extension of investiga-

[1]Different concepts of time consistency were introduced in the literature over the years, it is beyond the scope of this paper to give a fair review of this subject.

tion of conditional distances between stochastic processes initiated in Pflug [13], Pflug and Pichler [14] (see also Esteban-Pérez and Morales [6]).

For the sake of simplicity we mainly consider the setting where the underlying measurable space is finite. This allows to concentrate on some basic conceptual questions while avoiding delicate questions of measurability. An extension to a general setting is outlined in Section 3.

We use the following terminology and notation throughout the paper. The extended real line $\mathbb{R}\cup\{-\infty\}\cup\{+\infty\}$ is denoted $\overline{\mathbb{R}}$. By $(\Omega,\mathcal{F})$ we denote a measurable space, i.e., $\mathcal{F}$ is a sigma algebra of subsets of Ω. By $\mathfrak{P}$ is denoted the set of all probability measures on the measurable space $(\Omega,\mathcal{F})$. We use the notation $Q \ll P$ to denote that $Q \in \mathfrak{P}$ is absolutely continuous with respect to $P \in \mathfrak{P}$. By the Radon–Nikodym theorem we have that $Q \ll P$ iff there exists a measurable function (density function) $f\colon \Omega \to \mathbb{R}_+$, denoted $f = dQ/dP$, such that $Q(A) = \int_A f dP$ for every $A \in \mathcal{F}$. For A, $B \in \mathcal{F}$ we denote by $P(A|B) = P(A\cap B)/P(B)$ the conditional probability (assuming $P(B) \neq 0$). A family of sets $\{\varpi_i\}_{i\in\mathcal{I}}$ is a *partition* of Ω if each $\varpi_i \in \mathcal{F}$ is nonempty such that $\Omega = \cup_{i\in\mathcal{I}}\varpi_i$ and $\varpi_i \cap \varpi_j = \emptyset$ for $i \neq j \in \mathcal{I}$. It is said that the partition is finite if the index set $\mathcal{I}$ is finite. By $\mathcal{Z}$ we denote an appropriate linear space of measurable variables $Z\colon \Omega \to \mathbb{R}$. If $\Omega = \{\omega_1,\dots,\omega_n\}$ is finite, we assume that $\mathcal{F}$ consists of all subsets of Ω. In that case any $Z\colon \Omega \to \mathbb{R}$ is measurable, and the space $\mathcal{Z}$ consists of all $Z\colon \Omega \to \mathbb{R}$. Also, for any $P \in \mathfrak{P}$ the corresponding probability vector p, with components $p_i = p(\omega_i)$, $i = 1,\dots,n$, is an element of Δ_n, where $p(\omega) = P(\{\omega\})$, $\omega \in \Omega$, and $\Delta_n \subset \mathbb{R}^n$ denotes the n-dimensional simplex, i.e., $\Delta_n \coloneqq \big\{p \in \mathbb{R}^n_+ \colon \sum_{i=1}^n p_i = 1\big\}$. For a process $\xi_1,\xi_2,\dots$ we denote by $\xi_{[t]} = (\xi_1,\dots,\xi_t)$ its history up to time t.

2 General analysis

Let $\mathfrak{M} \subset \mathfrak{P}$ be a (nonempty) set of probability measures. The distributionally robust functional $\mathcal{R}\colon \mathcal{Z} \to \mathbb{R}$ associated with $\mathfrak{M}$ is defined as

$$\mathcal{R}(Z) \coloneqq \sup_{Q\in\mathfrak{M}} \Big\{ \mathbb{E}_Q[Z] = \int_\Omega Z(\omega)dQ(\omega)\Big\}. \tag{2.1}$$

In the literature on distributional robustness the set $\mathfrak{M}$ is often referred to as the *ambiguity set*. We assume that for every $Q \in \mathfrak{M}$ the expectation (integral) $\mathbb{E}_Q[Z]$ is well-defined and $\mathcal{R}(Z)$ is finite valued for all $Z \in \mathcal{Z}$, where $\mathcal{Z}$ is an appropriate linear space of measurable variables $Z\colon \Omega \to \mathbb{R}$. Given a sigma subalgebra $\mathcal{G}$ of $\mathcal{F}$ we would like to construct a conditional counterpart of $\mathcal{R}$.

There are two somewhat natural ways to define conditional counterparts of distributionally robust functionals, namely the distributionally robust and the risk averse approaches. In both approaches, the respective conditional functional is a $\mathcal{G}$-measurable variable. The distributionally robust approach is to consider the conditional expectation counterpart of the distributionally robust functional:

$$\mathcal{R}^\star_{|\mathcal{G}}(Z) \coloneqq \sup_{Q\in\mathfrak{M}} \mathbb{E}_{Q|\mathcal{G}}[Z]. \tag{2.2}$$

This approach was used by several authors (e.g., Föllmer and Schied [7, p. 306], Shapiro [21]). Recall that the conditional expectation $\mathbb{E}_{Q|\mathcal{G}}[Z]$ is a $\mathcal{G}$-measurable random variable, and that any two versions of $\mathbb{E}_{Q|\mathcal{G}}[Z]$ can be different from each other on a set Q-measure zero. Because of that there are technical issues in defining a precise meaning of the right-hand side of (2.2), we will discuss this later.

Of course, if the set $\mathfrak{M} = \{P\}$ is a singleton, then by the definition we have that $\mathcal{R}^\star_{|\mathcal{G}}(\cdot) = \mathbb{E}_{P|\mathcal{G}}(\cdot)$. In general,j $\mathcal{R}^\star_{|\mathcal{G}}$ can be viewed as a somewhat natural counterpart of the distributionally robust functional $\mathcal{R}$. However, it turns out that such definition of conditional functionals is different from the one used in the theory of risk measures (this was already pointed out in Pichler and Shapiro [16]). In order to give an intuitive insight, while avoiding delicate technical issues, we assume in the remainder of this section that the sigma algebra $\mathcal{G}$ is finite. The general case will be considered in Section 3.

Finite sigma subalgebra. Assume that $\mathcal{G}$ is *finite*. Since $\mathcal{G}$ is finite, it is generated by a finite partition $\{\varpi_i\}_{i\in\mathcal{I}}$ of the set Ω such that $Z\colon \Omega \to \mathbb{R}$ is $\mathcal{G}$-measurable iff $Z(\cdot)$ is constant on every set ϖ_i, $i \in \mathcal{I}$. The sets ϖ_i, $i \in \mathcal{I}$, are atoms (elementary events) of $\mathcal{G}$. For a $\mathcal{G}$-measurable Z we denote by

$$Z(\varpi_i) \tag{2.3}$$

its value on ϖ_i, which does not depend on the particular choice of $\omega \in \varpi_i$.

Consider a probability measure (distribution) $P \in \mathfrak{P}$. The conditional probability $P(\cdot\,|\,\varpi_i)$, conditional on the event ϖ_i, assuming that $P(\varpi_i) \neq 0$, is

$$P_{|\varpi_i}(A) := P(A|\varpi_i) = \frac{P(A \cap \varpi_i)}{P(\varpi_i)}, \quad A \in \mathcal{F}. \tag{2.4}$$

Note that $P_{|\varpi_i}$ is supported on the set ϖ_i. The conditional expectation of a random variable $Z\colon \Omega \to \mathbb{R}$ is a $\mathcal{G}$-measurable variable $\mathbb{E}_{P|\mathcal{G}}[Z]\colon \Omega \to \mathbb{R}$ taking a constant value on each elementary event ϖ_i, $i \in \mathcal{I}$, of $\mathcal{G}$ with

$$\mathbb{E}_{P|\mathcal{G}}[Z](\varpi_i) = \frac{1}{P(\varpi_i)} \int_{\varpi_i} Z(\omega)dP(\omega) = \int_{\varpi_i} Z(\omega)dP_{|\varpi_i}(\omega), \text{ for } P(\varpi_i) > 0. \tag{2.5}$$

If $P(\varpi_i) = 0$, then $\mathbb{E}_{P|\mathcal{G}}[Z]$ can be arbitrary on ϖ_i. That is, different versions of $\mathbb{E}_{P|\mathcal{G}}[Z]$ do coincide on such ϖ_i that $P(\varpi_i) > 0$, and can be arbitrary on ϖ_i when $P(\varpi_i) = 0$. Also, let us observe that the value of $\mathbb{E}_{P|\mathcal{G}}[Z](\cdot)$ on ϖ_i does not depend on $Z(\omega)$ for $\omega \notin \varpi_i$.

Assume that for every $i \in \mathcal{I}$ there exists $Q \in \mathfrak{M}$ such that $Q(\varpi_i) \neq 0$. Then the conditional distributionally robust functional is defined for $Z \in \mathcal{Z}$, as a $\mathcal{G}$-measurable variable $\mathcal{R}^\star_{|\mathcal{G}}(Z)\colon \Omega \to \mathbb{R}$, taking a constant value on each elementary event ϖ_i of $\mathcal{G}$, with

$$\mathcal{R}^\star_{|\mathcal{G}}(Z)(\varpi_i) = \sup_{Q\in\mathfrak{M},\, Q(\varpi_i)\neq 0} \Big\{ \underbrace{\frac{1}{Q(\varpi_i)} \int_{\varpi_i} Z(\omega)dQ(\omega)}_{\mathbb{E}_{Q|\mathcal{G}}[Z](\varpi_i)} \Big\}, \quad i \in \mathcal{I}. \tag{2.6}$$

Recall that if $Q(\varpi_i) = 0$, then $\mathbb{E}_{Q|\mathcal{G}}[Z](\varpi_i)$ can be arbitrary. Therefore, if $Q(\varpi_i) = 0$ for all $Q \in \mathfrak{M}$, then $\mathcal{R}^{\star}_{|\mathcal{G}}(Z)(\varpi_i)$ is not clearly defined, we will try to avoid such cases. For this reason, we make the following assumption.

Assumption 2.1. *For every $i \in \mathcal{I}$ there exists $Q \in \mathfrak{M}$ such that $Q(\varpi_i) \neq 0$.*

Remark 2.1. In case the set $\Omega = \{\omega_1, \dots, \omega_n\}$ is finite, each $Q \in \mathfrak{M}$ can be identified with the respective probability vector $q \in \Delta_n$, having components $q_i = Q(\{\omega_i\})$, $i = 1, \dots, n$, and the set $\mathfrak{M}$ can be considered as a subset of the simplex Δ_n. Without loss of generality we can assume that the set $\mathfrak{M} \subset \Delta_n$ is convex and closed. In terms of convex analysis, the distributionally robust functional $\mathcal{R}$ can be viewed as the support function of the set $\mathfrak{M} \subset \Delta_n$. Therefore, we have that if $\mathfrak{M}$, $\mathfrak{M}' \subset \Delta_n$ are closed convex sets, then the corresponding distributionally robust functionals do coincide iff $\mathfrak{M} = \mathfrak{M}'$. This can be extended to infinite dimensional settings with an appropriate definition of topology on the space of measures. $\square$

An alternative approach, of conditional risk measures, is the following. For $i \in \mathcal{I}$ consider a (nonempty) set $\mathfrak{A}_i$ of probability measures on $(\Omega, \mathcal{F})$ supported on the set ϖ_i. For $Z \in \mathcal{Z}$ consider the function $\mathcal{R}_{|\mathcal{G}}(Z) \colon \Omega \to \mathbb{R}$ taking a constant value on each elementary event ϖ_i of $\mathcal{G}$ with

$$\mathcal{R}_{|\mathcal{G}}(Z)(\varpi_i) \coloneqq \sup_{Q \in \mathfrak{A}_i} \mathbb{E}_Q[Z], \quad i \in \mathcal{I}. \tag{2.7}$$

Note that each set $\mathfrak{A}_i$ can be assumed to be convex. We can view $\mathcal{R}_{|\mathcal{G}}$ as a conditional risk measure with the right-hand side of (2.7) defining its value conditional on node $i \in \mathcal{I}$ at stage one.

For singleton sets $\mathfrak{A}_i \coloneqq \{P_{|\varpi_i}\}$, $i \in \mathcal{I}$, we have that $\mathcal{R}_{|\mathcal{G}}(\cdot) = \mathbb{E}_{P|\mathcal{G}}(\cdot)$. Recall that if the set $\mathfrak{M} = \{P\}$ is the singleton, then $\mathcal{R}^{\star}_{|\mathcal{G}}(\cdot) = \mathbb{E}_{P|\mathcal{G}}(\cdot)$ as well. For a general set $\mathfrak{M}$ we can use this for each $Q \in \mathfrak{M}$ in order to construct the respective sets $\mathfrak{A}_i$. That is, define

$$\mathfrak{A}_i \coloneqq \left\{ Q_{|\varpi_i} \,\middle|\, Q \in \mathfrak{M} \text{ and } Q(\varpi_i) \neq 0 \right\}, \quad i \in \mathcal{I}. \tag{2.8}$$

For sets $\mathfrak{A}_i$ defined in (2.8), $\mathcal{R}_{|\mathcal{G}}(\varpi_i)$ can be written as the supremum in the right-hand side of (2.6), that is for such sets $\mathcal{R}_{|\mathcal{G}}(\cdot) = \mathcal{R}^{\star}_{|\mathcal{G}}(\cdot)$. Conversely, consider conditional risk measure $\mathcal{R}_{|\mathcal{G}}$ determined by the respective (arbitrary, nonempty) sets $\mathfrak{A}_i$. Define

$$\mathfrak{M}_\nu \coloneqq \sum_{i \in \mathcal{I}} \nu_i \, \mathfrak{A}_i, \tag{2.9}$$

where $\nu_i > 0$ are positive probabilities, i.e., $\sum_{i \in \mathcal{I}} \nu_i = 1$. For such set $\mathfrak{M} = \mathfrak{M}_\nu$ we have that each $\mathfrak{A}_i$ coincides with the right-hand side of (2.8). That is, we have the following (cf. Shapiro et al. [23, Section 6.5.1]).

Proposition 2.1. *For any positive probabilities ν_i, $i \in \mathcal{I}$, and the conditional distributionally robust functional $\mathcal{R}^{\star}_{|\mathcal{G}}$ associated with the set $\mathfrak{M} = \mathfrak{M}_\nu$ of the form (2.9), we have that $\mathcal{R}_{|\mathcal{G}}(\cdot) = \mathcal{R}^{\star}_{|\mathcal{G}}(\cdot)$.*

Remark 2.2. The conditional functional $\mathcal{R}^\star_{|\mathcal{G}}$ is determined by the specified set $\mathfrak{M} \subset \mathfrak{P}$ of probability measures, while $\mathcal{R}_{|\mathcal{G}}$ is associated with the specified sets $\mathfrak{A}_i$. To the set $\mathfrak{M}$ correspond sets $\mathfrak{A}_i$ defined by formula (2.8). Conversely, with a family of sets $\mathfrak{A}_i$ we can associate the corresponding set $\mathfrak{M} = \mathfrak{M}_\nu$ by using (2.9) for any vector $\nu \in \Delta_n$ with positive components. Hence, there is a relation between constructions of $\mathcal{R}^\star_{|\mathcal{G}}$ and $\mathcal{R}_{|\mathcal{G}}$. The difference between the distributionally robust approach (2.2) and the risk averse approach (2.7) to construction of conditional functionals is a way how the respective ambiguity sets are constructed. $\square$

Remark 2.3. As in Remark 2.1, we have that the conditional risk measure $\mathcal{R}_{|\mathcal{G}}$ is uniquely determined by the sets $\mathfrak{A}_i$. That is, if $\mathfrak{A}_i$ and $\mathfrak{A}'_i$ are two families of closed convex sets of probability measures supported on ϖ_i, then the respective conditional risk measures $\mathcal{R}_{|\mathcal{G}}$ do coincide iff $\mathfrak{A}_i = \mathfrak{A}'_i$, $i \in \mathcal{I}$. On the other hand, it can happen that two *different* closed, convex set $\mathfrak{M}$, $\mathfrak{M}' \subset \mathfrak{P}$ define the same sets $\mathfrak{A}_i$ via (2.8), and hence the same distributionally robust functional associated with the corresponding ambiguity set $\mathfrak{M}_\nu$ (see Example 3.2 below). $\square$

In both constructions the corresponding conditional functionals $\mathcal{R}^\star_{|\mathcal{G}}$ and $\mathcal{R}_{|\mathcal{G}}$ satisfy the axioms of coherent risk mappings, i.e., both are subadditive, monotone, translation equivariant and positively homogeneous. The following example from Pichler and Shapiro [16, Example 3.2] demonstrates their possible difference.

Example 2.1. Let Ω be finite and $P \in \mathfrak{P}$ be a (reference) probability measure with the respective probabilities $p(\omega) > 0$, $\omega \in \Omega$. For $\alpha \in (0,1)$ let $\mathfrak{M} \subset \mathfrak{P}$ be the set of probability measures Q such that $q(\omega)/p(\omega) \le \alpha^{-1}$ for $\omega \in \Omega$. This set $\mathfrak{M}$ is the dual set of the Average Value-at-Risk measure

$$\mathcal{R}(Z) := \mathsf{AV@R}_{\alpha,P}(Z) = \inf_{\tau\in\mathbb{R}} \left\{\tau + \alpha^{-1}\,\mathbb{E}_P[Z-\tau]_+\right\}. \tag{2.10}$$

Let $\mathcal{G}$ be a subalgebra of $\mathcal{F}$ defined by partition $\{\varpi_i\}_{1\le i\le m}$. Suppose that $\alpha \le 1-P(\varpi_i)$, $i = 1,\dots,m$. Then for any $i \in \mathcal{I}$, with $\mathcal{I} = \{1,\dots,m\}$, and $\bar\omega \in \varpi_i$, there exists $Q \in \mathfrak{M}$ such that $q(\bar\omega) > 0$ and $q(\omega) = 0$ for $\omega \in \varpi_i \setminus \{\bar\omega\}$. Indeed, take $q(\bar\omega) := p(\bar\omega)$, $q(\omega) = 0$ for $\omega \in \varpi_i \setminus \{\bar\omega\}$ and $q(\omega) := \kappa\, p(\omega)$ for $\omega \in \Omega \setminus \varpi_i$, where $\kappa := \frac{1-p(\bar\omega)}{1-P(\varpi_i)}$. Note that $p(\bar\omega) + \sum_{\omega\notin\varpi_i} \kappa\, p(\omega) = 1$ and $\kappa \le \alpha^{-1}$ since $\alpha \le 1-P(\varpi_i)$.

For such $Q \in \mathfrak{M}$ we have that $Q(\varpi_i) = q(\bar\omega)$ and the corresponding conditional probability $q(\bar\omega)/Q(\varpi_i) = 1$. It follows that for every $i \in \mathcal{I}$ the set $\mathfrak{A}_i$ of the form (2.8) consists of *all* probability measures supported on ϖ_i, and hence the conditional distributionally robust counterpart of the $\mathsf{AV@R}$ is

$$\mathsf{AV@R}^\star_{\alpha,P|\mathcal{G}}(Z)(\varpi_i) = \max_{\omega\in\varpi_i} Z(\omega). \tag{2.11}$$

On the other hand consider the dual sets $\mathfrak{A}_i$ associated with the respective conditional Average Value-at Risk. That is, for $i \in \mathcal{I}$ consider the conditional probabilities $p'(\omega) := P(\{\omega\}|\varpi_i)$. Then the corresponding set $\mathfrak{A}_i$ is formed by probabilities Q supported on ϖ_i such that $q(\omega)/p'(\omega) \le \alpha^{-1}$, $\omega \in \varpi_i$. We have that

$$\mathsf{AV@R}_{\alpha,P|\mathcal{G}}(Z)(\varpi_i) = \inf_{\tau\in\mathbb{R}} \left\{\tau + \alpha^{-1}\,\mathbb{E}_{P(\cdot|\varpi_i)}[Z-\tau]_+\right\}. \tag{2.12}$$

The corresponding set $\mathfrak{M}_\nu = \{Q = \sum_{i\in\mathcal{I}} \nu_i Q_i |\, Q_i \in \mathfrak{A}_i,\ i \in \mathcal{I}\}$ is contained in the set $\mathfrak{M}$, and hence

$$\mathsf{AV@R}_{\alpha,P|\mathcal{G}}(Z) \le \mathsf{AV@R}^\star_{\alpha,P|\mathcal{G}}(Z), \quad Z \in \mathcal{Z}. \tag{2.13}$$

3 The general case

We consider now the general case of possibly infinite subalgebra $\mathcal{G}$. In order to proceed we make the following assumption.

Assumption 3.1. *There is a probability measure $P \in \mathfrak{P}$ (viewed as the reference measure) such that every $Q \in \mathfrak{M}$ is absolutely continuous with respect to P.*

In case the space Ω is finite we can use reference measure with all probabilities of the elementary events being positive. Then any probability measure is absolutely continuous with respect to such reference measure. Of course for a general measurable space the situation is more delicate. Unless stated otherwise we make probabilistic statements with respect to the reference measure P. By writing $Z \succeq Z'$ for two measurable variables $Z, Z' \colon \Omega \to \mathbb{R}$ we mean that $Z \ge Z'$ almost surely, i.e., $Z(\omega) \ge Z'(\omega)$ for P-almost every $\omega \in \Omega$. Note that if $Z \ge Z'$ almost surely with respect to P, then $Z \ge Z'$ almost surely with respect to any measure $Q \in \mathfrak{P}$ absolutely continuous with respect to P.

Definition 3.1. *The* essential supremum *of a (nonempty) set $\mathfrak{X}$ of measurable functions $X \colon \Omega \to \overline{\mathbb{R}}$, denoted* $\operatorname{ess\,sup} \mathfrak{X}$*, is a measurable function $X^* \colon \Omega \to \overline{\mathbb{R}}$ such that: (i) for any $X \in \mathfrak{X}$ it follows that $X^* \succeq X$, (ii) if Y is a measurable function such that $Y \succeq X$ for all $X \in \mathfrak{X}$, then $Y \succeq X^*$.*

The essential infimum, denoted $\operatorname{ess\,inf} \mathfrak{X}$, is defined similarly. The essential supremum (essential infimum) always exists and is uniquely defined in the almost sure sense, i.e., two versions of $\operatorname{ess\,sup} \mathfrak{X}$ can be different from each other on a set of P-measure zero (e.g., [7, Section A.5]). Note that definition of the essential supremum (essential infimum) depends on the specified probability measure P. Also, it is defined for a set of measurable functions, we apply this mainly to $\mathcal{G}$-measurable functions.

Let us make the following observation. The distributionally robust functional (2.1) can be written in the following form

$$\mathcal{R}(Z) = \inf \{y |\, y \ge \mathbb{E}_Q[Z],\ Q \in \mathfrak{M}\}. \tag{3.1}$$

We use this observation for defining $\mathcal{R}^\star_{|\mathcal{G}}$ in the general setting of possibly infinite $\mathcal{G}$. That is

$$\mathcal{R}^\star_{|\mathcal{G}}(Z) := \operatorname{ess\,inf} \{Y |\, Y \colon \Omega \to \overline{\mathbb{R}} \text{ is } \mathcal{G}\text{-measurable},\ Y \succeq \mathbb{E}_{Q|\mathcal{G}}[Z],\ Q \in \mathfrak{M}\}. \tag{3.2}$$

If $\mathcal{G}$ is finite, then the corresponding essential infimum becomes the pointwise infimum, and definitions (3.2) and (2.6) are equivalent. The condition $Q(\varpi_i) \neq 0$ ensures that $\mathbb{E}_{Q|\mathcal{G}}[Z](\varpi_i)$ is uniquely defined, and Assumption 2.1 ensures that $\mathcal{R}^\star_{|\mathcal{G}}(Z)(\cdot) > -\infty$. In the general case of possibly infinite $\mathcal{G}$, definition (3.2) of $\mathcal{R}^{\star}_{|\mathcal{G}}(Z)$ is equivalent to the one in Pichler and Shapiro [16, Definition 3.9], where it was defined as the essential supremum of the smallest versions of $\mathbb{E}_{Q|\mathcal{G}}(Z)$. We can ensure that $\mathcal{R}^{\star}_{|\mathcal{G}}(Z) > -\infty$ almost surely by the natural condition that $P \in \mathfrak{M}$.

Construction of conditional risk averse functionals. We proceed now to defining the conditional risk measure $\mathcal{R}_{|\mathcal{G}}$. We deal with the following construction. Assume that the measurable space $(\Omega, \mathcal{F})$ is given as the product of measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, i.e., $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$. We denote by $\mathfrak{P}_1$ and $\mathfrak{P}_2$ the sets of probability measures on the respective measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$. For a probability measure $Q \in \mathfrak{P}$ we denote by $Q_1 \in \mathfrak{P}_1$ the respective marginal probability measure on $(\Omega_1, \mathcal{F}_1)$, that is $Q_1(A) = Q(A \times \Omega_2)$ for $A \in \mathcal{F}_1$. The marginal probability measure $Q_2 \in \mathfrak{P}_2$ is defined in the similar way. We assume that $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are *Polish spaces*.[2]

Consider sigma subalgebra $\mathcal{G}$ of $\mathcal{F}$ consisting of the sets $A \times \Omega_2$, $A \in \mathcal{F}_1$, that is

$$\mathcal{G} \coloneqq \{A \times \Omega_2 \colon A \in \mathcal{F}_1\}. \tag{3.3}$$

It is straightforward to verify that indeed such defined $\mathcal{G}$ satisfies the axioms of sigma algebra. Note that the elements of subalgebra $\mathcal{G}$ are determined by sets (events) $A \in \mathcal{F}_1$, and in that sense $\mathcal{G}$ can be identified with the sigma algebra $\mathcal{F}_1$, we write this as $\mathcal{G} \equiv \mathcal{F}_1$. A random variable $Z(\omega', \omega'')$, $\omega = (\omega', \omega'') \in \Omega_1 \times \Omega_2$, is $\mathcal{G}$-measurable iff $Z(\omega', \cdot)$ is constant on Ω_2 for every $\omega' \in \Omega_1$, and $Z(\cdot, \omega'')$ is $\mathcal{F}_1$-measurable for every $\omega'' \in \Omega_2$. Therefore, with some abuse of the notation, we write a $\mathcal{G}$-measurable variable as a function $Z(\omega')$ of $\omega' \in \Omega_1$. We also use notation $Z_{\omega'}(\omega'') := Z(\omega', \omega'')$ viewed as random variable $Z_{\omega'} : \Omega_2 \to \mathbb{R}$.

Remark 3.1. If Ω_1 is *finite*, then $\mathcal{G}$ is finite with the respective atoms (elementary events) $\{\omega'\} \times \Omega_2$, $\omega' \in \Omega_1$. When both Ω_1 and Ω_2 are finite, this can be viewed as a scenario tree with the nodes $\omega' \in \Omega_1$ at stage one, and the children nodes $(\omega', \omega'')_{\omega'' \in \Omega_2}$, of node ω', at stage $t = 2$. For a probability measure $Q \in \mathfrak{P}$ and $\omega' \in \Omega_1$ we can view $Q(\{\omega'\} \times \{\omega''\})$ as the probability of moving from to node ω' and from this node further to its child node (ω', ω''). Some of these probabilities can be zero, so effectively the number of child nodes of ω' can be less than the cardinality of Ω_2. $\square$

Suppose that there is a reference probability measure $P_1 \in \mathfrak{P}_1$ with $\mathsf{supp}(P_1) = \Omega_1$. By saying for "a.e. $\omega' \in \Omega_1$" we mean that this holds almost surely with respect to P_1. Assume that for a.e. $\omega' \in \Omega_1$, is given a (nonempty) set $\mathcal{A}_{\omega'} \subset \mathfrak{P}_2$ of probability measures. For $Q_{\omega'} \in \mathcal{A}_{\omega'}$ consider the corresponding expectation

$$\mathbb{E}_{Q_{\omega'}}[Z_{\omega'}] = \int_{\Omega_2} Z(\omega', \omega'') dQ_{\omega'}(\omega''). \tag{3.4}$$

[2]Polish space is a separable complete metric space equipped with its Borel sigma algebra. In particular, any closed subset of $\mathbb{R}^m$ is a Polish space.

We say that $Q_{\omega'} \in \mathcal{A}_{\omega'}$ is a *measurable selection* of $\mathcal{A}_{\omega'}$ if, for every $Z \in \mathcal{Z}$, the integral in (3.4) is well-defined and finite for almost every $\omega' \in \Omega_1$, and the mapping (function) $\omega' \mapsto \mathbb{E}_{Q_{\omega'}}[Z_{\omega'}]$ is $\mathcal{F}_1$-measurable. Then we define

$$\mathcal{R}_{|\mathcal{G}}(Z) := \operatorname{ess\,inf} \left\{ Y \,\middle|\, \begin{array}{l} Y\colon \Omega_1 \to \overline{\mathbb{R}} \text{ is measurable, } Y(\omega') \geq \mathbb{E}_{Q_{\omega'}}[Z_{\omega'}] \text{ a.s.} \\ \text{and } Q_{\omega'} \text{ is a measurable selection of } \mathcal{A}_{\omega'} \end{array} \right\}, \tag{3.5}$$

where 'measurable' means with respect to the sigma algebra $\mathcal{F}_1$ (recall that $\mathcal{G} \equiv \mathcal{F}_1$).

If $\mathcal{F}_1$ is finite, then any $Y\colon \Omega_1 \to \overline{\mathbb{R}}$ is measurable, any selection $Q_{\omega'} \in \mathcal{A}_{\omega'}$ is measurable (provided the corresponding integral is well-defined and finite), and the essential infimum coincides with the pointwise infimum. Therefore, for finite $\mathcal{F}_1$, definitions (3.5) and (2.7) of $\mathcal{R}_{|\mathcal{G}}$ are equivalent (compare (3.4) with (2.5)). In general, we should be careful with verification of existence of measurable selections, we will discuss this later (in particular see Remark 3.3 below).

The relation (2.9) between the sets $\mathfrak{A}_i$ and the corresponding set of probability measures on $(\Omega, \mathcal{F})$ can be extended to the considered setting in the following way.

Definition 3.2 (Regular Probability Kernel)**.** *A function* $\mathbb{K}\colon \mathcal{F}_2 \times \Omega_1 \to [0,1]$ *is said to be a* Regular Probability Kernel *(RPK) of a probability measure* $Q \in \mathfrak{P}$ *if the following properties hold:* (i) $\mathbb{K}(\cdot|\omega')$ *is a probability measure for* Q_1*-almost every* $\omega' \in \Omega_1$*,* (ii) *for every* $B \in \mathcal{F}_2$ *the function* $\mathbb{K}(B|\cdot)$ *is* $\mathcal{F}_1$*-measurable,* (iii) *for every* $A \in \mathcal{F}_1$ *and* $B \in \mathcal{F}_2$ *it follows that*

$$Q(A \times B) = \int_A \mathbb{K}(B|\omega') dQ_1(\omega'). \tag{3.6}$$

In particular, $Q_2(B) = \int_{\Omega_1} \mathbb{K}(B|\omega') dQ_1(\omega')$ *is the respective marginal probability measure.*

The Disintegration Theorem (e.g. Dellacherie and Meyer [4, III-70]) ensures existence of the RPK for a wide class of measurable spaces, in particular if $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ are Polish spaces. Therefore, we assume existence of the RPK for every $Q \in \mathfrak{P}$. The function $\mathbb{K}(B|\cdot)$ is defined up to Q_1-measure zero, and is uniquely determined on the support, $\mathsf{supp}(Q_1)$, of Q_1. The RPK is associated with a specified $Q \in \mathfrak{P}$, we sometimes write $\mathbb{K}_Q$ to emphasize this.

Remark 3.2. In order to develop an intuition, suppose that Ω_1 is finite consisting of elements (atoms) ω'_i, $i \in \mathcal{I}$, with $\mathcal{I}$ being a finite index set. Then we can define

$$\mathbb{K}(B|\omega') := Q\big(\{\omega'\} \times B \,\big|\, \{\omega'\} \times \Omega_2\big) = \frac{Q(\{\omega'\} \times B)}{Q_1(\{\omega'\})}, \text{ if } Q_1(\{\omega'\}) \neq 0. \tag{3.7}$$

That is, $\mathbb{K}(B|\omega')$ is the conditional probability of the event $\{\omega'\} \times B \in \mathcal{F}$ given the event $\{\omega'\} \times \Omega_2 \in \mathcal{G}$. Here the support of Q_1 is

$$\mathsf{supp}(Q_1) = \{\omega' \in \Omega_1 \colon Q_1(\{\omega'\}) \neq 0\},$$

and as it was already mentioned, $\mathbb{K}(B|\cdot)$ is uniquely defined on $\mathsf{supp}(Q_1)$.

Consider an atom $\varpi = \{\omega'\} \times \Omega_2$ of the subalgebra $\mathcal{G}$. In accordance with (2.4) the corresponding probability measure $Q_{|\omega'}$, supported on ϖ, is defined as

$$Q_{|\omega'}(\{\omega'\} \times B) = \mathbb{K}(B|\omega'), \ B \in \mathcal{F}_2, \tag{3.8}$$

provided that $Q_1(\{\omega'\}) \neq 0$, i.e., $\omega' \in \mathsf{supp}(Q_1)$. $\square$

Remark 3.3. For $Q \in \mathfrak{P}$ and $\omega' \in \mathsf{supp}(Q_1)$ we can define a probability measure $Q_{|\omega'} \in \mathfrak{P}_2$ as

$$Q_{|\omega'}(B) := \mathbb{K}_Q(B|\omega'), \ B \in \mathcal{F}_2. \tag{3.9}$$

We have that[3] (compare with (2.5))

$$\mathbb{E}_{Q|\mathcal{G}}[Z](\omega') = \int_{\Omega_2} Z(\omega', \omega'') dQ_{|\omega'}(\omega''), \quad \omega' \in \mathsf{supp}(Q_1), \tag{3.10}$$

and $\mathbb{E}_{Q|\mathcal{G}}[Z](\omega')$ can be arbitrary for $\omega' \in \Omega_1 \setminus \mathsf{supp}(Q_1)$. By property (ii) of the RPK we have that if $Q_{|\omega'}$, defined in (3.9), belongs to $\mathcal{A}_{|\omega'}$, $\omega' \in \Omega_1$, then $Q_{|\omega'}$ is a measurable selection of $\mathcal{A}_{|\omega'}$. $\square$

For $\nu \in \mathfrak{P}_1$ with $\mathsf{supp}(\nu) = \Omega_1$, define

$$\mathfrak{M}_\nu := \left\{ Q \in \mathfrak{P} \,\middle|\, Q_1 = \nu, \ \mathbb{K}_Q(\cdot|\omega') \in \mathcal{A}_{\omega'} \text{ for a.e. } \omega' \in \Omega_1 \right\}. \tag{3.11}$$

For finite Ω_1 we have that for all $\omega' \in \Omega_1$,

$$\mathcal{A}_{\omega'} = \left\{ \mathbb{K}_Q(\cdot \mid \omega') \,\middle|\, Q \in \mathfrak{M}_\nu \right\}, \tag{3.12}$$

and (3.11) defines the same set $\mathfrak{M}_\nu$ as in (2.9) (the condition there that all probabilities ν_i are positive ensues that the support of the corresponding probability measure coincides with Ω_1). In general suppose that (3.12) holds for a.e. $\omega' \in \Omega_1$. Then for $\mathfrak{M} = \mathfrak{M}_\nu$ we have that $\mathcal{R}^\star_{|\mathcal{G}}(\cdot) = \mathcal{R}_{|\mathcal{G}}(\cdot)$ (compare with Proposition 2.1).

Example 3.1. Let $\Omega_1 = \Omega_2 = [0, 1]$ and $\mathcal{F}_1$ and $\mathcal{F}_2$ be the respective Borel sigma algebras. Let the reference measure P be the uniform probability measure on $[0, 1] \times [0, 1]$. It follows that the marginal distributions P_1 and P_2 are given by the uniform probability measure on the interval $[0, 1]$. Suppose that the ambiguity set $\mathfrak{M}$ consists of probability measures Q absolutely continuous with respect to P and such that the set of densities $dQ/dP \leq \alpha^{-1}$ a.s. for some $\alpha \in (0, 1)$. That is, here the distributionally robust functional $\mathcal{R} = \mathsf{AV@R}_{\alpha,P}$, defined on the space $\mathcal{Z} = L_1(\Omega, \mathcal{F}, P)$ of P-integrable functions.

For $Q \in \mathfrak{M}$ with the respective density $f = dQ/dP$, the conditional expectation with respect to $\mathcal{G} \equiv \mathcal{F}_1$ is

$$\mathbb{E}_{Q|\mathcal{G}}[Z](\omega') = \int_0^1 Z(\omega', \omega'') f(\omega', \omega'') d\omega''. \tag{3.13}$$

[3]Since $\mathbb{E}_{Q|\mathcal{G}}[Z]$ is $\mathcal{G}$-measurable, we write $\mathbb{E}_{Q|\mathcal{G}}[Z](\omega')$ as a function of ω' as it was mentioned before.

For a given $\omega' \in [0,1]$ and any measurable set $A \subset [0,1]$ we can choose $f = dQ/dP$, $Q \in \mathfrak{M}$, such that $f(\omega', \omega'') = \alpha^{-1}$ for $\omega'' \in A$, and $f(\omega', \omega'') = 0$ for $\omega'' \in [0,1] \setminus A$. It follows that (compare with (2.11))

$$\mathsf{AV@R}^{\star}_{|\mathcal{G}}(Z)(\cdot) = \alpha^{-1} \operatorname*{ess\,sup}_{\omega'' \in [0,1]} Z(\cdot, \omega''). \tag{3.14}$$

In order to define the conditional $\mathsf{AV@R}_{\alpha,P|\mathcal{G}}$, consider the sets

$$\mathcal{A}_{\omega'} := \big\{Q_{\omega'} \ll P_2 \colon dQ_{\omega'}/dP_2 \le \alpha^{-1} \text{ a.s.}\big\}. \tag{3.15}$$

The corresponding set $\mathfrak{M}_\nu$ consists of probability measures Q absolutely continuous with respect to P with densities $f = dQ/dP \le \alpha^{-1}$ such that $\int_0^1 f(\omega', \omega'') d\omega'' = 1$ for a.e. ω', and such that $Q_1 = \nu$, i.e., $\nu(A) = \int_A f(\omega', \omega'') d\omega'$, $A \in \mathcal{F}_1$, $\omega'' \in [0,1]$. We have that $\mathfrak{M}_\nu \subset \mathfrak{M}$ and the inequality (2.13) holds here as well. $\square$

Rectangularity. The derivations simplify considerably in the following, so called rectangular, setting. Suppose that the ambiguity set is of the following form

$$\mathfrak{M} := \{Q = Q_1 \times Q_2 |\, Q_1 \in \mathcal{M}_1, Q_2 \in \mathcal{M}_2\}, \tag{3.16}$$

where $\mathcal{M}_1 \subset \mathfrak{P}_1$ and $\mathcal{M}_2 \subset \mathfrak{P}_2$ are sets of marginal distributions (measures). That is, any $Q = Q_1 \times Q_2 \in \mathfrak{M}$ is the direct product of its marginal distributions, and

$$\mathbb{E}_{Q|\mathcal{G}}[Z](\omega') = \int_{\Omega_2} Z(\omega', \omega'') dQ_2(\omega'') = \mathbb{E}_{Q_2}[Z_{\omega'}]. \tag{3.17}$$

It follows that in the rectangular case

$$\mathcal{R}^{\star}_{|\mathcal{G}}(Z)(\omega') = \sup_{Q_2 \in \mathcal{M}_2} \int_{\Omega_2} Z(\omega', \omega'') dQ_2(\omega''). \tag{3.18}$$

Also in the rectangular case the Regular Probability Kernel $\mathbb{K}_Q(\cdot|\omega') = Q_2$ is independent of ω', and the sets $\mathcal{A}_{\omega'}$ are given by $\mathcal{A}_{\omega'} := \mathcal{M}_2$ for all ω'. It follows that in the rectangular case $\mathcal{R}^{\star}_{|\mathcal{G}} = \mathcal{R}_{|\mathcal{G}}$. The measurable selections in the definition of $\mathcal{R}_{|\mathcal{G}}$ are ensured by the respective Regular Probability Kernels.

3.1 Strict monotonicity

It follows from the definition (2.1) that the distributionally robust functional $\mathcal{R}$ is monotone,[4] i.e., if $Z' \succeq Z$, then $\mathcal{R}(Z') \ge \mathcal{R}(Z)$. It is said that the distributionally robust functional $\mathcal{R}$ is *strictly* monotone if $Z' \succeq Z$ and $Z' \ne Z$ implies that $\mathcal{R}(Z') \ne \mathcal{R}(Z)$. In other words if $Z' \ge Z$ a.s. and $Z' > Z$ with positive probability, then $\mathcal{R}(Z') > \mathcal{R}(Z)$. Similarly, the conditional functional $\mathcal{R}^{\star}_{|\mathcal{G}}$ (conditional functional $\mathcal{R}_{|\mathcal{G}}$) is said to be *strictly* monotone if $Z' \succeq Z$ and $Z' \ne Z$ implies that $\mathcal{R}^{\star}_{|\mathcal{G}}(Z') \ne \mathcal{R}^{\star}_{|\mathcal{G}}(Z)$ (that $\mathcal{R}_{|\mathcal{G}}(Z') \ne \mathcal{R}_{|\mathcal{G}}(Z)$).

[4] Recall that the inequality '$\succeq$' is understood in the almost sure sense with respect to the reference measure, and that it is assumed that the ambiguity set $\mathfrak{M}$ is convex and closed.

Strict monotonicity is important in investigation of time consistency and derivation of dynamic equations (cf. Shapiro [22]).

We say that probability measure $Q \in \mathfrak{P}$ is *strictly positive* if $Q(A) > 0$ for any $A \in \mathcal{F}$ such that $P(A) > 0$, i.e., if P is absolutely continuous with respect to Q. If Ω is finite, this simply means that all probabilities $q(\omega) = Q(\{\omega\})$, $\omega \in \Omega$, are positive. It is known that $\mathcal{R}$ is strictly monotone iff every $Q \in \mathfrak{M}$ is strictly positive (cf., [22]) (recall that it is assumed that every $Q \in \mathfrak{M}$ is absolutely continuous with respect to the reference measure P). It is shown in [16, Proposition 3.14] that if the distributionally robust functional $\mathcal{R}$ is strictly monotone, then the corresponding conditional distributionally robust functional $\mathcal{R}^{\star}_{|\mathcal{G}}$ is strictly monotone. The converse of that is not true in general, it could happen that $\mathcal{R}^{\star}_{|\mathcal{G}}$ is strictly monotone while $\mathcal{R}$ is not strictly monotone (see Example 3.2 below).

Suppose that Ω is finite and every set $\mathfrak{A}_i$, used in the definition of $\mathcal{R}_{|\mathcal{G}}$, is convex and closed. Then $\mathcal{R}_{|\mathcal{G}}$ is strictly monotone iff for every $i \in \mathcal{I}$ and every $Q \in \mathfrak{A}_i$ it holds that $q(\omega) > 0$ for every $\omega \in \varpi_i$. Clearly this condition holds iff, for any positive probabilities ν_i, every probability measure of the set $\mathfrak{M}_\nu = \sum_{i\in\mathcal{I}} \nu_i \mathfrak{A}_i$ is strictly positive on Ω. On the other hand, consider the distributionally robust functional $\mathcal{R}$ associated with an ambiguity set $\mathfrak{M}$ of probability measures. Clearly if every $Q \in \mathfrak{M}$ is strictly positive on Ω, then the respective conditional probabilities $Q(\cdot \mid \varpi_i)$ are strictly positive on ϖ_i. Therefore, we have the following.

Proposition 3.1. *Let Ω be finite. Then the conditional risk functional $\mathcal{R}_{|\mathcal{G}}$ is strictly monotone iff, for any positive probabilities ν_i, the corresponding distributionally robust functional $\mathcal{R}$, associated with the ambiguity set $\mathfrak{M} := \mathfrak{M}_\nu$, is strictly monotone.*

Let us consider the following simple example.

Example 3.2. Suppose that $\Omega = \{\omega_1, \dots, \omega_n\}$ is finite, $\mathcal{G} = \mathcal{F}$ and Assumption 2.1 holds. In that case the corresponding partition is given by the singletons $\varpi_i = \{\omega_i\}$, $i = 1, \dots, n$, with $|\mathcal{I}| = n$. Here Assumption 2.1 means that for every $\omega \in \Omega$ there exists $Q \in \mathfrak{M}$ such that $q(\omega) \neq 0$. For such Q we have that $\mathbb{E}_{Q|\mathcal{G}}[Z](\omega) = Z(\omega)$ for any $Z \in \mathcal{Z}$. It follows that $\mathcal{R}^{\star}_{|\mathcal{G}}(Z) = Z$ for any $Z \in \mathcal{Z}$, and hence $\mathcal{R}^{\star}_{|\mathcal{G}}(Z') \neq \mathcal{R}^{\star}_{|\mathcal{G}}(Z)$ if $Z \neq Z'$. It follows that $\mathcal{R}^{\star}_{|\mathcal{G}}(Z)$ is strictly monotone, while $\mathcal{R}$ could be not strictly monotone.

Because of Assumption 2.1 and by (2.8) we have here that $\mathfrak{A}_i = \{p_i\}$ is the singleton, with $p_i = 1$, $i = 1, \dots, n$. Consequently, the corresponding set $\sum_{i=1}^n \nu_i \mathfrak{A}_i$, defined in (2.9), is the singleton consisting of the probability vector ν, and can be different from the above ambiguity set $\mathfrak{M}$. It follows that $\mathcal{R}_{|\mathcal{G}} = \mathbb{E}_{|\mathcal{G}}$ and $\mathcal{R}_{|\mathcal{G}}(Z) = Z$ for any $Z \in \mathcal{Z}$. The distributionally robust functional associated with $\sum_{i=1}^n \nu_i \mathfrak{A}_i$ is just the expectation corresponding to the probability vector ν. Since it is assumed that the probabilities ν_i, $i = 1, \dots, n$, are positive, this expectation functional is strictly monotone. $\square$

3.2 Law invariance

Denote by $F_Z(z) := P(Z \leq z)$ the cumulative distribution function (cdf) of random variable Z, with respect to the reference measure P. It is said that $Z, Z' \in \mathcal{Z}$ are distributionally equivalent if their cdfs do coincide, i.e., $F_Z(z) = F_{Z'}(z)$ for any $z \in \mathbb{R}$.

Definition 3.3. *It is said that a functional* $\mathcal{R}\colon \mathcal{Z} \to \mathbb{R}$ *is* law invariant *if* $\mathcal{R}(Z) = \mathcal{R}(Z')$ *for any distributionally equivalent* $Z, Z' \in \mathcal{Z}$.

We are going to extend this concept to *conditional* law invariance. In order to get an insight, while avoiding delicate measurability issues, unless stated otherwise we assume below that Ω is *finite*.

For general probability spaces it makes sense to talk about law invariant risk measures when the reference probability space $(\Omega, \mathcal{F}, P)$ is atomless. For finite space Ω it makes sense to consider the law invariance when the reference probability measure assigns equal weights to all elementary events, i.e., $p(\omega) = 1/n$ for every $\omega \in \Omega = \{\omega_1, \dots, \omega_n\}$. In that case Z,$Z'\colon \Omega \to \mathbb{R}$ are distributionally equivalent, with respect to the reference probability measure P, iff there exists a permutation π of the set $\{1, \dots, n\}$ such that $Z'(\cdot) = Z(\pi(\cdot))$.

Definition 3.4. *We say that the set* $\mathfrak{M}$ *is* permutationally invariant *if it satisfies the following property: if* $Q \in \mathfrak{M}$ *and* π *is a permutation of the set* $\Omega = \{1, \dots, n\}$*, then* $Q' \in \mathfrak{M}$ *where* $q'(\cdot) = q(\pi(\cdot))$. *We say that the set* $\mathfrak{A}_i$*, of probability measures supported on* ϖ_i*, is* permutationally invariant *if for any* $Q \in \mathfrak{A}_i$ *and any permutation* π *of the elements of* ϖ_i*, it follows that* $Q' \in \mathfrak{A}_i$ *where* $q'(\omega) = q(\pi(\omega))$ *for* $\omega \in \varpi_i$.

In the above setting we have that the distributionally robust functional $\mathcal{R}$ is law invariant iff its ambiguity set $\mathfrak{M}$ is permutationally invariant. We can consider conditional law invariance of the conditional functionals.

Definition 3.5. *It is said that the conditional functional* $\mathcal{R}_{|\mathcal{G}}$ *is conditionally law invariant if for every* $i \in \mathcal{I}$ *and any distributionally equivalent* Z *and* Z' *such that* $Z(\omega) = Z'(\omega) = 0$ *for all* $\omega \notin \varpi_i$ *it follows that* $\mathcal{R}_{|\mathcal{G}}(Z) = \mathcal{R}_{|\mathcal{G}}(Z')$.

We have that $\mathcal{R}_{|\mathcal{G}}$ is conditionally law invariant iff each set $\mathfrak{A}_i$ is permutationally invariant. The conditional law invariance is a property inherited by conditional counterparts of law invariant risk measures. It could be noted that typically the ambiguity set $\mathfrak{M}$ of the form (2.9) is not permutationally invariant even if each $\mathfrak{A}_i$ is permutationally invariant. On the other hand if the set $\mathfrak{M}$ is permutationally invariant, then every respective set $\mathfrak{A}_i$, defined in (2.8), is permutationally invariant. That is, we have the following.

Proposition 3.2. *Assume the setting of the reference measure* $p_i = 1/n$, $i = 1, \dots, n$, *and let the sets* $\mathfrak{A}_i$*, used in the definition of* $\mathcal{R}_{|\mathcal{G}}$*, be of the form* (2.8). *In that setting we have that if the distributionally robust functional* $\mathcal{R}$ *is law invariant, then its conditional counterpart* $\mathcal{R}_{|\mathcal{G}}$ *is conditionally law invariant.*

Conditional law invariance can be extended to the general setting in the following way. Let $\mathbb{K}_P$ be the RPK of the reference measure P and $P_{|\omega'} = \mathbb{K}_P(\cdot|\omega')$ be the corresponding probability measure on $(\Omega_2, \mathcal{F}_2)$. Then we can define the conditional (on $\mathcal{G} \equiv \mathcal{F}_1$) counterpart of the cdf F_Z as (compare with (3.9))[5]

$$F_{Z|\omega'}(z) := P_{|\omega'}(Z_{\omega'} \le z), \;\; \text{a.e. } \omega' \in \Omega_1. \tag{3.19}$$

That is, $F_{Z|\omega'}$ is the cdf of $Z_{\omega'} : \Omega_2 \to \mathbb{R}$ associated with the probability measure $P_{|\omega'}$.

Suppose that $\mathcal{R}$ is law invariant. Then it can be considered as a function of the cdf F_Z, i.e., we can view $\mathcal{R}(Z) = \mathcal{R}(F_Z)$. We can define the respective (conditionally law invariant) $\mathcal{R}_{|\mathcal{G}}(Z)$ as the corresponding function of the cdf $F_{Z|\omega'}$, that is

$$\mathcal{R}_{|\mathcal{G}}(Z)(\omega') := \mathcal{R}(F_{Z|\omega'}), \;\; \text{a.e. } \omega' \in \Omega_1. \tag{3.20}$$

This definition is in line with the usual approach to defining conditional counterparts of coherent risk measures in the law invariant case. It coincides with the definition (3.5) in the setting of Remark 3.3 when the sets $\mathcal{A}_{\omega'}$ are given in the form (3.12).

3.3 Conditional ϕ-divergence

Consider a convex lower semicontinuous function $\phi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ such that $\phi(1) = 0$ and $\phi(x) = +\infty$ for $x < 0$. For probability measures $Q, P \in \mathfrak{P}$ such that $Q \ll P$, the ϕ-divergence is defined as (cf. Csiszár [3], Morimoto [12])

$$D_\phi(Q\|P) := \mathbb{E}_P[\phi(dQ/dP)] = \int \phi(dQ/dP)dP. \tag{3.21}$$

In particular, for $\phi(x) := x \log x - x + 1$, $x \ge 0$, this becomes the Kullback–Leibler (KL) divergence of Q from P. If Ω is finite it can be written as

$$D_{KL}(Q\|P) = \sum_{\omega \in \Omega} p(\omega)\big(q(\omega)/p(\omega)\big) \log\big(q(\omega)/p(\omega)\big) = \sum_{\omega \in \Omega} q(\omega) \log\big(q(\omega)/p(\omega)\big). \tag{3.22}$$

Note that $D_\phi(\cdot\|P)$ is law invariant with respect to P. Assume in the reminder of this section is that Ω is finite. As it was pointed out before, for a finite space Ω it makes sense to consider law invariance when the reference probability measure assigns equal weights to all elementary events. That is, suppose that $p(\omega) = 1/n$ for every $\omega \in \Omega = \{\omega_1, \ldots, \omega_n\}$. Then we have that $D_\phi(Q'\|P) = D_\phi(Q\|P)$ for $Q' \ll P$ and $Q \ll P$ if there is a permutation $\pi : \Omega \to \Omega$ such that $q'(\cdot) = q(\pi(\cdot))$, i.e., if Q' and Q are distributionally equivalent with respect to P. This motivates to consider the following conditional counterpart of $D_\phi(\cdot\|P)$.

Definition 3.6. *Let $P \in \mathfrak{P}$ be such that $P(\varpi_i) > 0$ for all $i \in \mathcal{I}$. The conditional counterpart of ϕ-divergence* (3.21)*, denoted $D_{\phi|\mathcal{G}}(Q\|P)$, is a $\mathcal{G}$-measurable variable defined as*

$$D_{\phi|\mathcal{G}}(Q\|P)(\varpi_i) := D_\phi(Q_{|\varpi_i}\|P_{|\varpi_i}), \;\; \text{for } Q(\varpi_i) > 0, \; i \in \mathcal{I}. \tag{3.23}$$

If $Q(\varpi_i) = 0$, we set $D_{\phi|\mathcal{G}}(Q\|P)(\varpi_i) := +\infty$.

[5] Recall that $Z_{\omega'}(\cdot) = Z(\omega', \cdot)$.

It could be noted that

$$D_{\phi|\mathcal{G}}(Q\|P)(\varpi_i) = \sum_{\omega\in\varpi_i} p'(\omega)\phi\big(q'(\omega)/p'(\omega)\big), \text{ for } Q(\varpi_i) > 0, \tag{3.24}$$

where $p'(\omega) := P(\{\omega\}|\,\varpi_i)$ and $q'(\omega) := Q(\{\omega\}|\,\varpi_i)$ are the respective conditional probabilities.

The ambiguity set associated with $D_\phi(\cdot\|P)$ is defined as

$$\mathfrak{M}_\epsilon := \{Q \in \mathfrak{P} \colon D_\phi(Q\|P) \le \epsilon\} \tag{3.25}$$

for some $\epsilon \ge 0$. This defines the corresponding distributionally robust functional $\mathcal{R}_\epsilon$ of the form (2.1). The conditional risk averse counterpart $\mathcal{R}_{\epsilon|\mathcal{G}}$ is determined by formula (2.7) with the sets $\mathfrak{A}_i$ of the form

$$\mathfrak{A}_i := \left\{Q_{|\varpi_i} \colon D_{\phi|\mathcal{G}}(Q\|P)(\varpi_i) \le \epsilon,\ Q \in \mathfrak{P}\right\},\ i \in \mathcal{I}. \tag{3.26}$$

This is an approach of "self similarity" to the construction of conditional counterparts of the respective distributionally robust functional. Note that if $Q(\varpi_i) = 0$, then $Q_{|\varpi_i}$ is not included in the set $\mathfrak{A}_i$ since in that case $D_{\phi|\mathcal{G}}(Q\|P)(\varpi_i) = +\infty$. Note also that $P_{|\varpi_i} \in \mathfrak{A}_i$, and hence the sets $\mathfrak{A}_i$ are nonempty.

We can also consider the conditional distributionally robust counterpart $\mathcal{R}^\star_{\epsilon|\mathcal{G}}$ determined by formula (2.2) with $\mathfrak{M} = \mathfrak{M}_\epsilon$. Note that if $Q \in \mathfrak{M}_\epsilon$, then $Q_{|\varpi_i}$ belongs to the set $\mathfrak{A}_i$ defined in (3.26). That is, the sets determined by formula (2.8) include the sets $\mathfrak{A}_i$ determined by (3.26). It follows that for conditional functionals determined by the ϕ-divergence approach, the following inequality holds for $\epsilon \ge 0$,

$$\mathcal{R}_{\epsilon|\mathcal{G}}(Z) \le \mathcal{R}^\star_{\epsilon|\mathcal{G}}(Z), \quad Z \in \mathcal{Z}. \tag{3.27}$$

For example, for $\alpha \in (0,1)$ consider function $\phi(x) := 0$ for $x \in [0, \alpha^{-1}]$, and $\phi(x) := +\infty$ otherwise. The corresponding set $\mathfrak{M}_\epsilon$, given in (3.25), consists of probability measures $Q \in \mathfrak{P}$ such that $dQ/dP \le \alpha^{-1}$. That is, $\mathfrak{M}_\epsilon$ is the dual set of the $\mathsf{AV@R}_{\alpha,P}$ risk measure for any $\epsilon > 0$. The conditional counterpart of $\mathsf{AV@R}_{\alpha,P}$ is obtained by using formula (2.7) with the sets $\mathfrak{A}_i$ of the form (3.26). On the other hand, as it was shown in Example 2.1, $\mathcal{R}^\star_{|\mathcal{G}}(Z)(\varpi_i) = \max_{\omega\in\varpi_i} Z(\omega)$ provided $\alpha \le 1 - P(\varpi_i)$.

Remark 3.4. It is possible to extend definition of the conditional risk measure $\mathcal{R}_{\epsilon|\mathcal{G}}$, associated with the conditional divergence $D_{\phi|\mathcal{G}}$, to a general setting. By duality arguments the distributionally robust functional $\mathcal{R}_\epsilon$ can be written as

$$\mathcal{R}_\epsilon(Z) = \inf_{\gamma,\lambda>0} \left\{\lambda\epsilon + \gamma + \lambda\mathbb{E}_P[\phi^*((Z-\gamma)/\lambda)]\right\}, \tag{3.28}$$

where ϕ^* is the conjugate of ϕ (cf. Ben-Tal and Teboulle [2], Bayraksan and Love [1]). Then the conditional counterpart of $\mathcal{R}_\epsilon$ can be written as

$$\mathcal{R}_{\epsilon|\mathcal{G}}(Z) = \inf_{\Gamma,\Lambda>0} \left\{\Lambda\epsilon + \Gamma + \Lambda\mathbb{E}_{P|\mathcal{G}}[\phi^*((Z-\Gamma)/\Lambda)]\right\}, \tag{3.29}$$

where Γ and Λ are $\mathcal{G}$-measurable variables now. For the $\mathsf{AV@R}_\alpha$ risk measure such approach to defining its conditional counterpart was already used in Ruszczyński and Shapiro [19]. $\square$

3.4 Conditional distances

In this section we discuss conditional counterparts of a distance between probability measures. In the divergence approach of Section 3.3 it was essential to assume that the probability measure Q is absolutely continuous with respect to the (reference) measure P. We do not make this assumption in this section. Let $\mathsf{d}(Q,P)$ be a distance (semi-distance) defined on the set $\mathfrak{P}$ of probability measures. We do not assume that $Q \ll P$ or $P \ll Q$. Also, the distance $\mathsf{d}(\cdot,P)$ does not have to be law invariant. The conditional counterpart of $\mathsf{d}(\cdot,\cdot)$ is defined in a way similar to the conditional divergence. Assume in the remainder of this section that the sigma subalgebra $\mathcal{G}$ is finite.

Definition 3.7. *The conditional counterpart of the distance* $\mathsf{d}(Q,P)$*, denoted* $\mathsf{d}_{|\mathcal{G}}(Q,P)$*, is a* $\mathcal{G}$*-measurable variable defined as*

$$\mathsf{d}_{|\mathcal{G}}(Q,P)(\varpi_i) \coloneqq \begin{cases} \mathsf{d}(Q_{|\varpi_i}, P_{|\varpi_i}) & \text{if } Q(\varpi_i) > 0 \text{ and } P(\varpi_i) > 0, \\ +\infty & \text{if } Q(\varpi_i) = 0 \text{ or } P(\varpi_i) = 0, \end{cases} \tag{3.30}$$

$i \in \mathcal{I}$.

As in the divergence approach, the respective ambiguity set can be defined as

$$\mathfrak{M}_\epsilon \coloneqq \{Q \in \mathfrak{P} \colon \mathsf{d}(Q,P) \le \epsilon\}, \tag{3.31}$$

for a reference measure $P \in \mathfrak{P}$ and $\epsilon \ge 0$. Of course, $P \in \mathfrak{M}_\epsilon$ and hence the set $\mathfrak{M}_\epsilon$ is nonempty. Suppose further that $P(\varpi_i) > 0$ for all $i \in \mathcal{I}$. Then the sets

$$\mathfrak{A}_i \coloneqq \left\{Q_{|\varpi_i} \colon \mathsf{d}_{|\mathcal{G}}(Q,P)(\varpi_i) \le \epsilon,\ Q \in \mathfrak{P}\right\},\ i \in \mathcal{I}, \tag{3.32}$$

are nonempty, and hence define the corresponding conditional risk functional $\mathcal{R}_{\epsilon|\mathcal{G}}$ in accordance with (2.7).

Remark 3.5. By the same arguments as in the divergence approach, the inequality (3.27) holds here as well for the respective conditional functionals $\mathcal{R}_{\epsilon|\mathcal{G}}$ and $\mathcal{R}^\star_{\epsilon|\mathcal{G}}$ associated with distance d and $\epsilon \ge 0$. □

Functional distance. A large class of distances (semi-distances) can be defined as follows. Let $\mathfrak{H}$ be a (nonempty) set of (test) measurable functions $h \colon \Omega \to \mathbb{R}$, and define

$$\mathsf{d}_{\mathfrak{H}}(Q,P) \coloneqq \sup_{h \in \mathfrak{H}} \left|\mathbb{E}_Q(h) - \mathbb{E}_P(h)\right|. \tag{3.33}$$

If the expectations in (3.33) are well-defined and the supremum is finite for any $Q, P \in \mathfrak{P}$, then $\mathsf{d}_{\mathfrak{H}}$ defines a semi-distance on $\mathfrak{P}$; if moreover $\mathsf{d}_{\mathfrak{H}}(Q,P) \ne 0$ for $Q \ne P$, then $\mathsf{d}_{\mathfrak{H}}$ is a distance. Its conditional counterpart is the $\mathcal{G}$-measurable variable defined by (3.30). If Ω is finite it can be written as

$$\mathsf{d}_{\mathfrak{H}|\mathcal{G}}(Q,P)(\varpi_i) = \sup_{h \in \mathfrak{H}} \Big| \sum_{\omega \in \varpi_i} \big(q'(\omega) - p'(\omega)\big) h(\omega) \Big|, \text{ if } Q(\varpi_i) > 0 \text{ and } P(\varpi_i) > 0, \tag{3.34}$$

$i \in \mathcal{I}$, where $p'(\omega) := P(\{\omega\}|\varpi_i)$ and $q'(\omega) := Q(\{\omega\}|\varpi_i)$ are the respective conditional probabilities. If $Q(\varpi_i) = 0$ or $P(\varpi_i) = 0$, then $\mathsf{d}_{\mathfrak{H}|\mathcal{G}}(Q, P)(\varpi_i) = +\infty$.

It follows from (3.33) and (3.34) that if $Z \in \mathfrak{H}$, then for any Q, $P \in \mathfrak{P}$,

$$\left|\mathbb{E}_Q(Z) - \mathbb{E}_P(Z)\right| \le \mathsf{d}_{\mathfrak{H}}(Q, P), \tag{3.35}$$

and

$$\left|\mathbb{E}_{Q|\mathcal{G}}(Z) - \mathbb{E}_{P|\mathcal{G}}(Z)\right| \le \mathsf{d}_{\mathfrak{H}|\mathcal{G}}(Q, P). \tag{3.36}$$

Example 3.3 (Wasserstein distance (of order one))**.** Let $\rho\colon \Omega \times \Omega \to \mathbb{R}_+$ be a metric (distance) on the set Ω. Define the set $\mathfrak{H}$ of functions $h\colon \Omega \to \mathbb{R}$ as

$$\mathfrak{H} := \left\{h\colon |h(\omega) - h(\omega')| \le \rho(\omega, \omega') \text{ for all } \omega, \omega' \in \Omega\right\}. \tag{3.37}$$

By (3.33) this defines the respective distance, denoted $\mathsf{d}_{W_1}(Q, P)$, between Q, $P \in \mathfrak{P}$. By Kantorovich–Rubinstein theorem this is the dual representation of Wasserstein distance of order one discussed in the next section.

Consider a Lipschitz continuous function $Z\colon \Omega \to \mathbb{R}$ modulus $L > 0$, i.e.,

$$|Z(\omega) - Z(\omega')| \le L\,\rho(\omega, \omega') \text{ for all } \omega, \omega' \in \Omega.$$

Then $L^{-1}Z \in \mathfrak{H}$, and hence we have by (3.35) that $\left|\mathbb{E}_Q(L^{-1}Z) - \mathbb{E}_P(L^{-1}Z)\right|$ is less than or equal to $\mathsf{d}_{W_1}(Q, P)$, and consequently

$$\left|\mathbb{E}_Q(Z) - \mathbb{E}_P(Z)\right| \le L \cdot \mathsf{d}_{W_1}(Q, P). \tag{3.38}$$

It follows that for the distributionally robust functional $\mathcal{R}_\epsilon$ associated with the respective ambiguity set $\mathfrak{M}_\epsilon$ of the form (3.31), the inequality

$$\left|\mathcal{R}_\epsilon(Z) - \mathbb{E}_P(Z)\right| \le L\,\epsilon \tag{3.39}$$

inequality holds for Lipschitz continuous Z with modulus L.

In accordance with (3.30), consider the conditional counterpart $\mathsf{d}_{W_1|\mathcal{G}}$ of the above distance. Furthermore, assuming that $P(\varpi_i) > 0$ for all $i \in \mathcal{I}$, consider the corresponding sets $\mathfrak{A}_i$ of the form (3.32) and the associated conditional risk averse functional $\mathcal{R}_{\epsilon|\mathcal{G}}$. Similar to (3.38) and (3.39), we have for Lipschitz continuous $Z\colon \Omega \to \mathbb{R}$ with modulus $L > 0$, that

$$\left|\mathbb{E}_{Q|\mathcal{G}}(Z)(\varpi_i) - \mathbb{E}_{P|\mathcal{G}}(Z)(\varpi_i)\right| \le L \cdot \mathsf{d}_{W_1|\mathcal{G}}(Q, P)(\varpi_i),\ i \in \mathcal{I}, \tag{3.40}$$

and

$$\left|\mathcal{R}_{\epsilon|\mathcal{G}}(Z)(\varpi_i) - \mathbb{E}_{P|\mathcal{G}}(Z)(\varpi_i)\right| \le L\,\epsilon \tag{3.41}$$

on every atom ϖ_i, $i \in \mathcal{I}$, of $\mathcal{G}$. $\square$

3.5 Conditional Wasserstein distances

The general Wasserstein distance allows to assess the difference of probability measures, even for different measurable spaces. That is, consider two (finite) sets $\Omega = \{\omega_1, \dots, \omega_n\}$ and $\Omega' = \{\omega'_1, \dots, \omega'_{n'}\}$, equipped with the respective sigma algebras $\mathcal{F}$ and $\mathcal{F}'$ of all their subsets, and let $\rho\colon \Omega \times \Omega' \to \mathbb{R}_+$ be a non-negatively valued function. Denote by $\mathfrak{P}$ and $\mathfrak{P}'$ the corresponding sets of all probability measures on $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$. For $P \in \mathfrak{P}$ and $P' \in \mathfrak{P}'$ set $p_i = P(\{\omega_i\})$ and $p'_j = P'(\{\omega'_j\})$, and let $\Pi(P, P')$ be the set of $n \times n'$ matrices $\pi = [\pi_{ij}]$ with non-negative entries π_{ij}, such that

$$\sum\nolimits_{j=1}^{n'} \pi_{ij} = p_i, \quad i = 1, \dots, n, \tag{3.42}$$

$$\sum\nolimits_{i=1}^{n} \pi_{ij} = p'_j, \quad j = 1, \dots, n'. \tag{3.43}$$

Definition 3.8 (Wasserstein distance)**.** *The general Wasserstein distance, of order $s \geq 1$, between measures $P \in \mathfrak{P}$ and $P' \in \mathfrak{P}'$ is defined as*

$$\mathsf{dl}_s(P, P') := \left(\inf_{\pi \in \Pi(P,P')} \sum_{i=1}^{n} \sum_{j=1}^{n'} \pi_{ij}\, \rho(\omega_i, \omega'_j)^s \right)^{1/s}, \tag{3.44}$$

where π_{ij} satisfies the constraints (3.42) *and* (3.43)*.*

Note that the measures P and P' can be defined on different measurable spaces, in which case $\mathsf{dl}_s(P, P')$ cannot be symmetric. Therefore, in the standard sense, dl_s is not a distance. It becomes a distance if $\Omega' = \Omega$ and ρ is a metric (distance) on Ω (cf. Villani [25]). Anyway, we can proceed to construction of conditional (nested) counterparts of dl_s in the spirit of the previous section. Let $\mathcal{G}$ and $\mathcal{G}'$ be subalgebras of $\mathcal{F}$ and $\mathcal{F}'$, respectively, generated by the corresponding partitions $\{\varpi_i\}_{i \in \mathcal{I}}$ and $\{\varpi'_j\}_{j \in \mathcal{I}'}$ of Ω and Ω', respectively.

Consider the partition $\{\varpi_i \times \varpi'_j\}_{i \in \mathcal{I}, j \in \mathcal{I}'}$ of $\Omega \times \Omega'$. This partition generates the corresponding subalgebra $\mathcal{G} \otimes \mathcal{G}'$ of $\mathcal{F} \otimes \mathcal{F}'$. As before we can consider the conditional probabilities $P(\cdot \,|\, \varpi_i)$ for $P \in \mathfrak{P}$ and $P'(\cdot \,|\, \varpi'_j)$ for $P' \in \mathfrak{P}'$, provided $P(\varpi_i) \neq 0$ and $P'(\varpi'_j) \neq 0$. The respective probability measures $P_{|\varpi_i}$ and $P'_{|\varpi'_j}$ are defined in (2.4). Recall that measures $P_{|\varpi_i}$ and $P'_{|\varpi'_j}$ are supported on ϖ_i and ϖ'_j, respectively, $i \in \mathcal{I}$, $j \in \mathcal{I}'$.

Definition 3.9 (Conditional Wasserstein distance)**.** *The conditional Wasserstein distance between the measures $P \in \mathfrak{P}$ and $P' \in \mathfrak{P}'$ of order $s \geq 1$ and with respect to the function $\rho\colon \Omega \times \Omega' \to \mathbb{R}_+$, is a $(\mathcal{G} \otimes \mathcal{G}')$-measurable variable, denoted $\mathsf{dl}_{s|\mathcal{G}\otimes\mathcal{G}'}(P, P')$, taking the values*

$$\mathsf{dl}_{s|\mathcal{G}\otimes\mathcal{G}'}(P, P')(\varpi_i \times \varpi'_j) := \begin{cases} \mathsf{dl}_s(P_{|\varpi_i}, P'_{|\varpi'_j}) & \textit{if } P(\varpi_i) > 0 \ \textit{ and } \ P'(\varpi'_j) > 0, \\ +\infty & \textit{if } P(\varpi_i) = 0 \ \textit{ or } \ P'(\varpi'_j) = 0, \end{cases} \tag{3.45}$$

$i \in \mathcal{I}$, $j \in \mathcal{I}'$.

That is, let ϖ_i and ϖ'_j be such that $P(\varpi_i) > 0$ and $P'(\varpi'_j) > 0$, and $m_i := |\varpi_i|$ and $m'_j := |\varpi'_j|$ be the respective cardinalities. Consider the set $\Pi\big(P_{|\varpi_i}, P'_{|\varpi'_j}\big)$ of $m_i \times m'_j$ matrices $\pi = [\pi_{k\ell}]$ with nonnegative entries $\pi_{k\ell}$ such that

$$\sum\nolimits_{\ell=1}^{m'_j} \pi_{k\ell} = P_{|\varpi_i}(\omega_k), \; k = 1, \dots, m_i, \tag{3.46}$$

$$\sum\nolimits_{k=1}^{m_i} \pi_{k\ell} = P'_{|\varpi'_j}(\omega'_\ell), \; \ell = 1, \dots, m'_j. \tag{3.47}$$

Then the value of $\mathsf{dl}_{s|\mathcal{G}\otimes\mathcal{G}'}(P, P')$ on the atom (elementary event) $\varpi_i \times \varpi'_j$ is

$$\mathsf{dl}_{s|\mathcal{G}\otimes\mathcal{G}'}(P, P')(\varpi_i \times \varpi'_j) = \left(\inf_{\pi \in \Pi\left(P_{|\varpi_i}, P'_{|\varpi'_j}\right)} \sum_{\omega_k \in \varpi_i} \sum_{\omega'_\ell \in \varpi'_j} \pi_{k\ell}\, \rho(\omega_k, \omega'_\ell)^s \right)^{1/s}. \tag{3.48}$$

In particular, if $\mathcal{G} = \{\emptyset, \Omega\}$ and $\mathcal{G}' = \{\emptyset, \Omega'\}$ are trivial, then $\mathsf{dl}_{s|\mathcal{G}\otimes\mathcal{G}'}$ coincides with the respective Wasserstein distance dl_s.

Remark 3.6. Suppose that $\Omega' = \Omega$ and ρ is a metric (distance) on Ω. Then, as mentioned before, $\mathsf{dl}_s(Q, P)$ defines a distance between measures $Q, P \in \mathfrak{P}$. Suppose further that the subalgebras $\mathcal{G} = \mathcal{G}'$, i.e., the respective partitions do coincide. Then, in accordance with Definition 3.7, we can consider the conditional counterpart, denoted $\mathsf{d}_{W_s|\mathcal{G}}$, of the distance dl_s. That is, $\mathsf{d}_{W_s|\mathcal{G}}(Q, P)$ is a $\mathcal{G}$-measurable variable taking value $\mathsf{d}_{W_s|\mathcal{G}}(Q, P)(\varpi_i) = \mathsf{dl}_s(Q_{|\varpi_i}, P_{|\varpi_i})$ on ϖ_i, provided that $Q(\varpi_i)$ and $P(\varpi_i)$ are positive. The Wasserstein distance $\mathsf{dl}_s(Q_{|\varpi_i}, P_{|\varpi_i})$ is computed in accordance with (3.44) with matrices $\pi \in \Pi(Q_{|\varpi_i}, P_{|\varpi_i})$ consisting of nonnegative elements satisfying the constraints of the form (3.46)–(3.47). We obtain that

$$\mathsf{d}_{W_s|\mathcal{G}}(Q, P)(\varpi_i) = \mathsf{dl}_{s|\mathcal{G}\otimes\mathcal{G}}(Q, P)(\varpi_i \times \varpi_i). \tag{3.49}$$

In particular, for $s = 1$ we have by (3.49) that the conditional distance $\mathsf{dl}_{1|\mathcal{G}\otimes\mathcal{G}}(\varpi_i \times \varpi_i)$ coincides with the conditional distance $\mathsf{d}_{W_1|\mathcal{G}}(\varpi_i)$ discussed in Example 3.3. Consequently, by (3.40) the following inequality holds

$$\big|\mathbb{E}_{Q|\mathcal{G}}(Z)(\varpi_i) - \mathbb{E}_{P|\mathcal{G}}(Z)(\varpi_i)\big| \le L \cdot \mathsf{dl}_{1|\mathcal{G}\otimes\mathcal{G}}(Q, P)(\varpi_i \times \varpi_i), \; i = 1, \dots, m, \tag{3.50}$$

for Q, $P \in \mathfrak{P}$ and Lipschitz continuous modulus L variables $Z \colon \Omega \to \mathbb{R}$. $\square$

4 Multistage setting

Consider the measurable space $(\Omega, \mathcal{F})$. We assume in this section that Ω is *finite*. Consider a filtration $\mathfrak{G} = \{\mathcal{G}_t\}_{1 \le t \le T}$. That is, $\mathcal{G}_1 \subset \ldots \subset \mathcal{G}_T$ is a sequence of increasing subalgebras[6] of $\mathcal{F}$ with $\mathcal{G}_T = \mathcal{F}$. Such filtration defines the corresponding scenario tree. In order for the root of this scenario tree to be deterministic we can assume that $\mathcal{G}_1 = \{\emptyset, \Omega\}$ is trivial.

[6] Since Ω is finite, sigma additivity of the subalgebras $\mathcal{G}_t$ follows from their finite additivity.

4.1 Nested risk functionals

Denote by $\mathcal{Z}_t$ the space of $\mathcal{G}_t$-measurable variables $Z\colon \Omega \to \mathbb{R}$, $t = 1, \dots, T$, in particular $\mathcal{Z}_T = \mathcal{Z}$. Consider the distributionally robust functional $\mathcal{R}(\cdot)$, defined in (2.1), associated with the ambiguity set $\mathfrak{M} \subset \mathfrak{P}$.

In accordance with (2.2), for every $t = 1, \dots, T$, we can define the conditional distributionally robust counterpart of $\mathcal{R}$, that is

$$\mathcal{R}^\star_{|\mathcal{G}_t}(Z) \coloneqq \sup_{Q\in\mathfrak{M}} \mathbb{E}_{Q|\mathcal{G}_t}[Z]. \tag{4.1}$$

Note that $\mathcal{R}^\star_{|\mathcal{G}_t}(Z)$ is a $\mathcal{G}_t$-measurable variable, and thus $\mathcal{R}^\star_{|\mathcal{G}_t}$ can be viewed as a mapping from $\mathcal{Z}_T$ to $\mathcal{Z}_t$. Restricted to the space $\mathcal{Z}_{t+1}$, this can be viewed as a one-step conditional mapping denoted $\mathcal{R}^\star_{t|\mathcal{G}_t}\colon \mathcal{Z}_{t+1} \to \mathcal{Z}_t$. Consider the corresponding nested (composite) mapping

$$\mathfrak{R}^\star_{t|\mathcal{G}_t}(Z) \coloneqq \mathcal{R}^\star_{t|\mathcal{G}_t} \circ \cdots \circ \mathcal{R}^\star_{T-1|\mathcal{G}_{T-1}}(Z). \tag{4.2}$$

By the tower property of the expectation operator we have that

$$\sup_{Q\in\mathfrak{M}} \mathbb{E}_{Q|\mathcal{G}_t}[Z] = \sup_{Q\in\mathfrak{M}} \mathbb{E}_{Q|\mathcal{G}_t}\big[\cdots \mathbb{E}_{Q|\mathcal{G}_{T-1}}[Z]\big] \le \sup_{Q\in\mathfrak{M}} \mathbb{E}_{Q|\mathcal{G}_t}\big[\cdots \sup_{Q\in\mathfrak{M}} \mathbb{E}_{Q|\mathcal{G}_{T-1}}[Z]\big], \tag{4.3}$$

and hence it follows (cf. Shapiro [21]) that

$$\mathcal{R}^\star_{|\mathcal{G}_t}(Z) \le \mathfrak{R}^\star_{t|\mathcal{G}_t}(Z),\ t = 1, \dots, T-1. \tag{4.4}$$

It is also possible to consider risk averse conditional mappings $\mathcal{R}_{|\mathcal{G}_t}\colon \mathcal{Z}_T \to \mathcal{Z}_t$, of the form (2.7), for specified respective sets $\mathfrak{A}_{ti}$. Restricted to the space $\mathcal{Z}_{t+1}$ this becomes a one-step conditional mapping $\mathcal{R}_{t|\mathcal{G}_t}\colon \mathcal{Z}_{t+1} \to \mathcal{Z}_t$, and the corresponding nested (composite) mapping is (cf. Shapiro et al. [23, Section 6.7.1])

$$\mathfrak{R}_{t|\mathcal{G}_t}(Z) \coloneqq \mathcal{R}_{t|\mathcal{G}_t} \circ \cdots \circ \mathcal{R}_{T-1|\mathcal{G}_{T-1}}(Z). \tag{4.5}$$

It is interesting to note that the inequality $\mathcal{R}_{|\mathcal{G}_t}(Z) \le \mathfrak{R}_{t|\mathcal{G}_t}(Z)$, similar to (4.4), is not guaranteed for all $Z \in \mathcal{Z}$. Such counterexamples exist, for instance, for the $\mathsf{AV@R}$ risk measure (cf. Pflug and Pichler [15]).

Example 4.1 (ϕ-divergence)**.** For example, it is possible to use the self-similar approach of conditional ϕ-divergence, discussed in Section 3.3, for construction of the corresponding one-step mappings. That is let P be a reference probability measure on $(\Omega, \mathcal{F})$ and $\epsilon_t \ge 0$, $t = 1, \dots, T$, be chosen constants. In accordance with (3.26) define sets $\mathfrak{A}_{ti}$ associated with $\mathcal{G}_t$ and ϵ_t. This defines the corresponding conditional functional (mapping) $\mathcal{R}_{\epsilon_t|\mathcal{G}_t}$, and hence the respective nested mapping $\mathfrak{R}_{t|\mathcal{G}_t}$. □

Similarly to the nested risk functional (4.2), the nested distances constitute a sequence of distances $\mathsf{d} = (\mathsf{d}_t)_{t=1}^T$, each d_t being measurable with respect to the subalgebra $\mathcal{G}_t$. In this way, nested distances resume the increasing structure of the subalgebras $\mathcal{G}_1 \subset \cdots \subset \mathcal{G}_T$. In the next section we discuss such construction with respect to the Wasserstein distance.

4.2 Nested Wasserstein distances

It is demonstrated above that distances of probability measures interact with risk measures, cf. (3.35) and (3.36). In the same way, conditional distances govern conditional risk measures, cf. (3.40) and (3.41) or (3.50).

In a multistage context, the nested risk measures are defined in (4.2). The corresponding inequalities, which bound the nested risk measures, remain valid by extending the distances to the nested distances. Introduced in Pflug [13], they extend the distances from random variables to distances of stochastic processes. The following section relates them to the conditional, nested risk functionals introduced above.

The multistage setting allows to consider the (general) Wasserstein distance $\mathsf{dl}_s(P, P')$ of the probability measures $P \in \mathfrak{P}$ and $P' \in \mathfrak{P}'$ (cf. Definition 3.8). The nested distance takes them into account and provides a process distance and is defined recursively.

Definition 4.1 (Nested Wasserstein distance). *Let $\big(\Omega, (\mathcal{G}_t)_{t=1}^T, P\big)$ and $\big(\Omega', (\mathcal{G}'_t)_{t=1}^T, P'\big)$ be filtered probability spaces. The nested distance is the stochastic process*

$$\mathsf{dl}_{s|\mathcal{G}_t\otimes\mathcal{G}'_t}(P, P'), \quad t = 1, \dots, T,$$

where $\mathsf{dl}_{s|\mathcal{G}_t\otimes\mathcal{G}'_t}(P, P')(\varpi_i, \varpi'_j)$ is the conditional Wasserstein distance (cf. Definition 3.9) using the distance $\mathsf{dl}_{s|\mathcal{G}_{t+1}\otimes\mathcal{G}'_{t+1}}(P, P')$ as the transportation cost function on the next stage $t+1$; at the final stage T, the distance is $\mathsf{dl}_{s|\mathcal{G}_T\otimes\mathcal{G}'_T} = \rho$.

Assume now that $\Omega = \Omega'$, $\mathcal{G}_t = \mathcal{G}'_t$ for all t, and ρ is a metric (distance) on Ω. The recursively defined nested distance and the recursive maps $\mathfrak{R}_t$ in (4.5) collaborate naturally. It is also possible to use the distance approach discussed in Section 3.4. In particular, consider the setting of Example 3.3. For a reference probability measure $P \in \mathfrak{P}$, consider the sets $\mathfrak{A}_{ti}$ of the form (3.32) associated with conditional distances $\mathsf{d}_{W_1|\mathcal{G}_t}$ and constants $\epsilon_t \geq 0$, $t = 1, \dots, T$. This defines the respective conditional functionals $\mathcal{R}_{\epsilon_t|\mathcal{G}_t}$, and hence the respective nested mapping $\mathfrak{R}_{t|\mathcal{G}_t}$. For the nested risk functional, we have the following regularity result.

Corollary 4.1 (Wasserstein distance (of order one)). *By* (3.41) *we have that for Lipschitz continuous modulus $L > 0$ function $Z\colon \Omega \to \mathbb{R}$, the following inequality holds*

$$\big|\mathcal{R}_{\epsilon_t|\mathcal{G}_t}(Z)(\cdot) - \mathbb{E}_{P|\mathcal{G}_t}(Z)(\cdot)\big| \leq L\,\epsilon_t, \tag{4.6}$$

and hence

$$\big|\mathfrak{R}_{\epsilon_t|\mathcal{G}_t}(Z)(\cdot) - \mathbb{E}_{P|\mathcal{G}_t}(Z)(\cdot)\big| \leq L(\epsilon_t + \dots + \epsilon_T). \tag{4.7}$$

Proof. We shall demonstrate the result for $\mathcal{G}$ at stage $t = 2$ first, as it generalizes analogously to all other stages. We have that

$$\mathcal{R}_{\epsilon_1}\big(\mathcal{R}_{\epsilon_2|\mathcal{G}}(Z)\big) - \mathbb{E}_P(Z) = \mathcal{R}_{\epsilon_1}\big(\mathcal{R}_{\epsilon_2|\mathcal{G}}(Z)\big) - \mathcal{R}_{\epsilon_1}\big(\mathbb{E}_{|\mathcal{G}}(Z)\big) + \mathcal{R}_{\epsilon_1}\big(\mathbb{E}_{|\mathcal{G}}(Z)\big) - \mathbb{E}_P\,\mathbb{E}_{|\mathcal{G}}(Z). \tag{4.8}$$

With (3.41) it holds that

$$\big|\mathcal{R}_{\epsilon_2|\mathcal{G}}(Z)(\varpi_i) - \mathbb{E}_{|\mathcal{G}}(Z)(\varpi_i)\big| \leq L \cdot \epsilon_2,$$

and by using monotonicity and translation invariance of $\mathcal{R}(\cdot)$ we conclude that

$$\left|\mathcal{R}_{\epsilon_1}\big(\mathcal{R}_{\epsilon_2|\mathcal{G}}(Z)\big) - \mathcal{R}_{\epsilon_1}\big(\mathbb{E}_{|\mathcal{G}}(Z)\big)\right| \le L\cdot\epsilon_2.$$

Further we have that

$$\left|\mathbb{E}_{P(\cdot|\varpi_i)}(Z) - \mathbb{E}_{P'(\cdot|\varpi_i)}(Z)\right| \le L\cdot \mathsf{d}_{W_1}\big(P(\cdot|\varpi_i), P'(\cdot|\varpi_i)\big),\ i=1,\dots,n.$$

That is, the variable

$$\varpi \mapsto \mathbb{E}_{P(\cdot|\varpi)}(Z)$$

is Lipschitz continuous with the constant L with respect to the distance on the atoms ϖ and ϖ', that is, the distance on the subsequent stage. It follows that

$$\mathcal{R}_{\epsilon_1}\big(\mathbb{E}_{|\mathcal{G}}(Z)\big) - \mathbb{E}_P\,\mathbb{E}_{|\mathcal{G}}(Z) \le L\cdot\epsilon_1.$$

The assertion follows from (4.8) by combining the above partial results. □

5 Summary and outlook

The risk averse extension of optimization under uncertainty and stochastic optimization naturally build on risk functionals. The modern theory of risk measures is well-developed and is based on two fundamental assumptions; namely, the existence of the reference probability measure and the absolute continuity, with respect to the reference measure, of all probability measures of the involved ambiguity sets (cf. Föllmer and Schied [7], Shapiro et al. [23]). The nested risk averse approach to stochastic programming was discussed already in Ruszczyński and Shapiro [19]. However, there are important examples in the distributionally robust settings where these basic assumptions do not hold. Without the absolute continuity assumption it is not clear how to define the conditional counterpart of the distributional robust functional (2.1).

This paper resumes the discussion of mathematical properties of the distributionally robust functionals, as their definition – from a mathematical perspective – is not entirely obvious and several versions are plausible. We assumed first that the considered subalgebras are finite, and hence in the nested formulation the corresponding scenario tree is finite. In various applications this is a natural modelling assumption. An extension to a general setting was outlined in Section 3. We demonstrated the connection between nested distributionally robust functionals and conditional formulations of (Wasserstein) distances between probability measures.

References

[1] G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. *Tutorials in Operations Research, INFORMS*, pages 1563–1581, 2015.

[2] A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17:449–476, 2007.

[3] I. Csiszár. Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von markoffschen Ketten. *Magyar. Tud. Akad. Mat. Kutato Int. Kozls*, 8:85–105, 1963.

[4] C. Dellacherie and P.-A. Meyer. *Probabilities and Potential.* North-Holland Publishing Co., Amsterdam, The Netherlands., 1988. URL `https://projecteuclid.org/euclid.bams/1183546371`.

[5] D. Ellsberg. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669, 1961. doi: 10.2307/1884324.

[6] A. Esteban-Pérez and J. M. Morales. Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming*, 2021. doi: 10.1007/s10107-021-01724-0. Adrian Esteban-Perez.

[7] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time.* de Gruyter Studies in Mathematics 27. Berlin, Boston: De Gruyter, 2004. ISBN 978-3-11-046345-3. doi: 10.1515/9783110218053. URL `http://books.google.com/books?id=cL-bZSOrqWoC`.

[8] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann. A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62, 2015. doi: 10.1007/s10107-015-0896-z.

[9] A. J. Keith and D. K. Ahner. A survey of decision making and optimization under uncertainty. *Annals of Operations Research*, 300(2):319–353, 2019. doi: 10.1007/s10479-019-03431-8.

[10] F. H. Knight. *Risk, Uncertainty and Profit.* Library of Economics and Liberty, 1921.

[11] Y. Liu, A. Pichler, and H. Xu. Discrete approximations and quantification in distributionally robust optimization. *Mathematics of Operations Research*, 44(1), 2018. doi: 10.1287/moor.2017.0911.

[12] T. Morimoto. Markov processes and the H-Theorem. *Journal of the Physical Society of Japan*, 18(3):328–331, 1963. doi: 10.1143/jpsj.18.328.

[13] G. Ch. Pflug. Version-independence and nested distributions in multistage stochastic optimization. *SIAM Journal on Optimization*, 20:1406–1420, 2009. doi: 10.1137/080718401.

[14] G. Ch. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012. doi: 10.1137/110825054.

[15] G. Ch. Pflug and A. Pichler. Time-inconsistent multistage stochastic programs: Martingale bounds. *European Journal of Operational Research*, 249(1):155–163, 2015. doi: 10.1016/j.ejor.2015.02.033.

[16] A. Pichler and A. Shapiro. Mathematical foundations of distributionally robust multistage optimization. *SIAM Journal on Optimization*, 31(4):3044–3067, 2021. doi: 10.1137/21M1390517.

[17] H. Rahimian and S. Mehrotra. Distributionally robust optimization: A review, 2021. URL `https://arXiv.org/abs/1908.05659`.

[18] F. Riedel. Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112(2):185–200, 2004. doi: 10.1016/j.spa.2004.03.004.

[19] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31(3):544–561, 2006. doi: 10.1287/moor.1060.0204.

[20] H. Scarf. A min-max solution of an inventory problem. In *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, 1958.

[21] A. Shapiro. Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research*, 219(3):719–726, 2012. doi: 10.1016/j.ejor.2011.11.005.

[22] A. Shapiro. Interchangeability principle and dynamic equations in risk averse stochastic programming. *Operations Research Letters*, 45(4):377–381, jul 2017. doi: 10.1016/j.orl.2017.05.008.

[23] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. MOS-SIAM Series on Optimization. SIAM, third edition, 2021. doi: 10.1137/1.9781611976595.

[24] H. Sun and H. Xu. Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(2):377–401, 2016. doi: 10.1287/moor.2015.0732. URL `http://dx.doi.org/10.1287/moor.2015.0732`.

[25] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. ISBN 0-821-83312-X. doi: 10.1090/gsm/058. URL `http://books.google.com/books?id=GqRXYFxe0l0C`.

[26] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.