RISK AND UTILITY IN THE DUALITY FRAMEWORK OF CONVEX ANALYSIS

R. Tyrrell Rockafellar[1]

Abstract

Measures of risk have grown in importance in expressing preferences between different manifestations of uncertain cost or loss in finance and engineering, but utility functions and expected utility have had a more traditional role. This article surveys how risk and utility are in fact more closely related than may have been appreciated by practitioners. The tools of convex analysis, including conjugate duality, are able to bring this out.

Keywords: *preferences under uncertainty, risk quantification, utility quantification, regret quantification, stochastic ambuiguity, distributional robustness, convex analysis, duality, quantiles, superquantiles*

July 23, 2018

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: http://www.math.washington.edu/~rtr/mypage.html

1 Preferences under uncertainty

Utility functions were developed in economics for the purpose of describing the preferences of individuals among various bundles of goods. Their basis in axioms involved choosing between "lotteries" in which the quantities in a bundle would be obtained with different probabilities. Utility theory, from that angle, was always connected with preferences under uncertainty, but in more recent times it has been important also in guiding financial decisions about portfolios of stocks and bonds. A portfolio put together on a given date at a given cost will have, at a targeted future date, a monetary value that can be viewed as a random variable. Preferences among various portfolio choices come down then to preferences among random variables of returns, and an important approach to that has been to make comparisons in terms of the expected utility of those returns with respect to a utility function for money.

Utility has not been the only way of looking at preferences in finance, however. An increasingly popular alternative has been the quantification of the risk inherent in a random variable, most conveniently oriented in this case not on the attractiveness of rewards but rather on the distaste for losses. There is by now a well developed theory of such quantification, which like utility theory can capture how the attitudes of one decision maker may differ from those of another.

Are these two notions, risk and utility, in competition, or do they fit together harmoniously in some larger pattern? Convex analysis provides a mathematical framework in which they not only fit together but also interact productively. Ben-Tal and Teboulle [5] were the first to recognize this on a fundamental level, but the subject has since been pushed into broader territory by Rockafellar and Uryasev in [20]. Here we offer a consolidated presentation which emphasizes conjugate duality and provides some new results. Duality famously ties risk to the contemplation of sets of probability measures instead of just a single specified probability measure, as first was demonstrated by Artzner et al. in [2]. Our framework is able from that perspective to coordinate such stochastic ambiguity, as applied to expected utility, with particular forms of risk, and this is one of the new contributions made here. It furthermore offers a different perspective on utility which is better suited to comparisons to a benchmark than the customary approach in finance, as seen for example in the treatment in references like the book of Föllmer and Schied [7].

To begin, we have to set the stage with a probability space $(\Omega, \mathcal{F}, P_0)$ where Ω is the set of "scenarios" or "future states" ω, $\mathcal{F}$ is a field of subsets of Ω, and P_0 is a probability measure on $\mathcal{F}$. The reason for the subscript 0 is that we will come to comparing other probability measures P to P_0, which itself may in general just be a nominal choice, perhaps arising empirically or from subjective guesswork. Any measureable function $X : \Omega \to I\!R$ can be interpreted as a *random variable* for which the cumulative distribution function F_X from $(-\infty, \infty) \to [0, 1]$ is given by

$$F_X(\tau) = \text{prob}[X(\omega) \leq \tau] = P_0[\{\, \omega \in \Omega \mid X(\omega) \leq \tau \}]. \tag{1.1}$$

The traditional spaces of random variables are

$$\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, P_0) = \begin{cases} \{\, X \mid E[|X|^p] < \infty \} & \text{for } 1 \leq p < \infty, \\ \{\, X \mid \sup |X| < \infty \} & \text{for } p = \infty, \end{cases} \tag{1.2}$$

where the expectation E of a random variable is its integral with respect to the probability measure P_0, and sup refers to the essential supremum of a function. But which of these might be "best" to work with?

The answer is not simple here because of the usual duality of $\mathcal{L}^p$ versus $\mathcal{L}^q$. The problem is that through consideration of alternative probability measures P we are led to restricting P to be absolutely continuous with respect to P_0 and expressing the expected value of X with respect to P in terms of the Radon-Nikodym derivative dP/dP_0 as

$$E_P[X] = E\Big[X\frac{dP}{dP_0}\Big] = \int_\Omega X(\omega)\frac{dP}{dP_0}(\omega)dP_0(\omega). \tag{1.3}$$

For $X \in \mathcal{L}^1$ this dictates restricting P to the class of probability measures such that $dP/dP_0 \in \mathcal{L}^\infty$, which is severe. Embracing the broadest generality in probability measures by having the derivatives dP/dP_0 be in $\mathcal{L}^1$, on the other hand, would seem to limit the random variables X to be essentially bounded, also perhaps not a very good idea.

In the face of this dilemma, we will follow [20] in taking $\mathcal{L}^2$ as the space to work with, the inner product there being

$$\langle X, Q\rangle = E[XQ] = \int_\Omega X(\omega)Q(\omega)dP_0(\omega). \tag{1.4}$$

For random variables $X \in \mathcal{L}^2$, of course, both the mean and variance,

$$\mu(X) = E[X], \qquad \sigma^2(X) = E[(X-\mu(X))^2] = E[X^2] - E[X]^2, \tag{1.5}$$

are well-defined and finite; and in fact this characterizes our choice. The probability measures P that will come into considerations as alternatives to P_0 will be those that are absolutely continuous with respect to P_0 and have densities dP/dP_0 belonging to $\mathcal{L}^2$. The set of all such density functions forms the "unit simplex" $\mathcal{P}_0$ of $\mathcal{L}^2$,

$$\begin{aligned}\mathcal{P}_0 &= \{\, Q \in \mathcal{L}^2 \,|\, Q \geq 0,\ E[Q] = 1\} \\ &= \{\, dP/dP_0 \in \mathcal{L}^2 \,|\, P \text{ absolutely continuous w.r.t. } P_0\}\end{aligned} \tag{1.6}$$

A sort of misalignment in our topic between minimization and maximization comes from the orientation of risk toward losses and utility toward gains. This will be reconciled later by inserting "regret" as an anti-utility, or disutility, suitable for minimization instead of maximization, but for now we will proceed just with risk.

2 How should "risk" be understood?

In thinking of a random variable X as standing for loss, or cost, or various other things that a decision maker would want to be lower[2] rather than higher, with negative values interpreted as "good," we come to the idea of quantifying risk as opposed to uncertainty in a random variable X. Risk definitely is involved with uncertainty, but there is an important distinction. An example is afforded by a lottery that offers the loss of \$1,000,000 with probability .99 and a loss of \$-1 (a reward of \$1) with probability .01. The uncertainty is very small, but the risk of participating in the lottery, as we are thinking about it here, is nearly \$1,000,000.

By a *risk quantifier*[3] we will mean a functional $\mathcal{R}$ that assigns to a random variable X a value $\mathcal{R}(X) \in (-\infty, \infty]$ standing as a surrogate for the "overall risk" deemed to be present in X. Some immediate and commonly employed examples are

$$\begin{aligned}&\mathcal{R}(X) = E[X] = \text{ expected loss ("average" loss)},\\ &\mathcal{R}(X) = \sup X = \text{ worst-case loss},\\ &\mathcal{R}(X) = E[X] + \lambda\sigma^2(X) = \text{ mean/variance loss with parameter } \lambda > 0,\\ &\mathcal{R}(X) = q_\alpha(X) = \text{ quantile loss at probability level } \alpha \in (0,1),\end{aligned} \tag{2.1}$$

[2] Often people speak of smaller instead of lower, but smaller means "closer to 0," while lower means "closer to $-\infty$."

[3] The usual terminology in finance is "measure of risk" or "risk measure," but here we experiment with trying to avoid conflict with the standard concept of a "measure" in mathematics.

where the α-quantile of X is defined by

$$q_\alpha(X) = \min\{\,\tau \mid F_X(\tau) \geq \alpha\} \tag{2.2}$$

and characterized in probability by

$$q_\alpha(X) \leq \tau \quad \Longleftrightarrow \quad \text{prob}[X > \tau] < 1 - \alpha. \tag{2.3}$$

In finance, the quantile loss is called the *value-at-risk* at the probability level α and denoted by $\mathrm{VaR}_\alpha(X)$. A less obvious concept, closely related to to the latter, is

$$\mathcal{R}(X) = \overline{q}_\alpha(X) = \text{ superquantile loss at level } \alpha \in (0,1), \tag{2.4}$$

which was developed as the *conditional value-at-risk*, $\mathrm{CVaR}_\alpha(X)$, in [18, 19] in defining it to be the average of X in its upper α-tail[4] but can also can be expressed by the averaging formula[5]

$$\overline{q}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 q_\beta(X)d\beta \tag{2.5}$$

and most valuably by

$$\overline{q}_\alpha(X) = \min_{C \in R}\left\{C + \frac{1}{1-\alpha}E[\,\max\{0, X - C\}]\right\}. \tag{2.6}$$

The last is a type of risk formula that will have an important role later in connecting risk to utility. An interesting feature of the expression in (2.6) is that the quantile $q_\alpha(X)$ is a value of C that attains the minimum,[6] so the same one-dimensional optimization problem serves to calculate both, and furthermore does so with no need of delving into conditional expectations.[7] The term "superquantile" as a substitute for "conditional value-at-risk" was proposed in [14] as being more suitable for applications outside of finance where "quantile" was already preferred to "value-at-risk."

Of course, not every functional $\mathcal{R}$ would deserve to be called a risk quantifier. What properties might be required in general, and to what extent do the examples just given meet the test? This issue was forcefully raised in finance by Artzner et al. in [2] in a critique of the widespread use of value-at-risk in assessing the safety of portfolios and in particular banking reserves. They took for the first time an axiomatic approach to risk and argued that a valid quantifier should be *coherent* in the sense of satisfying[8]

$$\begin{array}{ll} (r1) & \mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X'), \\ (r2) & \mathcal{R}(\lambda X) = \lambda \mathcal{R}(X) \;\text{ when }\; \lambda > 0, \\ (r3) & \mathcal{R}(X) \leq \mathcal{R}(X') \;\text{ when }\; X \leq X', \\ (r4) & \mathcal{R}(X + C) = \mathcal{R}(X) + C \;\text{ for constants } C, \end{array} \tag{2.7}$$

where $X \leq X'$ means that $\text{prob}[X > X'] = 0$. For them, axiom (r1) was critical under the interpretation that $\mathcal{R}(X)$ stands for the amount of cash that should be held in reserve to cover possible losses in a portfolio X. If two portfolios X and X' had $\mathcal{R}(X + X') > \mathcal{R}(X) + \mathcal{R}(X')$, that would signify

[4] This is the conditional expectation of X with respect to X exceeding its α-quantile as long as the probability of X actually equaling its α-quantile is 0. But when that probability is not 0 the definiion involves a more careful interpretation of the α-tail distribution associated with X, as set forth in [19].

[5] On the basis of this formula, the concept has been called "average value-at-risk" by Föllmer and Schied [7].

[6] In general the argmin set, if not actually a singleton, is a closed interval having the quantile as its left endpoint.

[7] The intimate connection between this formula and the very concept of the cumulative distribution function for a random variable X has been explained in [15].

[8] They couched their conditions in a context of gains instead of losses, as we express them here.

that the portfolio obtained by combining them required more cash in reserve than the uncombined portfolios and thus was actually riskier. That situation, counter to the virtues of diversification, is one of the big troubles they identified in value-at-risk, i.e., in using quantiles to quantify risk. It is not the only trouble, though: quantiles can behave discontinuously, even inevitably so in the case of a *finite* probability space, i.e, when Ω is a finite discrete set.

Axioms (r1) and (r2) together are equivalent to $\mathcal{R}$ being *sublinear* in the sense that

$$\mathcal{R}\Big(\sum_{i=1}^{m} \lambda_i X_i\Big) \leq \sum_{i=1}^{m} \lambda_i \mathcal{R}(X_i) \ \text{ for } \ \lambda_i > 0, \tag{2.8}$$

and then, in particular, $\mathcal{R}$ is *convex*. That is where convex analysis was first brought into risk theory in finance. Axiom (r3) seems utterly desirable, but in fact it stands in the way of another practice long followed in portfolio optimization in finance, namely concentrating on mean-variance loss criteria. With that, it is actually possible to have X deemed less risky than X' even though the outcomes of X are worse than the outcomes of X' with probability 1! Actually, mean-variance loss does not even satisfy (r1) or (r2), but that can be remedied by switching to *mean-devation* loss, obtained by putting the standard deviation $\sigma(X)$ in place of the variance σ^2. On the other hand, it can be questioned whether the linear scaling axiom (r2) is really appropriate, the suggestion being that the sublinearity (r1)+(r2) might better be replaced by the direct assumption of convexity.

The ground-breaking coherency axioms (2.7) were only articulated in [2] for finite-valued $\mathcal{R}$, no ∞ admitted, and for a *finite* probability space, in which case the $\mathcal{L}^p$ spaces of random variables are finite-dimensional and differ only in norm (not topology). Then the convexity of $\mathcal{R}$ coming from the sublinearity in (r1)+(r2) implies the *continuity* of $\mathcal{R}$. To cover general probability spaces, however, $\mathcal{R}(X)$ should be allowed to be ∞, with the case of $\mathcal{R}(X) = \sup X$ being a prime candidate, and then a topological assumption, namely lower semicontinuity, needs to brought in to serve in place of the no-longer-automatic continuity.

These reflections on the subject lead to a broader set of axioms:

$$\begin{array}{ll} (R1) & \mathcal{R} \ \text{ is lower semicontinuous, possibly taking on } \infty, \\ (R2) & \mathcal{R}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{R}(X) + \lambda\mathcal{R}(X') \ \text{ for } \ \lambda \in (0,1), \\ (R3) & \mathcal{R}(X) \leq \mathcal{R}(X') \ \text{ when } \ X \leq X', \\ (R4) & \mathcal{R}(C) = C \ \text{ for constants } \ C, \end{array} \tag{2.9}$$

with (R2) being the explicit assumption of convexity. Under that convexity the simpler (and easier to defend) condition (R4) in place of (r4) is actually equivalent to (r4), cf. [20]. An additional condition to contemplate in the mix as a sharpening of (R4) is

$$(R4') \quad \mathcal{R}(X) > E[X] \ \text{ for nonconstant } X, \tag{2.10}$$

which is called *aversity* and insists on a positive "risk premium" being tied to uncertainty. It was first brought up in [21] and turned out to be crucial for the scheme developed in [20].

Where does this leave us with examples? Mean/variance and mean/deviation fail (R3) as with (r3), and quantile risk fails (R2) as it previously failed (r1). The "risk neutral" quantifier $\mathcal{R}(X) = E[X]$ satisfies (R1)–(R4) but lacks the aversity in (R4′). But superquantile risk and worst-case risk (its limit as $\alpha \nearrow 1$) satisfy *all* the conditions. Another, quite different looking example satisfying all the conditions is

$$\mathcal{R}(X) = \log E[\exp X] = \ \text{ log-exponential loss}, \tag{2.11}$$

which, in contrast to the ones just mentioned is truly just convex and not also sublinear, i.e., does not have (r1)+(r2). Furthermore any sum

$$\mathcal{R}(X) = \lambda_1 \mathcal{R}_1(X) + \cdots + \lambda_m \mathcal{R}_m(X) \ \text{ with } \ \lambda_i > 0, \ \lambda_1 + \cdots + \lambda_m = 1, \tag{2.12}$$

in which each $\mathcal{R}_i$ satisfies (R1)–(R4) again satisfies (R1)–(R4), and if at least one $\mathcal{R}_i$ has the aversity in (R4′), then $\mathcal{R}$ has it as well.[9]

For the broad class of risk quantifiers characterized by (R1)–(R4) and maybe also (R4′), a fundamental issue of convex analysis can be raised, namely duality. Whenever we have a convex functional $\mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty]$ that is lower semicontinuous and $\not\equiv \infty$, the formula

$$\mathcal{R}^*(Q) = \sup_X \{ \langle X, Q \rangle - \mathcal{R}(X)\} \tag{2.13}$$

defines a *conjugate* convex functional $\mathcal{R}^* : \mathcal{L}^2 \to (-\infty, \infty]$ that likewise is lower semicontinuous and $\not\equiv \infty$, and then $\mathcal{R}^{**} = \mathcal{R}$, or in other words,

$$\mathcal{R}(X) = \sup_Q \{ \langle X, Q \rangle - \mathcal{R}^*(Q)\}, \tag{2.14}$$

For our class of risk quantifiers $\mathcal{R}$, what are the conjugates $\mathcal{R}^*$ and what do they tell us about risk?

The answer is beautifully informative and, in the domain of conjugate duality — with its extensive catalog of dualization of properties — is a routine exercise to obtain, although such background was not familiar to researchers in finance. It takes us back to considering alternative probability measures P to our nominal probability measure P_0 in the context of their densities $Q = dP/dP_0$ in the probability simplex $\mathcal{P}_0$ in (1.6).

Theorem 1: risk dualization. *The class of functionals $\mathcal{J}$ on $\mathcal{L}^2$ that come up as conjugates $\mathcal{R}^*$ of risk quantifiers $\mathcal{R}$ satisfying (R1)–(R4) is characterized by*

$$\begin{array}{ll} (J1) & \mathcal{J} \text{ is lower semicontinuous,} \\ (J2) & \mathcal{J}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{J}(X) + \lambda \mathcal{J}(X') \text{ for } \lambda \in (0,1), \\ (J3) & \mathcal{P}_0 \supset \mathcal{Q} = \operatorname{dom} \mathcal{J} = \{ Q \mid \mathcal{J}(Q) < \infty \}, \\ (J4) & \inf\limits_{Q \in \mathcal{Q}} \mathcal{J}(Q) = 0. \end{array} \tag{2.15}$$

The extra condition (R4′) corresponds in this to

$$(J4') \quad \mathcal{J}(1) = 0 \text{ (entailing } 1 \in \mathcal{Q}\text{), but } \partial \mathcal{J}(1) \text{ contains no nonconstant } X. \tag{2.16}$$

The subclass consisting of the conjugates $\mathcal{R}^*$ of functionals $\mathcal{R}$ which are sublinear instead of just convex as in (R2) is identified with $\mathcal{J}$ being $\equiv 0$ on $\operatorname{dom} \mathcal{J}$. This subclass thus consists of the functionals $\mathcal{J}$ of the form

$$\mathcal{J} = \delta_{\mathcal{Q}} \text{ for some nonempty closed convex set } \mathcal{Q} \subset \mathcal{P}_0, \tag{2.17}$$

where $\delta_{\mathcal{Q}}$ denotes, as usual, the indicator of $\mathcal{Q}$, having the value 0 on $\mathcal{Q}$ but ∞ outside.

Proof. For the most part, these facts are contained in the Envelope Theorem of [20] and also known elsewhere, but the treatment of (J4) and (J4′) has not been seen in this form. The equivalence of (J4) with (R4) under conjugacy is obvious from having $\mathcal{R}(X) = \sup_Q \{ E[XQ] - \mathcal{J}(Q)\}$ in taking $X = C$ for a constant C, since $E[CQ] = C$. The "1" in (J4′) refers to the constant function 1 as an element of $\mathcal{L}^2$, which is the density dP_0/dP_0. Since $\mathcal{R}(X) = \sup_Q \{ E[XQ] - \mathcal{J}(Q)\}$, having $\mathcal{J}(1) = 0$ says in conjunction with (J3) that $\mathcal{R}(X) \geq E[X]$. Subgradients of $\mathcal{J}$ are defined by

$$X \in \partial \mathcal{J}(Q) \iff \mathcal{J}(Q') \geq \mathcal{J}(Q) + \langle X, Q' - Q \rangle = \mathcal{J}(Q) + E[XQ'] - E[XQ] \text{ for all } Q',$$

or equivalently through conjugacy, $\mathcal{R}(X) = E[XQ] - \mathcal{J}(Q)$, so having $X \in \partial \mathcal{J}(1)$ with $\mathcal{J}(1) = 0$ amounts to having $\mathcal{R}(X) = E[X]$. Forbidding this for nonconstant X corresponds to (R4′). □

[9] This rule carries over from discrete sums to "continuous sums" with respect to a "weighting measure," see [1, 21, 17]

The main revelation here is that selecting a risk quantifier $\mathcal{R}$ corresponds in a unique way through duality to selecting a nonempty convex subset set $\mathcal{Q}$ of $\mathcal{P}_0$ along with a nonnegative convex expression $\mathcal{J}(Q)$ on $\mathcal{Q}$ such that (J4) holds and the level sets $\{\, Q \in \mathcal{Q} \,|\, \mathcal{J}(Q) \leq c\}$ for $c \in I\!R$ are closed.[10] In this way, from (2.14), a representation of $\mathcal{R}$ is obtained in the form

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} \{\, E[XQ] - \mathcal{J}(Q)\} \tag{2.18}$$

which has a rewarding interpretation in probability. As a subset of the probability simplex $\mathcal{P}_0$ in (1.6), $\mathcal{Q}$ corresponds to *a set* $\mathcal{P}$ *of probability measures* for which (2.18) can be written via (1.3) as

$$\mathcal{R}(X) = \sup_{P \in \mathcal{P}} \{\, E_P[X] - \mathcal{J}(dP/dP_0)\}. \tag{2.19}$$

When $\mathcal{R}$ is sublinear, we are in the case of (2.17) and the representation comes down to

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} E[XQ] = \sup_{P \in \mathcal{P}} E_P[X]. \tag{2.20}$$

Having $1 \in \mathcal{Q}$ as in (J4′) translates to having $P_0 \in \mathcal{P}$, inasmuch as $dP/dP_0 = 1$ means $P = P_0$.

The formula $\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}_0} E[XQ]$ for an arbitrary subset $\mathcal{Q}_0 \neq \emptyset$ of $\mathcal{P}_0$, not necessarily convex and possibly just finite, still yields a risk quantifier satisfying (R1)–(R4), and different choices of such $\mathcal{Q}_0$ can give the same $\mathcal{R}$. But then (2.20) holds also for $\mathcal{Q}$ being the closed convex hull of $\mathcal{Q}_0$, which is uniquely the largest set of densities Q that can serve in this manner.

Dualization in the form of (2.20) holding for a closed convex set $\mathcal{Q} \subset \mathcal{P}_0$ is illustrated by

$$\begin{array}{lcl} \mathcal{R}(X) = E[X] & \longleftrightarrow & \mathcal{Q} = \{1\}, \\ \mathcal{R}(X) = \sup X & \longleftrightarrow & \mathcal{Q} = \mathcal{P}_0, \\ \mathcal{R}(X) = \overline{q}_\alpha(X) & \longleftrightarrow & \mathcal{Q} = \{\, Q \in \mathcal{P}_0 \,|\, Q \leq (1-\alpha)^{-1}\}. \end{array} \tag{2.21}$$

An example in the broader picture of $\mathcal{J}$ not just being an indicator function as in (2.17) is

$$\mathcal{R}(X) = \log E[\exp X] \quad \longleftrightarrow \quad \mathcal{Q} = \mathcal{P}_0,\ \mathcal{J}(Q) = E[Q \log Q] \ \text{ on } \ \mathcal{P}_0, \tag{2.22}$$

with $Q(\omega)\log Q(\omega)$ taken to be 0 when $Q(\omega) = 0$. This is striking because

$$E[Q \log Q] \ \text{ for } \ Q = dP/dP_0 \ \text{ is the Kullback-Leibler distance of } P \text{ from } P_0, \tag{2.23}$$

also known as the *relative entropy* of P with respect to P_0. Many more examples are available in [20]; see also [17]. Observe that in the general case of when $\mathcal{J}$ is not just an indicator, (J4′) makes $\mathcal{J}(dP/dP_0)$ give a sort of penalty related to how far P is from P_0; the Kullback-Leibler distance in (2.23) is just one illustration of that phenomenon.

The chief lesson in this is that the systematic approach to "risk" as axiomatized by (R1)-(R4) intrinsically leads through duality to the kind of worst-case analysis in (2.19), or more specially (2.20), in which some collection of probability measures is brought into action, not just P_0, to achieve "robustness." This brings out a fundamental connection betweek risk quantifiers and "robust optimization," which is a term that has become popular for problem formulations in which a worst-case expression is minimized or constrained [3, 4]. In recent years this idea has pursued more specifically under the heading of "distributionally robust optimization" in which special schemes for constructing collections

[10] Extending $\mathcal{J}$ from $\mathcal{Q}$ to all of $\mathcal{L}^2$ by assigning it the value ∞ outside of $\mathcal{Q}$ results, under this closedness condition, in $\mathcal{J}$ being lower semicontinuous.

of alternative probability measures are explored, cf. [10, 23, 8].[11] The term "stochastic ambiguity" is likewise often used in situations where multiple probability measures are under consideration. Both distributional robustness and stochastic ambiguity are therefore at the heart of risk theory. Interestingly, stochastic ambiguity can also enter on a higher level and then be coordinated with the stochastic ambiguity in risk, as will be seen in Section 4.

3 How should utility be understood?

Utility theory in economics is traditionally occupied with preferences among bundles of goods, but here we are focusing on preferences among scalar random variables such as arise in financial optimization or reliability engineering. We have been looking so far at loss-oriented random variables X with the idea that X will be preferable to X' when, with respect to a risk quantifier $\mathcal{R}$ that reflects our interests, $\mathcal{R}(X) < \mathcal{R}(X')$. In taking up utility, however, we switch to the opposite orientation, signaled here notationally (to reduce confusion) by Y as the symbol for a random variable for which we like outcomes to be higher rather than lower. The space of random variables is, as always, $\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{F}, P_0)$.

Without yet pinning down any specifics, we can build our discussion around the notion that a *utility quantifier* is a possibly extended-real-valued functional $\mathcal{U}$ on $\mathcal{L}^2$ which is aimed at aiding decisions by marking Y as preferred to Y' when $\mathcal{U}(Y) > \mathcal{U}(Y')$. Examples of this already in use will lead us to identifying specific axioms to place on such $\mathcal{U}$, although an obvious one from the start ought to be monotonicity: $Y \geq Y'$ should imply $\mathcal{U}(Y) \geq \mathcal{U}(Y')$.

The prime example of $\mathcal{U}$ to begin with is simple *expected utility*:

$$\mathcal{U}(Y) = E[u(Y)] = \int_\Omega u(Y(\omega))dP_0(\omega) \text{ for a function } u : (-\infty, \infty) \to [-\infty, \infty). \tag{3.1}$$

Here $-\infty$ has been allowed as a value of u to cover cases where the natural domain of u is not the whole real line, as for instance if $u(y) = \log y$ or $u(y) = \sqrt{y}$; then u is extended by $-\infty$. Clearly we should want the utility function u to be nondecreasing, since that will lead to the expected utility being monotone in assigning preferences. What else? In line with utility theory more generally, we should likely want the set of Y' preferred to Y to be convex. That essentially dictates in (3.1) that u should be a concave function, and then $\mathcal{U}$ will be concave.

But expected utility as in (3.1) is not the only approach to utility preferences. We can turn to *stochastic ambiguity* to get examples of the kind

$$\mathcal{U}(Y) = \inf_{P \in \mathcal{P}} E_P[u(Y)] = \inf_{P \in \mathcal{P}} E\Big[u(Y)\frac{dP}{dP_0}\Big] \tag{3.2}$$

for a set $\mathcal{P}$ of probability measures P that is contemplated as potentially having to be faced instead of just P_0. Clearly this could tie in with the risk ideas at the end of Section 2, and indeed we will come back to that in Section 4. The point for now is that (3.2) presents candidates beyond those in (3.1) which are definitely worthy of including in our picture of utility quantifiers.

Still another line of thinking leads to "relative utility" in the sense of comparing outcomes to some benchmark.[12] Thus, there may be some random variable B with respect to which we are interested in

$$\mathcal{U}(Y) = E[u(B+Y)]-E[u(B)] = \int_\Omega u_B(Y(\omega), \omega)dP_0(\omega) \text{ for } u_B(y, \omega) = u(B(\omega)+y)-u(B(\omega)). \tag{3.3}$$

[11] Researchers in that subject don't seem aware that it is a branch of the general theory of coherent measures of risk initiated much earlier in [2].

[12] In finance, for example one might want to compare the gains Y from a portfolio to a well known stock market index.

This is not of course covered by (3.1), because we do not just have a function $u_B(y)$ unless $B(\omega)$ is actually a constant b independent of ω.

In (3.3) we have $\mathcal{U}(0) = 0$, and that brings up something further. Back in the case of (3.1), there is a utility function u, but there is some arbitrariness in choosing it for getting preferences. The same standards for when Y is preferred to Y' would be captured by selecting any b with $u(b) > -\infty$ and replacing $u(y)$ by $u_b(y) = u(b+y) - u(b)$. Thus, no loss of generality is incurred by supposing in (3.1) that $u(0) = 0$, in which case $\mathcal{U}(0) = 0$ there as well. Moreover this would transfer to the setting of stochastic ambiguity in (3.2).

In this vein of "normalizing" a utility function u, we can also contemplate rescaling, since that too would not affect the preferences coming from it. This can have an important effect on relating utility to expectation, in connection with already having $u(0) = 0$. For instance in the case of u being differentiable at 0, if $u'(0) > 0$, as would be natural from the standpoint of monotonicity, we can divide by u by $u'(0)$ to normalize to having $u'(0) = 1$. With the concavity then implying $u(y) \leq y$ for all y, we thus can arrange that $\mathcal{U}(Y) \leq E[Y]$. Even without differentiability at 0, we can anyway divide u by some value between its right derivative $u'_+(0)$ and its left derivative $u'_-(0)$ (which exist in consequence of concavity) to get the same result.

These examples and considerations suggest working, in the random variable context here, at least, with the following set of conditions on a utility quantifier $\mathcal{U} : \mathcal{L}^2 \to [-\infty, \infty)$:

$$\begin{array}{ll} (U1) & \mathcal{U} \text{ is upper semicontinuous, possibly taking on } -\infty, \\ (U2) & \mathcal{U}((1-\lambda)Y + \lambda Y') \geq (1-\lambda)\mathcal{U}(Y) + \lambda\mathcal{U}(Y') \text{ for } \lambda \in (0,1), \\ (U3) & \mathcal{U}(Y) \geq \mathcal{U}(Y') \text{ when } Y \geq Y', \\ (U4) & \mathcal{U}(0) = 0, \end{array} \tag{3.4}$$

with the sometime addition of

$$(U4') \quad \mathcal{U}(Y) < E[Y] \text{ for all } Y \neq 0. \tag{3.5}$$

This perspective on utility differs, of course, from the common one in finance that mainly focuses on technical features of utility functions, like HARA, which have more to do with mathematical simplications than expressing true preferences, cf. [7]. It aims rather at a relative form of utility preferences which, for instance, builds on (3.3) for a benchmark return B through dividing also by $u'(B(\omega))$ to achieve (U4').

Our next step is dualizing the conditions (U1)–(U4) and (U4'). This could be done while keeping to the concavity in (U2), but it will really be better to revert to convexity, which entails reverting to the loss-oriented context of random variables X in Section 2. To this end we set up a one-to-one correspondence between functionals $\mathcal{U}$ and $\mathcal{V}$ through the relations

$$\mathcal{V}(X) = \mathcal{U}_*(X) := -\mathcal{U}(-X), \qquad \mathcal{U}(Y) = \mathcal{V}_*(Y) = -\mathcal{V}(-Y), \tag{3.6}$$

calling $\mathcal{V}$ the *regret quantifier* associated with $\mathcal{U}$ as utility quantifier. The conditions on $\mathcal{U}$ in (3.4) and (3.5) translate for its counterpart $\mathcal{V} : \mathcal{L}^2 \to (-\infty, \infty]$ into:

$$\begin{array}{ll} (V1) & \mathcal{V} \text{ is lower semicontinuous, possibly taking on } \infty, \\ (V2) & \mathcal{V}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{V}(X) + \lambda\mathcal{V}(X') \text{ for } \lambda \in (0,1), \\ (V3) & \mathcal{V}(X) \geq \mathcal{V}(X') \text{ when } X \geq X', \\ (V4) & \mathcal{V}(0) = 0, \end{array} \tag{3.7}$$

with the sometime addition of

$$(V4') \quad \mathcal{V}(X) > E[X] \text{ for all } X \neq 0. \tag{3.8}$$

Once more we investigate what happens with these conditions under the conjugacy relations

$$\mathcal{V}^*(Q) = \sup_X \{ \langle X, Q \rangle - \mathcal{V}(X)\}, \qquad \mathcal{V}(X) = \sup_Q \{ \langle X, Q \rangle - \mathcal{V}^*(Q)\}, \tag{3.9}$$

A distinction will be that instead of dual elements Q only in the simplex $\mathcal{P}_0$, we will have them in the nonnegative orthant

$$\mathcal{L}^2_+ = \{ Q \in \mathcal{L}^2 \,|\, Q \geq 0\}. \tag{3.10}$$

Theorem 2: regret dualization. *The class of functionals $\mathcal{K}$ on $\mathcal{L}^2$ that come up as conjugates $\mathcal{V}^*$ of regret quantifiers $\mathcal{V}$ satisfying (V1)–(V4) is characterized by*

$$\begin{array}{ll} (K1) & \mathcal{K} \text{ is lower semicontinuous,} \\ (K2) & \mathcal{K}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{K}(X) + \lambda\mathcal{K}(X') \text{ for } \lambda \in (0,1), \\ (K3) & \mathcal{L}^2_+ \supset \mathcal{M} = \operatorname{dom}\mathcal{K} = \{ Q \,|\, \mathcal{K}(Q) < \infty\}, \\ (K4) & \inf\limits_{Q \in \mathcal{M}} \mathcal{K}(Q) = 0. \end{array} \tag{3.11}$$

The extra condition (V4′) corresponds in this to

$$(K4') \quad \mathcal{K}(1) = 0 \text{ (entailing } 1 \in \mathcal{M}\text{), but } \partial\mathcal{K}(1) \text{ contains no nonzero } X. \tag{3.12}$$

The subclass consisting of the conjugates $\mathcal{V}^*$ of functionals $\mathcal{V}$ which are sublinear instead of just convex as (V2) is identified with $\mathcal{K}$ being $\equiv 0$ on $\mathcal{M}$. This subclass thus consists of the functionals $\mathcal{K}$ of the form

$$\mathcal{K} = \delta_{\mathcal{M}} \text{ for some nonempty closed convex set } \mathcal{M} \subset \mathcal{L}^2_+. \tag{3.13}$$

Proof. Conditions (K1) and (K2) merely echo (V1) and (V2) under the conjugacy between $\mathcal{V}$ and $\mathcal{K}$. To confirm that (V3) implies (K3), we note that if $Q \in \mathcal{M}$ then $\sup_X \{ E[XQ] - \mathcal{V}(X)\} < \infty$. Taking $c \in I\!R$ to be the supremum, we in particular have for any $Z \in \mathcal{L}^2_+$ that $\mathcal{V}(-Z) \geq -E[ZQ] - c$, but also from (V3) that $\mathcal{V}(-Z) \leq \mathcal{V}(0) = 0$. Therefore $E[ZQ] \geq -c$ for all $Z \in \mathcal{L}^2_+$, which requires $Q \in \mathcal{L}^2_+$. In the other direction, if $\mathcal{K}$ satisfies (K3), then from writing the second formula in (3.9) as

$$\mathcal{V}(X) = \sup_{Q \in \mathcal{M}} \{ E[XQ] - \mathcal{K}(Q)\} \tag{3.14}$$

we see that $\mathcal{V}(X+Z) \geq \mathcal{V}(X)$ when $Z \in \mathcal{L}^2_+$, which is (V3). From (3.14) it is obvious as well that (K4) corresponds to (V4). To understand (K4′), start by observing that $\mathcal{K}(1) = 0$ means under conjugacy that $\sup_X \{ E[X\,1] - \mathcal{V}(X)\} = 0$, which is the same as $\mathcal{V}(X) \geq E[X]$ for all X, with equality holding for $X = 0$ by (V3). The subgradient condition means that $\mathcal{V}(X) = E[X\,1] - \mathcal{K}(1)$ only when $X = 0$, which in view of $\mathcal{K}(1)$ being 0 ensures that $\mathcal{V}(X) > E[X]$ for all $X \neq 0$. □

A particular category to look at more closely in these relationships is expected utility as in (3.1) and its regret counterpart. What conditions on the utility function u on $(-\infty, \infty)$ make $\mathcal{U}(Y) = E[u(Y)]$ satisfy (U1)–(U4)? It is elementary that this is true when

$$\begin{array}{ll} (u1) & u \text{ is upper semicontinuous, possibly taking on } -\infty, \\ (u2) & u((1-\lambda)y + \lambda y') \geq (1-\lambda)u(y) + \lambda u(y') \text{ for } \lambda \in (0,1), \\ (u3) & u(y) \geq u(y') \text{ when } y \geq y', \\ (u4) & u(0) = 0, \end{array} \tag{3.15}$$

and that (U4′) will be satisfied as well when

$$(u4') \quad u(y) < y \text{ for all } y \neq 0. \tag{3.16}$$

Moreover these conditions are not just sufficient but necessary, as seen by considering constants $Y \equiv y$.

On the side of regret in place of utility, it all works the same way. Expected regret takes the form

$$\mathcal{V}(X) = E[v(X)] = \int_\Omega v(X(\omega))dP_0(\omega) \text{ for a function } v : (-\infty, \infty) \to (-\infty, \infty]. \tag{3.17}$$

The conditions on v that are both necessary and sufficient for $\mathcal{V}$ to satisfy (V1)–(V4) are

$$\begin{array}{ll} (v1) & v \text{ is lower semicontinuous, possibly taking on } \infty, \\ (v2) & v((1-\lambda)x + \lambda x') \le (1-\lambda)v(x) + \lambda v(x') \text{ for } \lambda \in (0,1), \\ (v3) & v(x) \le v(x') \text{ when } x \ge x', \\ (v4) & v(0) = 0, \end{array} \tag{3.18}$$

and (V4′) is associated with

$$(v4') \quad v(x) > x \text{ for all } x \neq 0. \tag{3.19}$$

On the other hand, the sublinearity of $\mathcal{V}$ corresponds to the sublinearity of v, which means in the one-dimensional setting and in coordination with (v1)–(v4) that

$$\begin{array}{l} \text{on } (-\infty, 0),\ v \text{ is linear with a slope } a \ge 0, \text{ while} \\ \text{on } (0, \infty),\ v \text{ is linear with a slope } b \ge a \text{ or } v \equiv \infty. \end{array} \tag{3.20}$$

The pairing of $\mathcal{U}$ and $\mathcal{V}$ in (3.6) corresponds to the pairing of u and v by

$$v(x) = u_*(x) = -u(-x), \qquad u(y) = v_*(y) = -v(-y), \tag{3.21}$$

which is easy to picture because it just means that the graphs of u and v are reflections of each other across a 45-degree line between the axes of $\mathbb{R}^2$.

Duality between $\mathcal{V}$ and $\mathcal{K} = \mathcal{V}^*$ as in Theorem 2 is easy in this context via [13]:

$$\mathcal{V}(X) = E[v(X)] \iff \mathcal{K}(Q) = E[k(Q)], \text{ where } k(q) = v^*(q) = \sup_x \{ xq - v(x) \}. \tag{3.22}$$

The conditions on k that correspond to (K1)–(K4) on $\mathcal{K}$ are

$$\begin{array}{ll} (k1) & k \text{ is lower semicontinuous,} \\ (k2) & k((1-\lambda)X + \lambda X') \le (1-\lambda)k(X) + \lambda k(X') \text{ for } \lambda \in (0,1), \\ (k3) & k(q) < \infty \implies q \ge 0, \\ (k4) & \inf_q k(q) = 0. \end{array} \tag{3.23}$$

The extra condition (v4′) corresponds in this to

$$(k4') \quad k \text{ is differentiable at 1 with } k'(1) = 0, \tag{3.24}$$

which in combination with (k4) means that $k(1) = 0$ and $\partial k(1) = \{0\}$. The case of sublinear $\mathcal{V}$, in which v has the form in (3.21), corresponds to k being the indicator of the interval $[a, b] \subset [0, \infty)$, or as the case may be, $[a, \infty)$.

4 Risk versus utility

One of our chief goals in this article is to clarify how risk and utility might be related to each other in our context, and this passage from utility quantifiers $\mathcal{U}$ via (3.6) to regret quantifiers $\mathcal{V}$ and their dualization sheds a lot of light on that. The regret conditions in (3.7) and (3.8) have turned out to be almost identical to the risk conditions in (2.9) and (2.10)! The only difference is seen in having constant/nonconstant in (R4) and (R4′) versus zero/nonzero in (V4) and (V4′). In the dualizations that is paralleled by having $Q \in \mathcal{P}_0$ in the risk case but only $Q \in \mathcal{L}^2_+$ in the regret case. Those seemingly small distinctions are nevertheless significant and have interesting consequences which we will explore next.

Theorem 3: risk from utility or regret. *Let $\mathcal{V}$ be a regret quantifier satisfying (V1)–(V4) + (V4′), thus being associated under (3.6) with a utility quantifier $\mathcal{U}$ satisfying (U1)–(U4) + (U4′). Let*

$$\mathcal{R}(X) = \inf_{C\in\boldsymbol{R}} \{ C + \mathcal{V}(X - C)\}, \qquad \mathcal{S}(X) = \operatorname*{argmin}_{C\in\boldsymbol{R}} \{ C + \mathcal{V}(X - C)\}. \tag{4.1}$$

Then $\mathcal{S} \neq \emptyset$ and $\mathcal{R}$ is a risk quantifier satisfying (R1)–(R4) + (R4′). In this relationship sublinearity for $\mathcal{V}$ produces sublinearity for $\mathcal{R}$. The dualizations $\mathcal{J} = \mathcal{R}^*$ and $\mathcal{K} = \mathcal{V}^*$ are moreover tied together by

$$\mathcal{J}(Q) = \begin{cases} \mathcal{K}(Q) & \text{if } E[Q] = 1, \\ \infty & \text{if } E[Q] \neq 1. \end{cases} \tag{4.2}$$

Proof. This was established in [20] under an additional assumption on $\mathcal{V}$, but that assumption was shown to be unnecessary in [16]. □

The rule in (4.1) provides a vast generalization of the formula (2.6) for the superquantile $\overline{q}_\alpha(X)$, in which

$$\overline{q}_\alpha(X) = \min_{C\in\boldsymbol{R}} \{ C + \mathcal{V}_\alpha(X - C)\} \text{ for } \mathcal{V}_\alpha(X) = \frac{1}{1-\alpha} E[\max\{0, X\}]. \tag{4.3}$$

The random variable $\max\{0, X\}$ gives the *absolute* loss in X, unbalanced by the desirable outcomes, if any, in which $X(\omega) < 0$. As mentioned earlier, the argmin set in (4.3), if not consisting of the quantile $q_\alpha(X)$ alone, is a closed interval with that quantile as its left endpoint. That exemplifies the nonemptiness of the set $\mathcal{S}$ in (4.1). Note that (4.3) falls into the case of sublinearity, which on the dual side has $\mathcal{J}$ and $\mathcal{K}$ indicators of the sets

$$\mathcal{Q}_\alpha = \{ Q \in \mathcal{P}_0 \,|\, Q \leq (1-\alpha)^{-1}\}, \qquad \mathcal{M}_\alpha = \{ Q \in \mathcal{L}^2_+ \,|\, Q \leq (1-\alpha)^{-1}\}, \tag{4.4}$$

which illustrate the rule in (4.2). An example beyond the sublinear case is furnished by the log-exponential risk quantifier:

$$\log E[\exp X] = \min_{C\in\boldsymbol{R}} \{ C + \mathcal{V}(X - C)\} \text{ for } \mathcal{V}(X) = E[\exp X - 1], \tag{4.5}$$

where the regret quantifier $\mathcal{V}$ dualizes to

$$\mathcal{K}(Q) = \begin{cases} E[Q \log Q - Q] & \text{when } Q \geq 0, \\ \infty & \text{otherwise.} \end{cases} \tag{4.6}$$

A remarkable feature of this choice of $\mathcal{V}$ is that it not only produces $\log E[\exp X]$ as the value of $\mathcal{R}(X)$ but also $C = \log E[\exp X]$ as the unique minimizing C in $\mathcal{S}(X)$. This is reminiscent of facts about the exponential function, like it being its own derivative.

The general formula in (4.1) can be given an appealing interpretation in terms of the risk of incurring a loss. Through (3.6) the regret $\mathcal{V}(X)$ is a kind of anti-utility which stands for the overall displeasure in being saddled with the potential losses in X (with negative losses acting as gains). These losses occur in the future, but it is possible to account for them to some extent in the present by writing off a selected amount C of loss as being certain. Then in place of $\mathcal{V}(X)$ we have, after the write-off, only the regret in the reduced random loss variable $X - C$. We can optimize by determining a value of C that makes the combination $C + \mathcal{V}(X - C)$ as low as possible, i.e., $C \in \mathcal{S}$. There does exist such C because $\mathcal{S} \neq \emptyset$. The minimizing C can be thought of as the amount of compensation that ought to be demanded for shouldering the obligations represented by X.

Everything in terms of a regret quantifier $\mathcal{V}$ can be restated in terms of a utility quantifier through (3.6). For instance, (4.1) can given the form

$$\mathcal{R}_*(Y) = \sup_{D \in \boldsymbol{R}} \{ D + U(Y - D)\}, \;\text{ where }\; \mathcal{R}_*(Y) = -\mathcal{R}(-Y) \tag{4.7}$$

in which the C in (4.1) is replaced by $D = -C$ when X is replaced by $Y = -X$. The convexity of $\mathcal{R}$ turns into the concavity of $\mathcal{R}_*$ and leads in dualization to recasting the formulas (2.18), (2.19) and (2.20) in terms of minimization. Mathematically this is a trivial exercise, but the different view in (4.7) is actually the original one in this subject. It is how the thinking went with Ben-Tal and Teboulle [5], who called a D giving the maximum in (4.7) the *optimized certainty equivalent* for Y. Their pioneering efforts only targeted expected utility as in (3.1), not necessarily "normalized," but that was enough to cover the crucial connection with some of the coherent risk quantifiers of major importance. The widening of the picture to other versions of $\mathcal{U}$, as facilitated in treatment by the introduction of "regret" as a reoriented partner to utility, was carried out in [20].

Without laying out all the parallel details in $\mathcal{U}$-$\mathcal{V}$ notation and formulation, we can take advantage of this prospect to understand more about stochastic ambiguity as seen by economists in utility terms. This brings us back to examing more closely utility quantifiers like the one in (3.2), which concern worst-case expected utility with respect to some collection of probability measures. In order not to disrupt the notational scheme for risk and that we have so far put in place, which already is based through duality on a form of stochastic ambiguity, we shift the formula in (3.2) to avoid using $P \in \mathcal{P}$ and instead incorporate uncertainty in terms of a collection $\bar{\mathcal{P}}$ of probability measures $\bar{P}$ and the associated set $\bar{\mathcal{Q}}$ of densities $\bar{Q} = d\bar{P}/dP_0$:

$$\mathcal{U}(Y) = \inf_{\bar{P} \in \bar{\mathcal{P}}} E_{\bar{P}}[u(Y)] = \inf_{\bar{Q} \in \bar{\mathcal{Q}}} E[u(Y)\bar{Q}]. \tag{4.8}$$

With no loss of generality (passing to the closed convex hull if necessary), we can suppose in this that $\bar{\mathcal{Q}}$ is a nonempty closed convex subset of the probability simplex $\mathcal{P}_0$. We then have

$$\mathcal{U}(Y) = -\bar{\mathcal{R}}(-u(Y)) = \bar{\mathcal{R}}_*(u(Y)) \;\text{ for the risk quantifier }\; \bar{\mathcal{R}}(X) = \sup_{\bar{Q} \in \bar{\mathcal{Q}}} E[X\bar{Q}]. \tag{4.9}$$

For complete rigor here we need to be sure that the expectations in (4.8) are well defined, of course, and that is true because u is concave.[13] Fancier versions of stochastic ambiguity in utility have been

[13] The concavity of u implies the existence of an affine function a that majorizes u, and then $u(Y)\bar{Q} \leq a(Y)\bar{Q}$. For $Y \in \mathcal{L}^2$ we have $a(Y) \in \mathcal{L}^2$. Since $\bar{Q} \in \mathcal{L}^2$ as well, we know that $a(Y)\bar{Q}$ is integrable, giving $\langle a(Y), \bar{Q}\rangle$. Having $u(Y)\bar{Q}$ bounded above by the integrable function $a(Y)\bar{Q}$ ensures that the expectation in (4.8) is well defined as either a real value or $-\infty$. The latter case occurs if and only if the set $\left\{ \omega \,\middle|\, u(Y(\omega))\bar{Q}(\omega) = -\infty \right\}$ has positive measure with respect to P_0 (under the usual interpretation that the product of $-\infty$ and 0 is 0).

studied in which (4.8) is replaced by

$$\mathcal{U}(Y) = \inf_{\bar{P}\in\bar{\mathcal{P}}} \{ E_{\bar{P}}[u(Y)] + c(\bar{P})\} \tag{4.10}$$

for some function c, or in our density-type formulation,

$$\mathcal{U}(Y) = \inf_{\bar{Q}\in\bar{\mathcal{Q}}} \{ E[u(Y)\bar{Q}] + \bar{\mathcal{J}}(\bar{Q})\} \tag{4.11}$$

for some function $\bar{\mathcal{J}}$. This type of utility quantifier was introduced by Maccheroni, Marinacci and Rustichini [11] for doing a better job at capturing the preferences of financial decision makers; in earlier work of Gilboa and Schmeidler [9] the extra term was not present. The main thing for us here is that a formula of type (4.11) fits perfectly with the ideas of risk and its dualization in Theorem 1, yielding the representations (2.18)–(2.19): we can rewrite (4.11) in those terms as

$$\mathcal{U}(Y) = \bar{\mathcal{R}}_*(u(Y)) = -\bar{\mathcal{R}}(-u(Y)) \text{ for } \bar{\mathcal{R}}(\bar{X}) = \sup_{\bar{Q}\in\bar{\mathcal{Q}}} \{ E[\bar{X}\bar{Q}] - \bar{\mathcal{J}}(\bar{Q})\} \tag{4.12}$$

under the assumption that $\bar{\mathcal{J}}$ satisfies (J1)–(J4) with $\operatorname{dom}\bar{\mathcal{J}} = \bar{\mathcal{Q}}$.

Strzalecki [24] has taken special interest in the case of (4.10) which, expressed as (4.11), is

$$\mathcal{U}(Y) = \inf_{\bar{Q}\in\mathcal{P}_0} \{ E[u(Y)\bar{Q}] + E[\bar{Q}\log\bar{Q}]\}. \tag{4.13}$$

The extra term then expresses, for the probability measure $\bar{P}$ having $d\bar{P}/dP_0 = \bar{Q}$, the Kullback-Leibler distance of $\bar{P}$ from P_0, or the relative entropy of $\bar{P}$ with respect to P_0. According to (2.22) this means that

$$\mathcal{U}(Y) = \bar{\mathcal{R}}_*(u(Y)) = -\bar{\mathcal{R}}(-u(Y)) \text{ for the risk quantifier } \bar{\mathcal{R}}(X) = \log E[\exp X]. \tag{4.14}$$

Two general questions emerge. Do utility quantifiers $\mathcal{U}$ defined in the manner of (4.12), with (4.9) and (4.14) as special cases, satisfy the conditions (U1)–(U4), and possibly (U4′), that we came up with earlier? And when they do, what can be said about the risk quantifier $\mathcal{R}$ that they produce through the formula in Theorem 3 and its dualization to $\mathcal{J}$?

In answering these questions we will proceed for convenience in the equivalent mode of regret instead of utility. This will take advantage of the pairing of the utility funtion u with a regret function v in (3.20) and the corresponding pairing of a utility quantifier $\mathcal{U}$ as in (4.11) with a regret quantifier

$$\mathcal{V}(X) = \bar{\mathcal{R}}(v(X)) \text{ for } \bar{\mathcal{R}}(\bar{X}) = \sup_{\bar{Q}\in\bar{\mathcal{Q}}} \{ E[\bar{X}\bar{Q}] - \bar{\mathcal{J}}(\bar{Q})\} \tag{4.15}$$

Theorem 4: basic properties of ambiguous expected utility/regret. *A regret quantifier $\mathcal{V}$ of the form in (4.15) satisfies (V1)–(V4) as long as the risk quantifier $\bar{\mathcal{R}}$ satisfies (R1)–(R4) (or its dual counterpart $\mathcal{J}$ satisfies (J1)–(J4)) and the underlying regret function v satisfies (v1)–(v4). It further satisfies (V4′) when $\bar{\mathcal{R}}$ satisfies (R4′) (or $\mathcal{J}$ satisfies (J4′)) and v satisfies (v4′), and then there is associated with $\mathcal{V}$ a risk quantifier $\mathcal{R}$ given by (4.1) as*

$$\mathcal{R}(X) = \min_C \{ C + \overline{\mathcal{R}}(v(X - C))\}, \tag{4.16}$$

which in turn satisfies (R1)–(R4)+(R4′).

In addition, $\mathcal{V}$ is sublinear when $\bar{\mathcal{R}}$ and v are sublinear, and then $\mathcal{R}$ is sublinear as well.

Proof. These claims are elementary to verify except for the one about (V4′). If v satisfies (v4′), we have $v(X) \geq X$ for any $X \in \mathcal{L}^2$. Then by (R3) we have $\bar{\mathcal{R}}(v(X)) \geq \bar{\mathcal{R}}(X)$. On the other hand, by (R4′) we have $\bar{\mathcal{R}}(X) \geq E[X]$, with equality holding only when X is constant. Thus, $\bar{\mathcal{R}}(v(X)) = E[X]$ cannot hold unless X is a constant C, in which case we are looking at $\bar{\mathcal{R}}(v(C)) = C$. Since $\bar{\mathcal{R}}(v(C)) = v(C)$ by (R4), but $v(C) > C$ for $C \neq 0$ by (v4′), this is only possible when $C = 0$. Thus, $\bar{\mathcal{R}}(v(X))$, which is $\mathcal{V}(X)$, cannot equal $E[X]$ unless 0, and we have (V4′). The properties about $\mathcal{R}$ come then from applying Theorem 3 in this setting. ◻

Working our way now toward dualization of the quantifiers $\mathcal{V}$ and $\mathcal{R}$ in Theorem 4, we must begin with close scrutiny of the special case of (4.15) where only a simple change of P_0 to a different probability measure $\bar{P}$ is involved, so that $\mathcal{V}$ just has the form:

$$\mathcal{V}_{\bar{Q}}(X) = E[v(X)\bar{Q}] = E_{\bar{P}}[v(X)] \text{ for fixed } \bar{P} \text{ and } \bar{Q} = d\bar{P}/dP_0. \tag{4.17}$$

Under the assumption that v satisfies (v1)–(v4), this is covered by Theorem 4 and therefore enjoys the dualization in Theorem 2. This would reduce to (3.21) for the function $k = v^*$ satisfying (3.22) if $\bar{P} = P_0$, corresponding to $\bar{Q} \equiv 1$, but here we want to allow $\bar{P}$ to be different from P_0.

Lemma. *For v satisfying (v1)—(v4) and its conjugate k satisfying (k1)–(k4), the convex functional $\mathcal{V}^*_{\bar{Q}}$ conjugate to $\mathcal{V}_{\bar{Q}}$ in (4.17) is*

$$\mathcal{V}^*_{\bar{Q}}(Q) = E[\overline{k}(\bar{Q}, Q)], \quad \text{where } \overline{k}(\bar{q}, q) = \begin{cases} \bar{q}k(\bar{q}^{-1}q) & \text{if } \bar{q} > 0, \\ 0 & \text{if } \bar{q} = 0,\ q = 0, \\ \infty & \text{otherwise,} \end{cases} \tag{4.18}$$

and therefore

$$\sup_Q \{ E[XQ] - E[\overline{k}(\bar{Q}, Q)]\} = \mathcal{V}_{\bar{Q}}(X). \tag{4.19}$$

Here $\overline{k}$ is a sublinear function on $I\!R \times I\!R$ for which the lower semicontinuous hull is

$$\mathrm{cl}\,\overline{k}(\bar{q}, q) = \begin{cases} \bar{q}k(\bar{q}^{-1}q) & \text{if } \bar{q} > 0, \\ k^\infty(q) & \text{if } \bar{q} = 0, \\ \infty & \text{if } \bar{q} < 0, \end{cases} \tag{4.20}$$

where

$$k^\infty(q) = \lim_{t\to\infty} \frac{k(tq)}{t} = \sup\{ xq \mid v(x) < \infty\}. \tag{4.21}$$

Moreover

$$\sup_Q \{ E[XQ] - E[\mathrm{cl}\,\overline{k}(\bar{Q}, Q)]\} = \mathcal{V}_{\bar{Q}}(X) + \delta(X \mid X \leq b), \tag{4.22}$$

where $\delta(X \mid X \leq b)$ is the indicator of X being $\leq b = \sup\{ x \mid v(x) < \infty\}$ with probability 1.

Proof. We are dealing in (4.17) with an integral functional on $\mathcal{L}^2$ having the form

$$X \mapsto \int_\Omega f(X(\omega), \omega) dP_0(\omega) \text{ for } f(x, \omega) = v(x)\bar{Q}(\omega)$$

and can utilize the rule in [13], according to which the conjugate convex functional on $\mathcal{L}^2$ is

$$Q \mapsto \int_\Omega f^*(Q(\omega), \omega) dP_0(\omega), \text{ where } f^*(\cdot, \omega) \text{ is conjugate to } f(\cdot, \omega).$$

The calculus of conjugates in convex analysis gives us

$$f^*(q,\omega)=\sup_x \{ xq-f(x,\omega)\} = \sup_x \{ xq-\bar{Q}(\omega)v(x)\} = [\bar{Q}(\omega)v]^*(q).$$

A further rule about multiplication in [12] says that

$$[\bar{q}v]^*(q)=\begin{cases} \bar{q}v^*(\bar{q}^{-1}q) \text{ if } \bar{q}>0,\\ 0 \text{ if } \bar{q}=0,\, q=0,\\ \infty \text{ otherwise.}\end{cases}$$

Since $v^*=k$, this confirms the claim in (4.18).

The formula then for $\mathrm{cl}\,\overline{k}$ is well known from convex analysis [12, Corollary 8.5.1]. The conjugate of $\mathrm{cl}\,\overline{k}(0,\cdot)$ is then the conjugate of k^∞, which is the indicator of the closure of $\mathrm{dom}\,v$, namely $\delta_{(-\infty,b]}$. (Recall that because v is nondecreasing and $v(0)=0$, the interval comprising $\mathrm{dom}\,v$ is not bounded from below.) In repeating the earlier integral functional argument to get the conjugate of $Q\mapsto \int_\Omega g(Q(\omega),\omega)dP_0(\omega)$ for $g(q,\omega)=\mathrm{cl}\,\overline{k}(\bar{Q}(\omega),q)$ we get the integral functional $X\to\int_\Omega g^*(X(\omega),\omega)dP_0(\omega)$ with

$$g^*(x,\omega)=\begin{cases} v(x)\bar{Q}(\omega) \text{ if } \bar{Q}(\omega)>0,\\ \delta_{(\infty,b]}(x) \text{ if } \bar{Q}(\omega)=0.\end{cases}$$

However, this is the same as $g^*(x,\omega)=v(x)\bar{Q}(\omega)+\delta_{(-\infty,b]}(x)$. That yields (4.22). $\square$

With this at our disposal we are ready for the broader dualizations.

Theorem 5: dualizations from ambiguous regret/utility. *Under the assumption that $\bar{R}$ satisfies (R1)–(R4) + (R4′) while v satisfies (v1)–(v4) + (v4′), the regret quantifier $\mathcal{V}$ in (4.15) has the dual representation*

$$\mathcal{V}(X)=\sup_{Q\in\mathcal{L}^2_+}\Big\{ E[XQ]-\mathcal{K}_0(Q)\Big\} \;\text{ for }\; \mathcal{K}_0(Q)=\inf_{\bar{Q}\in\bar{\mathcal{Q}}}\Big\{ E[\mathrm{cl}\,\overline{k}(\overline{Q},Q)]+\bar{\mathcal{J}}(\bar{Q})\Big\} \tag{4.23}$$

with $\overline{k}$ coming from(4.18), where $\mathcal{K}_0$ is a convex functional on $\mathcal{L}^2$, and consequently the conjugate functional $\mathcal{K}=\mathcal{V}^*$ is the lower semicontinuous hull $\mathrm{cl}\,\mathcal{K}_0$ of $\mathcal{K}_0$. The risk quantifier $\mathcal{R}$ associated with $\mathcal{V}$ by Theorem 3 then has the dual representation

$$\mathcal{R}(X)=\sup_{Q\in\mathcal{P}_0}\Big\{ E[XQ]-\mathcal{K}(Q)\Big\},$$

which corresponds to the conjugate functional $\mathcal{J}=\mathcal{R}^*$ being $\mathcal{K}$ plus the indicator $\delta(Q\,|\,E[Q]=1)$.

Proof. From (4.15) and (4.17) we have

$$\mathcal{V}(X)=\sup_{\bar{Q}\in\bar{\mathcal{Q}}}\{\mathcal{V}_{\bar{Q}}(Q)-\bar{\mathcal{J}}(\bar{\mathcal{Q}})\} \tag{4.24}$$

where, by the Lemma, $\mathcal{V}_{\bar{Q}}(X)$ is given by (4.19). In fact, in this situation it makes no difference if we replace $\mathcal{V}_{\bar{Q}}(Q)$ here by $\mathcal{V}_{\bar{Q}}(X)+\delta(X\,|\,X\le b)$. The reason is that 1 is one of the elements in $\bar{\mathcal{Q}}$ in consequence of (R4′) for $\bar{\mathcal{R}}$ (through its counterpart (J4′) for $\bar{\mathcal{J}}$), and for that choice of $\bar{\mathcal{Q}}$ one has $\mathcal{V}_{\bar{Q}}(Q)=E[v(X)]=\infty$ unless $v(X)\in\mathrm{dom}\,v$ almost surely, implying $v(X)\le b$ almost surely (for b as introduced in the Lemma). Thus, nothing is changed in (4.24) if we express $\mathcal{V}_{\bar{Q}}(Q)$ by the supremum in (4.22) instead of the one in (4.17). Therefore, in writing (4.15) as

$$\mathcal{V}(Q) = \sup_{\bar{Q}\in\bar{\mathcal{Q}}} \{ \mathcal{V}_{\bar{Q}}(Q) - \bar{\mathcal{J}}(\bar{Q})\}$$

we can replace $\mathcal{V}_{\bar{Q}}(Q)$ by the supremum on the left side of (4.22) and obtain

$$\begin{aligned}\mathcal{V}(Q) &= \sup_{\bar{Q}\in\bar{\mathcal{Q}}} \Big\{ \sup_Q \{ E[XQ] - E[\mathrm{cl}\,\overline{k}(\bar{Q},Q)]\} - \bar{\mathcal{J}}(\bar{Q})\Big\} \\ &= \sup_{\bar{Q}\in\bar{\mathcal{Q}}} \sup_Q \Big\{ E[XQ] - E[\mathrm{cl}\,\overline{k}(\bar{Q},Q)]\} - \bar{\mathcal{J}}(\bar{Q})\Big\} \\ &= \sup_Q \Big\{ E[XQ] - \inf_{\bar{Q}\in\bar{\mathcal{Q}}} \{ E[\mathrm{cl}\,\overline{k}(\bar{Q},Q)] + \bar{\mathcal{J}}(\bar{Q})\}\Big\},\end{aligned}$$

which is (4.23) since $\overline{k}(\bar{q},q) = \infty$ when $q > 0$. This says that $\mathcal{K}_0^* = \mathcal{V}$, and therefore $\mathcal{V}^* = \mathcal{K}_0^{**} = \mathrm{cl}\,K_0$. The assertion about $\mathcal{R}$ follows then from (4.2) in Theorem 3. □

Perhaps further analysis, invoking additional assumptions, could do away with the closure operation, letting $\mathcal{K} = \mathcal{K}_0$, but we leave that for future work.

An interesting example of the relationships in Theorem 5 can be obtained by taking v to be the function underlying the α-superquantile,

$$v(x) = (1-\alpha)^{-1}\max\{0,x\}, \qquad k(q) = \delta_{[0,(1-\alpha)^{-1}]}(q), \ \text{ where } \ \alpha\in(0,1). \tag{4.25}$$

Then $k^\infty = \delta_0$, so that there is no difference between $\mathrm{cl}\,\overline{k}$ and $\overline{k}$, with

$$k(\bar{q},q) = \begin{cases} 0 \ \text{ if } \ \bar{q}>0,\ \bar{q}\geq(1-\alpha)q, \\ 0 \ \text{ if } \ \bar{q}=0,\ q=0, \\ \infty \ \text{ otherwise.}\end{cases} \tag{4.26}$$

This yields in (4.23) that

$$\mathcal{K}_0(Q) = \inf\{ \bar{\mathcal{J}}(\bar{Q}) \,|\, \bar{Q}\geq(1-\alpha)Q\}, \tag{4.27}$$

where the condition $\bar{Q}\in\bar{\mathcal{Q}}$ has been omitted as superfluous since it just corresponds to $\bar{\mathcal{J}}(\bar{Q}) < \infty$.

References

[1] Acerbi, C., "Spectral measures of risk: a coherent reprsentation of subjective risk aversion," *Journal of Banking and Finance* **26** (2002), 1505–1518.

[2] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D., "Coherent measures of risk," *Mathematical Finance* **9** (1999), 203–227.

[3] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A., *Robust Optimization*, Princeton University Press, 2009.

[4] Ben-Tal, A., and Nemirovski, A., "Robust convex optimization," *Mathematics of Operations Research* **23** (1998), 769–805.

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

[6] Dencheva, D., and Ruszczỳnski, A., "Common mathematical foundations of expected utility and dual utility theories," *SIAM J. Optimization* **24** (2013), 381–405.

[7] Föllmer, H., and Schied, A., *Stochastic Finance*, 2nd edition, de Gruyter, New York, 2004.

[8] Gao, R., and Kleywegt, A. J., Sistrigutiionally robust stochastic optimization with Wasserstein distance," arXiv:1604.02199 (2016).

[9] Gilboa, I.,and Schmeidler, D., "Maxmin expected utility with non-unique prior," *J. of Mathematical Economics* **18** (1989), 141–153.

[10] Goh, J., and Sim, M., "Distributionally robust optimization and its tractable approximations," *Operations Research* **58** (2010), 902–917.

[11] Maccheroni, F., Marinacci, M., and Rustichini, A., "Ambiguity aversion, robustness, and the variational representation of preferences," *Econometrica* **74** (2006), 1147–1498

[12] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, 1970.

[13] Rockafellar, R.T., "Integrals which are convex functionals," *Pacific J. Mathematics* **24** (1968), 525–539.

[14] Rockafellar, R. T., and Royset, J. O., "On buffered failure probability in design and optimization of structures," *J. Reliability Engineering and System Safety* **99** (2010), 499–510.

[15] Rockafellar, R.T., and Royset, J.O., "Random variables, monotone relations and convex analysis," *Mathematical Programming B* [2014].

[16] Rockafellar, R.T., and Royset, J.O., "Measures of residual risk with connections to regression, risk tracking, surrogate models and ambiguity," *SIAM J. Optimization* **28** (2015), 1179–1208.

[17] Rockafellar, R.T., and Royset, J.O., "Superquantile/CVaR risk measures: second-order theory," *Annals of Operations Research* **262** (2018), 3–29.

[18] Rockafellar, R. T. and Uryasev, S., "Optimization of conditional value-at-risk," *J. Risk* **2** (2000), 21–42.

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

[20] Rockafellar, R. T. and Uryasev, S. "The fundamental risk quadrangle in risk management, optimization and statistical estimation," *Surveys in Operations Research and Management Science* **18** (2013), 33–53.

[21] Rockafellar, R. T., Uryasev, S., and Zabarankin, M., "Generalized deviations in risk analysis," *Finance and Stochastics* **10** (2006), 51–74.

[22] Rockafellar, R.T., and Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1997.

[23] Wiesemann, W., Kuhn, D., and Sim, M., "Distributionally robust convex optimization," *Operations Research* **62** (2014), 1358–1376.

[24] Strzalecki, T.,' "Axiomatic foundations of multiplier preferences," *Econometrica* **79** (2011), 47–73.