11 October 2011

THE FUNDAMENTAL RISK QUADRANGLE IN RISK MANAGEMENT, OPTIMIZATION AND STATISTICAL ESTIMATION

R. Tyrrell Rockafellar, [1] ***Stan Uryasev*** [2]

RESEARCH REPORT # 2011-5

Risk Management and Financial Engineering Lab
Center for Applied Optimization
Department of Industrial and Systems Engineering
University of Florida, Gainesville, FL 32611

Abstract

Measures of risk can be used to extract from a random variable that stands for a hazard cost, or loss, a single quantity which can substitute for it in risk management and optimization. Risk in this sense can be derived from penalty expressions of regret about the mix of potential outcomes of such a random variable by means of trade-offs between an up-front level of cost and the uncertain residual.

Statistical estimation is inevitably a partner with risk management in handling hazard variables, which may be known only through a data base, but a much deeper connection has come to light with statistical theory itself, in particular regression. Very general measures of error can associate with any hazard variable a "statistic" along with a "deviation" which quantifies the variable's nonconstancy. Measures of deviation, on the other hand, are known to be paired closely with measures of risk exhibiting aversity. A direct correspondence can furthermore be identified between measures of error and measures of regret, or its flip side as utility. The fundamental quadrangle of risk developed here puts all of this together in a unified scheme.

Keywords: *risk quadrangle, risk measures, deviation measures, error measures, regret, utility, entropy, coherency, value-at-risk, conditional value-at-risk, quantiles, superquantiles, generalized regression.*

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*.
[2]University of Florida, ISE Department, P.O. Box 116595, 303 Weil Hall, Gainesville, FL 32611-6595;
E-mail: *uryasev@ufl.edu*

1 Introduction

The challenges of dealing with risk pervade many areas of management and engineering. The decisions that have to be made in risky situations must nonetheless confront constraints on their consequences, no matter how uncertain those consequences may be. Furthermore, the decisions need to be open to comparisons which enable some kind of optimization to take place.

When uncertainty is modeled probabilistically with random variables, practical challenges arise about estimating properties of those random variables and their interrelationships. Information may come from empirical distributions generated by sampling, or there may only be databases representing information accumulated somehow or other in the past. Standard approaches to statistical analysis and regression in terms of expectation, variance and covariance may then be brought in. But the prospect is now emerging of a vastly expanded array of tools which can be finely tuned to reflect the various ways that risk may be assessed and, at least to some extent, controlled.

This paper is aimed at promoting and developing such tools in a new paradigm we call the *risk quadrangle*, which is shown in Diagram 1. It brings together several different lines of research and methodology which have, until now, been pursued separately and without a full view of their interplay.

$$\begin{array}{ccccc} & & \text{risk } \mathcal{R} \longleftrightarrow \mathcal{D} \text{ deviation} & & \\ \text{optimization} & & \uparrow\downarrow \quad \mathcal{S} \quad \downarrow\uparrow & & \text{estimation} \\ & & \text{regret } \mathcal{V} \longleftrightarrow \mathcal{E} \text{ error} & & \end{array}$$

Diagram 1: The Fundamental Risk Quadrangle

The context is that of random variables that can be thought of as standing for uncertain "costs" or "losses" in the broadest sense, not necessarily monetary (with a negative "cost" corresponding perhaps to a "reward"). The language of cost gives the orientation that we would like the outcomes of these random variables to be lower rather than higher, or to be held below some threshold. All sorts of indicators that may provide signals about hazards can be viewed from this perspective. The quadrangle elements provide "quantifications" of them which can be utilized for various purposes.

It will help, in understanding the quadrangle, to begin at the upper left corner, where $\mathcal{R}$ is a so-called *measure of risk*. The specific sense of this needs clarification, since there are so many angles to "risk." In denoting a random "cost" by X and a constant by C, the key question is how to give meaning to a statement that X is "more or less" $\leq C$. The role of a risk measure $\mathcal{R}$ is to aggregate the overall uncertain cost in X into a single numerical value $\mathcal{R}(X)$ in order to

$$\text{model "}X \text{ more or less} \leq C\text{" by the inequality } \mathcal{R}(X) \leq C.$$

There are familiar ways of doing this. One version could be that X is $\leq C$ on average, as symbolized by $\mu(X) \leq C$ with $\mu(X)$ the mean value, or in the equivalent alternative notation of expectations, $EX \leq C$. Then $\mathcal{R}(X) = \mu(X) = EX$. A tighter version could be $\mu(X) + \lambda\sigma(X) \leq C$ with λ giving a positive multiple of the standard deviation $\sigma(X)$ so as to provide on a safety margin reminiscent of a confidence level in statistics; then $\mathcal{R}(X) = \mu(X) + \lambda\sigma(X)$. The alternative idea that the inequality should hold at least with a certain probability $\alpha \in (0, 1)$ corresponds to $q_\alpha(X) \leq C$ with $q_\alpha(X)$ denoting the α-quantile of X, whereas insisting that $X \leq C$ almost surely can be written as $\sup X \leq C$ with $\sup X$ standing for the essential supremum of X. Then $\mathcal{R}(X) = q_\alpha(X)$ or $\mathcal{R}(X) = \sup X$, respectively. However, these examples are only the initial possibilities among many for which pros and cons need to be appreciated. Axioms laying out sensible standards for a measure of risk, such the *coherency* introduced in Artzner et al. [1999], are vital for that.

Another idea in dealing with uncertainty in a random variable X is to quantify its nonconstancy through a *measure of deviation* $\mathcal{D}$, with $\mathcal{D}(X)$ then being a generalization of $\sigma(X)$. Again, axioms have to be articulated. That holds similarly for the notion of a *measure of regret* $\mathcal{V}$, which specifies a value $\mathcal{V}(X)$ standing for the net displeasure perceived in the potential mix of outcomes of a random "cost" X which may sometimes be > 0 (bad) and sometimes ≤ 0 (OK or better).[3]

Regret comes up in penalty approaches to constraints in stochastic optimization and, in mirror image, corresponds to measures of "utility" $\mathcal{U}$ in a context of gains Y instead of losses X, which is typical in economics: $\mathcal{V}(X) = -\mathcal{U}(-X), \mathcal{U}(Y) = -\mathcal{V}(-Y)$. Regret naturally has $\mathcal{V}(0) = 0$, so in this pairing we have to focus on utility measures that have $\mathcal{U}(0) = 0$; we say then that $\mathcal{U}$ is a *measure of relative utility*. The interpretation is that, in applying $\mathcal{U}$ to Y, we are thinking of Y not as absolute gain but gain relative to some threshold, e.g., $Y = Y_0 - B$ where Y_0 is absolute gain and B is a benchmark. Focusing on relative utility in this sense is a positive feature of the quadrangle scheme because it can help to capture the sharp distinction in attitude toward outcomes above or below a benchmark that is increasingly acknowledged as influencing the preferences of decision makers.[4]

Measures of deviation $\mathcal{D}$ and measures of regret $\mathcal{V}$ are deeply related to measures of risk $\mathcal{R}$, and one of our tasks will to bring this all out. Especially important will be a one-to-one correspondence between measures of deviation and measures of risk under "aversity," regardless of coherency.

$\mathcal{R}(X)$ provides a numerical surrogate for the overall hazard in X,
$\mathcal{D}(X)$ measures the "nonconstancy" in X as its uncertainty,
$\mathcal{E}(X)$ measures the "nonzeroness" in X,
$\mathcal{V}(X)$ measures the "regret" in facing the mix of outcomes of X,
$\mathcal{S}(X)$ is the "statistic" associated with X through $\mathcal{E}$ or equivalently $\mathcal{V}$.

Diagram 2: The Quantifications in the Quadrangle.

In the realm of statistical estimation, broader approaches to regression than classical "least-squares" are central to our theme. Regression is a way of approximating a random variable Y by a function $f(X_1, \ldots, X_n)$ of one or more other random variables X_j for purposes of anticipating outcome properties or trends. It requires a way of measuring how far the random difference $Z_f = Y - f(X_1, \ldots, X_n)$ is from 0. This leads us to speak of a *measure of error* $\mathcal{E}$ as assigning to a random variable X a value $\mathcal{E}(X)$ that quantifies the nonzeroness in X. Classical examples would be the $\mathcal{L}^p$-norms

$$||X||_1 = E|X|, \quad ||X||_p = [E(|X|^p)]^{1/p} \text{ for } p \in (1, \infty), \quad ||X||_\infty = \sup |X|,$$

but there is much more to think of besides norms. There may be incentive for using asymmetric error measures $\mathcal{E}$ that look at more than just $|Z_f|$. Indeed, when Y has cost or hazard orientation, underestimations $Y - f(X_1, \ldots, X_n) > 0$ may be more dangerous than overestimations $Y - f(X_1, \ldots, X_n) < 0$.

[3] In financial terms, if X and $\mathcal{V}(X)$ have units of money, $\mathcal{V}(X)$ can be the compensation deemed appropriate for taking on the burden of the uncertain loss X.

[4] This will be discussed in more detail in Section 4 in the case of utility expressions $\mathcal{U}(Y) = E[u(Y)]$ for an underlying function u. Havine $\mathcal{U}(0) = 0$ corresponds to having $u(0) = 0$, which can be achieved by selecting a benchmark and shifting the graph of a given "absolute" utility so that benchmark point is at the origin of $\mathbb{R}^2$.

There is every reason then to think that the choice of a measure $\mathcal{E}$ might best be tailored to such dangers in estimation through the choice of a risk measure $\mathcal{R}$, but how? That could rely on the partnering between deviation measures and risk measures $\mathcal{R}$ and the following idea for "projecting" error measures $\mathcal{E}$ onto deviation measures $\mathcal{D}$.

Given an error measure $\mathcal{E}$ and a random variable X, one can look for a constant C that is nearest to X in the sense of minimizing $\mathcal{E}(X - C)$. The resulting minimum "$\mathcal{E}$-distance," denoted by $\mathcal{D}(X)$, turns out to define a deviation measure (under assumptions explained later). The C value giving the minimum, denoted by $\mathcal{S}(X)$, can be called the "*statistic*" associated with X by $\mathcal{E}$. The case of $\mathcal{E}(X) = ||X||_2$ produces $\mathcal{S}(X) = \mu(X)$ and $\mathcal{D}(X) = \sigma(X)$, but many other possibilities will soon come on stage.

Remarkably, the "statistic" story does not end there, because the scheme for projecting $\mathcal{E}$ to $\mathcal{D}$ is echoed by a certainty-uncertainty *trade-off formula* which projects a regret measure $\mathcal{V}$ onto a risk measure $\mathcal{R}$. This formula, in which $C + \mathcal{V}(X - C)$ is minimized over C, generalizes a rule in Rockafellar and Uryasev [2000], Rockafellar and Uryasev [2002], for VaR-CVaR computations. It extends the insights gained beyond that by Ben-Tal and Teboulle [2007] in a context of expected utility, and lines up with still broader expressions for risk in Krokhmal [2007]. Under a simple relationship between $\mathcal{V}$ and $\mathcal{E}$, the optimal C value in the trade-off is the same statistic $\mathcal{S}(X)$ as earlier, but that conceptual bond has been missed until now.

Altogether, we arrive in this way at a "quadrangle" of quantifications having the descriptions in Diagram 2 and the interconnections in Diagram 3.[5] More details will be furnished in Section 3, after the assumptions needed to justify the relationships have been explained.

$$\mathcal{D}(X) = \mathcal{R}(X) - EX, \qquad \mathcal{R}(X) = EX + \mathcal{D}(X)$$

$$\mathcal{E}(X) = \mathcal{V}(X) - EX, \qquad \mathcal{V}(X) = EX + \mathcal{E}(X)$$

$$\mathcal{D}(X) = \min_C \{\, \mathcal{E}(X - C) \,\}, \qquad \mathcal{R}(X) = \min_C \{\, C + \mathcal{V}(X - C) \,\}$$

$$\mathcal{S}(X) = \operatorname*{argmin}_C \{\, \mathcal{E}(X - C) \,\} = \operatorname*{argmin}_C \{\, C + \mathcal{V}(X - C) \,\}$$

Diagram 3: The General Relationships

Note that Diagram 3 says nothing about the *inverse problem* of identifying for a given risk measure $\mathcal{R}$ a "natural" regret measure $\mathcal{V}$ that yields it in trade-off, or for a given deviation measure $\mathcal{D}$ a "natural" error measure $\mathcal{E}$ that projects to it (these two questions being equivalent through the pairing between $\mathcal{R}$ and $\mathcal{D}$ and the connection at the bottom between $\mathcal{V}$ and $\mathcal{E}$). That is a large topic with many good answers in examples in Section 2 and broader principles in Section 3.

Additional interpretations and results are developed in Section 4. Section 5 explains how the risk quadrangle enters applications in optimization under uncertainty and generalized statistical regression.

Duality is another large topic, and it will occupy our attention in Section 6. Each of the quantifiers $\mathcal{R}$, $\mathcal{D}$, $\mathcal{E}$, $\mathcal{V}$, has a dual expression in the presence of "closed convexity," a property we will build into them in Section 3. That leads to characterization of risk through "risk envelopes" and expressions of "generalized entropy," and it thereby sheds additional light on modeling motivations.

[5]The "argmin" notation refers to the C values that achieve the "min."

Expectation Quadrangles. Many examples, but by no means all, will fall into the category that we call the *expectation case* of the risk quadrangle. The special feature in this case is that

$$\mathcal{E}(X) = E[e(X)], \qquad \mathcal{V}(X) = E[v(X)], \qquad \mathcal{U}(Y) = E[u(y)], \tag{1.1}$$

for functions e and v on $(-\infty, \infty)$ related to each other by

$$e(x) = v(x) - x, \qquad v(x) = e(x) + x, \tag{1.2}$$

and on the other hand, v corresponding to relative utility u through

$$v(x) = -u(-x), \qquad u(y) = -v(-y). \tag{1.3}$$

The $\mathcal{V} \leftrightarrow \mathcal{E}$ correspondence in Diagram 3 holds under (1.2), while (1.3) ensures that $\mathcal{V}(X) = -\mathcal{U}(-X)$ and $\mathcal{U}(Y) = -\mathcal{V}(-Y)$. The consequences for the $\mathcal{S}$, $\mathcal{R}$ and $\mathcal{D}$ components of the quadrangle, as generated by the other formulas in Diagram 3, will be discussed in Section 4.

Expected utility is a central notion in decision analysis in economics and likewise in finance, cf. Föllmer and Schied [2004]. Expected error expressions similarly dominate much of statistics, cf. Gneiting [2011]. Expectation quadrangles provide the connection to those bodies of theory in the development undertaken here.

However, the quadrangle scheme also reveals serious limitations of the expectation case. Many interesting examples do not fit into it, as will be clear in the sampling of Section 2. Departure from expected utility and expected error is therefore inevitable, if the quadrangle relationships we are exploring are to reach their full potential for application.

2 Some Examples Showing the Breadth of the Scheme

Before going into technical details, we will look at an array of examples aimed at illustrating the scope and richness of the quadrangle scheme and the interrelationships it reveals. In each case the elements correspond to each other in the manner of Diagram 3. Some of the connections are already known, but they have not all previously been placed in a single, comprehensive picture.

The first example ties classical safety margins in the risk measure format to the standard tools of least-squares regression. It centers on the mean value of X as the statistic. The scaling factor $\lambda > 0$ allows the safety margin to come into full play.

Example 1: The Mean-Based Quadrangle (with $\lambda > 0$ as a scaling parameter)

$\mathcal{S}(X) = EX = \mu(X) =$ mean

$\mathcal{R}(X) = EX + \lambda\,\sigma(X) =$ safety margin tail risk

$\mathcal{D}(X) = \lambda\,\sigma(X) =$ standard deviation, scaled

$\mathcal{V}(X) = EX + \lambda||X||_2 = L^2$-regret, scaled

$\mathcal{E}(X) = \lambda||X||_2 = L^2$-error, scaled

Already here we have an example outside the expectation case. Perhaps that may seem a bit artificial, because the $\mathcal{L}^2$-norm could be replaced by its square. That would produce a modified quadrangle giving the same statistic:

Example 1′: Variance Version of Example 1

$$\mathcal{S}(X) = EX = \mu(X)$$
$$\mathcal{R}(X) = EX + \lambda\,\sigma^2(X)$$
$$\mathcal{D}(X) = \lambda\,\sigma^2(X)$$
$$\mathcal{V}(X) = EX + \lambda||X||_2^2 = E[v(X)] \ \text{ for } v(x) = x + \lambda x^2$$
$$\mathcal{E}(X) = \lambda||X||_2^2 = E[e(x)] \ \text{ for } e(x) = \lambda x^2$$

However, some properties would definitely change. The first version has $\mathcal{R}(X+X') \leq \mathcal{R}(X)+\mathcal{R}(X')$, which is a property often promoted for measures of risk, but this fails for the second version.

The next example combines newer statistical elements with concepts coming recently from risk management. By tying "conditional value-at-risk," on the optimization side, to quantile regression as pioneered in Koenker and Bassett [1978], it underscores the unity that might go unrecognized without the risk quadrangle scheme.

The key in this case is provided by the (cumulative) distribution function $F_X(x) = \text{prob}\{\,X \leq x\,\}$ of a random variable X and the quantile values associated with it. If, for a probability level $\alpha \in (0,1)$, there is a unique x such that $F_X(x) = \alpha$, that x is the α-quantile $q_\alpha(X)$. In general, however, there are two values to consider as extremes:

$$q_\alpha^+(X) = \inf\{\,x \,|\, F_X(x) > \alpha\,\}, \qquad q_\alpha^-(X) = \sup\{\,x \,|\, F_X(x) < \alpha\,\}. \tag{2.1}$$

It is customary, when these differ, to take the lower value as "the" α-quantile, noting that, because F_X is right-continuous, this is the lowest x such that $F_X(x) = \alpha$. Here, instead, we will consider the entire *interval* between the two competing values as the quantile,

$$q_\alpha(X) = [q_\alpha^-(X), q_\alpha^+(X)], \tag{2.2}$$

bearing in mind that this interval usually collapses to a single value. That approach will fit better with our way of defining a "statistic" by the argmin notation. Also important to understand, in our context of interpreting X as a "cost" or "loss," is that the notion of *value-at-risk* in finance coincides with quantile. There is an upper value-at-risk $\text{VaR}_\alpha^+(X) = q_\alpha^+(X)$ along with a lower value-at-risk $\text{VaR}_\alpha^-(X) = q_\alpha^-(X)$, and, in general, a value-at-risk *interval* $\text{VaR}_\alpha(X) = [\text{VaR}_\alpha^+(X), \text{VaR}_\alpha^-(X)]$ identical to the quantile interval $q_\alpha(X)$.

Besides value-at-risk, the example coming under consideration involves the *conditional value-at-risk* of X at level $\alpha \in (0,1)$ as defined by

$$\text{CVaR}_\alpha(X) = \text{ expectation of } X \text{ in its } \alpha\text{-tail}, \tag{2.3}$$

which is also expressible by

$$\text{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\tau^1 \text{VaR}_\tau(X)d\tau. \tag{2.4}$$

The second formula is due to Acerbi [2002] in different terminology, while the first formula follows the pattern in Rockafellar and Uryasev [2000], where "conditional value-at-risk" was coined.[6] Due to applications of risk theory in areas outside of finance, such as reliability engineering, we believe it is

[6] The α-tail distribution of X corresponds to the upper part of the distribution of X having probability $1-\alpha$. The interpretation of this for the case when F_X has a jump at the α quantile is worked out in Rockafellar and Uryasev [2002].

advantageous to maintain, parallel to value-at-risk and quantile, the ability to refer to the conditional value-at-risk $\mathrm{CVaR}_\alpha(X)$ equally as the *superquantile* $\bar{q}_\alpha(X)$.

We will be helped here and later by the notation

$$X = X_+ - X_- \text{ with } X_+ = \max\{0, X\},\ X_- = \max\{0, -X\}.$$

Example 2: The Quantile-Based Quadrangle (at any confidence level $\alpha \in (0,1)$)

$$\mathcal{S}(X) = \mathrm{VaR}_\alpha(X) = q_\alpha(X) = \text{quantile}$$

$$\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X) = \bar{q}_\alpha(X) = \text{superquantile}$$

$$\mathcal{D}(X) = \mathrm{CVaR}_\alpha(X - EX) = \bar{q}_\alpha(X - EX) = \text{superquantile-deviation}$$

$$\mathcal{V}(X) = \tfrac{1}{1-\alpha} EX_+ = \text{a penalty expression for regret as scaled average loss}^7$$

$$\mathcal{E}(X) = E[\tfrac{\alpha}{1-\alpha} X_+ + X_-] = \text{normalized Koenker-Bassett error}$$

This is an expectation quadrangle with

$$e(x) = \frac{\alpha}{1-\alpha}\max\{0,x\} + \max\{0,-x\}, \quad v(x) = \frac{1}{1-\alpha}\max\{0,x\}, \quad u(y) = \frac{1}{1-\alpha}\min\{0,y\}.$$

The original Koenker-Bassett error expression differs from the one here by a positive factor. Adjustment is needed to make it project to the desired $\mathcal{D}$.

The special case of Example 2 in which the quantile is the median is worth looking at directly. It corresponds to the error measure being the $\mathcal{L}^1$-norm in contrast to Example 1, where the error measure was the $\mathcal{L}^2$-norm.

Example 3: The Median-Based Quadrangle (the quantile case for $\alpha = 1/2$)

$$\mathcal{S}(X) = \mathrm{VaR}_{1/2}(X) = q_{1/2}(X) = \text{median}$$

$$\mathcal{R}(X) = \mathrm{CVaR}_{1/2}(X) = \bar{q}_{1/2}(X) = \text{“supermedian” (average in tail above median)}$$

$$\mathcal{D}(X) = \mathrm{CVaR}_{1/2}(X - EX) = \bar{q}_{1/2}(X - EX) = \text{median-deviation}$$

$$\mathcal{V}(X) = 2E[X_+] = \mathcal{L}^1\text{-regret}$$

$$\mathcal{E}(X) = E|X| = \mathcal{L}^1\text{-error}$$

For the sake of comparison, it is instructive to ask what happens if the error measure $\mathcal{E}$ is the $\mathcal{L}^\infty$-norm. This leads to our fourth example, which targets the case where X is (essentially) bounded.

Example 4: The Range-Based Quadrangle

$$\mathcal{S}(X) = \tfrac{1}{2}[\sup X + \inf X] = \text{center of range of } X \text{ (if bounded)}$$

$$\mathcal{R}(X) = EX + \tfrac{1}{2}[\sup X - \inf X] = \text{range-buffered risk}$$

$$\mathcal{D}(X) = \tfrac{1}{2}[\sup X - \inf X] = \text{radius of the range of } X \text{ (maybe } \infty)$$

$$\mathcal{V}(X) = EX + \sup|X| = \mathcal{L}^\infty\text{-regret}$$

$$\mathcal{E}(X) = \sup|X| = \mathcal{L}^\infty\text{-error}$$

This is *not* an expectation quadrangle.

7 Average loss as “regret,” and as inspiration for the terminology we are introducing here more broadly, goes back to Dembo and King [1992] in stochastic programming.

The example offered next identifies both as the statistic and as the risk the "worst cost" associated with X. It can be regarded as the limit of the quantile-based quadrangle as $\alpha \to 1$.

Example 5: The Worst-Case-Based Quadrangle

$$\mathcal{S}(X) = \sup X = \text{top of the range of } X \text{ (maybe } \infty)$$

$$\mathcal{R}(X) = \sup X = \text{yes, the same as } \mathcal{S}(X)$$

$$\mathcal{D}(X) = \sup X - EX = \text{span of the upper range of } X \text{ (maybe } \infty)$$

$$\mathcal{V}(X) = \begin{cases} 0 & \text{if } X \leq 0 \\ \infty & \text{if } X \not\leq 0 \end{cases} = \text{worst-case-regret}$$

$$\mathcal{E}(X) = \begin{cases} E|X| & \text{if } X \leq 0 \\ \infty & \text{if } X \not\leq 0 \end{cases} = \text{worst-case-error}$$

This is another expectation quadrangle but with functions of unusual appearance:

$$e(x) = \begin{cases} x & \text{if } x \leq 0 \\ \infty & \text{if } x > 0 \end{cases}, \qquad v(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \infty & \text{if } x > 0 \end{cases}, \qquad u(y) = \begin{cases} -\infty & \text{if } y < 0 \\ 0 & \text{if } y \geq 0 \end{cases}.$$

The "range" of X here is its essential range, i.e., the smallest closed interval in which outcomes must lie with probability 1. Thus, in Example 5, the inequality $\mathcal{R}(X) \leq C$ gives the risk-measure code for insisting that $X \leq C$ with probability 1. The regret measure $\mathcal{V}(X)$ assigns infinite penalty when this is violated, but no disincentive otherwise.

An interesting generalization of Example 5 can be made in which the appraisal of "worst" is distributed over different visions of the future which are tied to a coarse level of probability modeling. The details will not be fully understandable until we begin posing risk in the rigorous framework of a probability space in Section 3 (and all the more in Section 6), but we proceed anyway here to an initial formulation. It depends on partitioning the underlying uncertainty about the future into several different "sets of circumstances" $k = 1, \ldots, r$ having no overlap[8] and letting

$$\begin{aligned} p_k = {} & \text{probability of the } k\text{th set of circumstances, with } p_k > 0,\ p_1 + \cdots + p_r = 1, \\ \sup\nolimits_k X = {} & \text{worst of } X \text{ under circumstances } k, \text{ for } k = 1, \ldots, r, \\ E_k X = {} & \text{conditional expectation of } X \text{ under circumstances } k. \end{aligned} \tag{2.5}$$

The last implies, of course, that $p_1 E_1 X + \cdots + p_r E_r X = EX$.

Example 6: Distributed-Worst-Case-Based Quadrangle (with respect to (2.5))

$$\mathcal{S}(X) = p_1 \sup\nolimits_1 X + \cdots + p_r \sup\nolimits_r X$$

$$\mathcal{R}(X) = p_1 \sup\nolimits_1 X + \cdots + p_r \sup\nolimits_r X = \text{yes, the same as } \mathcal{S}(X)$$

$$\mathcal{D}(X) = p_1[\sup\nolimits_1 X - E_1 X] + \cdots + p_r[\sup\nolimits_r X - E_r X]$$

$$\mathcal{V}(X) = \begin{cases} 0 & \text{if } p_1 \sup_1 X + \cdots + p_r \sup_r X \leq 0, \\ \infty & \text{otherwise} \end{cases}$$

$$\mathcal{E}(X) = \begin{cases} E|p_1 E_1 X + \cdots + p_r E_r X| & \text{if } p_1 \sup_1 X + \cdots + p_r \sup_r X \leq 0, \\ \infty & \text{otherwise} \end{cases}$$

Again, this is *not* an expectation quadrangle. Moreover, unlike the previous cases, the quantifiers in Example 6 are not "law-invariant," i.e., their effects on X depend on more than just the distribution function F_X. It should be noted that expectations only enter the elements on the right side of this

[8] Technically this refers to "events" as measurable subsets Ω_k of the probability space Ω introduced in the next section.

quadrangle. As far as optimization is concerned, by itself, there are no assumptions about probability structure other than the first line of (2.6). This can be regarded as a compromise between the starkness of Example 5 and a full-scale probability model.

An expectation quadrangle which, in a sense, interpolates between Examples 1′ and 3, looks at an error expression like the one in Huber's modification of least-squares regression in order to mollify the influence of outliers. Here we introduce a scaling parameter $\beta > 0$ and make use of the β-*truncation* function

$$T_\beta(x) = \begin{cases} \beta & \text{when } x \geq \beta, \\ x & \text{when } -\beta \leq x \leq \beta, \\ -\beta & \text{when } x \leq -\beta. \end{cases}$$

Example 7: The Truncated-Mean-Based Quadrangle (with scaling parameter $\beta > 0$)

$$\mathcal{S}(X) = \mu_\beta(X) = \text{value of } C \text{ such that } E[T_\beta(X - C)] = 0$$

$$\mathcal{R}(X) = \mu_\beta(X) + E[v(X - \mu_\beta(X))] \ \text{ for } v \text{ as below}$$

$$\mathcal{D}(X) = E[e(X - \mu_\beta(X))] \ \text{ for } e \text{ as below}$$

$$\mathcal{V}(X) = E[v(X)] \ \text{ with } v(x) = \begin{cases} -\frac{\beta}{2} & \text{when } x \leq -\beta \\ x + \frac{1}{2\beta}x^2 & \text{when } |x| \leq \beta \\ 2x - \frac{\beta}{2} & \text{when } x \geq \beta \end{cases}$$

$$\mathcal{E}(X) = E[e(X)] \ \text{ with } e(x) = \begin{cases} |x| - \frac{\beta}{2} & \text{when } |x| \geq \beta \\ \frac{1}{2\beta}x^2 & \text{when } |x| \leq \beta \end{cases} \qquad \text{Huber-type error}$$

In the limit of $\mu_\beta(X)$ as $\beta \to \infty$, we end up with just $EX = \mu(X)$, as in Examples 1 and 1′. For the deviation measure $\mathcal{D}$ in Example 7, one can think of $2\beta\mathcal{D}(X)$ as a sort of β-modification of $\sigma^2(X)$ which approaches that variance as $\beta \to \infty$.

The next quadrangle looks very different. The log-exponential risk measure it incorporates is a recognized tool in risk theory in finance, but its connection with a form of generalized regression, by way of the the right side of the quadrangle, has not previously been contemplated.

Example 8: The Log-Exponential-Based Quadrangle

$$\mathcal{S}(X) = \log E[\exp X] = \text{expression dual to Boltzmann-Shannon entropy}$$ [9]

$$\mathcal{R}(X) = \log E[\exp X] = \text{yes, the same as } \mathcal{S}(X)$$

$$\mathcal{D}(X) = \log E[\exp(X - EX)] = \text{log-exponential deviation}$$

$$\mathcal{V}(X) = E[\exp X - 1] \text{ regret} \quad \longleftrightarrow \quad \text{utility } \mathcal{U}(Y) = E[1 - \exp(-Y)]$$

$$\mathcal{E}(X) = E[\exp X - X - 1] = \text{exponential error}$$

The regret in Example 8 is paired with an expected utility expression that is commonly employed in finance: we are in the expectation case with

$$v(x) = \exp x - 1 \quad \longleftrightarrow \quad u(y) = 1 - \exp(-y).$$

Such utility pairing is seen also in the example coming next, which fits the expectation case with

$$e(x) = \begin{cases} \log \frac{1}{1-x} - x & \text{if } x < 1 \\ \infty & \text{if } x \geq 1, \end{cases} \quad v(x) = \begin{cases} \log \frac{1}{1-x} & \text{if } x < 1 \\ \infty & \text{if } x \geq 1, \end{cases} \quad u(y) = \begin{cases} \log(1+y) & \text{if } y > -1 \\ -\infty & \text{if } y \leq -1. \end{cases}$$

[9]The "exp" notation is adopted so as not to conflict with the convenient use of "e" for error integrands in (1.1).

Example 9: The Rate-Based Quadrangle

$$\mathcal{S}(X) = r(X) = \text{ unique } C \geq \sup X - 1 \text{ such that } E\Big[\tfrac{1}{1-X+C}\Big] = 1$$

$$\mathcal{R}(X) = r(X) + E\Big[\log \tfrac{1}{1-X+r(X)}\Big]$$

$$\mathcal{D}(X) = r(X) + E\Big[\log \tfrac{1}{1-X+r(X)} - X\Big]$$

$$\mathcal{V}(X) = E\Big[\log \tfrac{1}{1-X}\Big] \text{ regret } \quad \longleftrightarrow \quad \text{utility } \mathcal{U}(Y) = E[\log(1+Y)]$$

$$\mathcal{E}(X) = E\Big[\log \tfrac{1}{1-X} - X\Big]$$

We have dubbed this quadrangle "rate-based" because, in the utility connection, $\log(1 + y)$ is an expression applied to a rate of gain y (which of necessity is > -1). Correspondingly in $\log \frac{1}{1-x}$, we are dealing with a rate of loss.

The remaining two examples in this section lie outside the expectation case and present a more complicated picture where error and regret are defined by an auxiliary operation of minimization. The first concerns "mixed" quantiles/VaR and superquantiles/CVaR, which have an interesting role in expressing preferences toward the handling of risk. The idea, from the risk measure perspective, is to study expressions of the type

$$\mathcal{R}(X) = \int_0^1 \text{CVaR}_\alpha(X) d\lambda(\alpha) \tag{2.6}$$

for any weighting measure λ on $(0, 1)$ (nonnegative with total measure 1). In particular, if λ is comprised of atoms with weights $\lambda_k > 0$ at points α_k for $k = 1, \dots, r$, with $\lambda_1 + \cdots + \lambda_r = 1$, one gets

$$\mathcal{R}(X) = \lambda_1 \text{CVaR}_{\alpha_1}(X) + \cdots + \lambda_r \text{CVaR}_{\alpha_r}(X). \tag{2.7}$$

The question is whether this can be placed in a full quadrangle in the format of Diagrams 1, 2 and 3.

Motivation comes from the fact that such risk measures have a representation as "spectral measures" in the sense of Acerbi [2002], which capture preferences in terms of "risk profiles."[10] We proved in [Rockafellar et al., 2006a, Proposition 5] (echoing our working paper Rockafellar et al. [2002]) that, as long as the weighting measure λ satisfies $\int_0^1 \alpha^{-1} d\lambda(\alpha) < \infty$, the $\mathcal{R}(X)$ in (2.7) can equivalently be expressed in the form

$$\mathcal{R}(X) = \int_0^1 \text{VaR}_\tau(X)\phi(\tau)d\tau \text{ with } \phi(\tau) = \int_{[\tau,1)} \alpha^{-1} d\lambda(\alpha), \tag{2.8}$$

where the function ϕ, defined on $(0, 1)$, gives the *risk profile.*[11]

The risk profile for a single "unmixed" risk measure CVaR_α is the function ϕ_α that has the value $1/(1 - \alpha)$ on $[\alpha, \infty)$ but 0 on $(0, \alpha)$; this corresponds to formula (2.4). Moreover the risk profile for a weighted CVaR sum as in (2.7) would be the step function $\phi = \lambda_1\phi_{\alpha_1} + \cdots + \lambda_r\phi_{\alpha_r}$.

Although the quadrangle that would serve for a general weighting measure in (2.6) is still a topic of research, the special case in (2.7) is accessible from the platform of Rockafellar et al. [2008], which will be widened in Section 4 (in the Mixing Theorem).

[10] Such profiles occur in "dual utility theory," a subject addressed by Yaari [1987] and Roell [1987].

[11] This function ϕ is left-continuous and nonincreasing with $\phi(0^+) < \infty$, $\phi(1^-) = 0$ and $\int_0^1 \phi(\tau)d\tau = 1$. Conversely, any function ϕ with those properties arises from a unique choice of λ as described.

Example 10: The Mixed-Quantile-Based Quadrangle
(for any confidence levels $\alpha_i \in (0,1)$ and weights $\lambda_k > 0$, $\sum_{k=1}^r \lambda_k = 1$)

$$\mathcal{S}(X) = \textstyle\sum_{k=1}^r \lambda_k q_{\alpha_k}(X) = \sum_{k=1}^r \lambda_k \mathrm{VaR}_{\alpha_k}(X) = \text{a mixture}^{12} \text{ of quantiles of } X$$

$$\mathcal{R}(X) = \textstyle\sum_{k=1}^r \lambda_k \bar{q}_{\alpha_k}(X) = \sum_{k=1}^r \lambda_k \mathrm{CVaR}_{\alpha_k}(X)$$
$$= \text{the corresponding mixture of superquantiles of } X$$

$$\mathcal{D}(X) = \textstyle\sum_{k=1}^r \lambda_k \bar{q}_{\alpha_k}(X - EX) = \sum_{k=1}^r \lambda_k \mathrm{CVaR}_{\alpha_k}(X - EX)$$
$$= \text{the corresponding mixture of superquantile deviations of } X$$

$$\mathcal{V}(X) = \min_{B_1,\dots,B_r} \Big\{ \textstyle\sum_{k=1}^r \lambda_k \mathcal{V}_{\alpha_k}(X - B_k) \,\Big|\, \sum_{k=1}^r \lambda_k B_k = 0 \Big\}$$
$$= \text{a derived balance of the regrets } \mathcal{V}_{\alpha_k}(X) = \tfrac{1}{1-\alpha_k} EX_+$$

$$\mathcal{E}(X) = \min_{B_1,\dots,B_r} \Big\{ \textstyle\sum_{k=1}^r \lambda_k \mathcal{E}_{\alpha_k}(X - B_k) \,\Big|\, \sum_{k=1}^r \lambda_k B_k = 0 \Big\}$$
$$= \text{a derived balance of the errors } \mathcal{E}_{\alpha_k}(X) = E[\tfrac{\alpha_k}{1-\alpha_k} X_+ + X_-\,]$$

The case of a general weighting measure may be approximated by this arbitrarily closely, as can very well be seen through the corresponding risk profiles. When the measure is concentrated in finitely many points, the corresponding profile function ϕ in (2.9) is a step function, and vice versa, as already noted. An arbitrary profile function ϕ (fulfilling the conditions indicated above in a footnote) can be approximated by a profile function that is a step function.

The final example offers something new as far as risk measures and potential applications in regression are concerned, although the "statistic" in question has already come up in mortgage pipeline hedging; see Aorda [2010].

Example 11: The Quantile-Radius-Based Quadrangle (for any $\alpha \in (1/2, 1)$)

$$\mathcal{S}(X) = \tfrac{1}{2}[q_\alpha(X) - q_{1-\alpha}(X)] = \tfrac{1}{2}[\mathrm{VaR}_\alpha(X) - \mathrm{VaR}_{1-\alpha}(X)]$$
$$= \text{the } \alpha\text{-quantile radius of } X, \text{ or } \tfrac{1}{2} \text{ two-tail-VaR}_\alpha \text{ of } X$$

$$\mathcal{R}(X) = EX + \tfrac{1}{2}[\bar{q}_\alpha(X) + \bar{q}_\alpha(-X)] = EX + \tfrac{1}{2}[\mathrm{CVaR}_\alpha(X) + \mathrm{CVaR}_\alpha(-X)]$$
$$= \text{reverted CVaR}_\alpha$$

$$\mathcal{D}(X) = \tfrac{1}{2}[\bar{q}_\alpha(X) + \bar{q}_\alpha(-X)] = \tfrac{1}{2}[\mathrm{CVaR}_\alpha(X) + \mathrm{CVaR}_\alpha(-X)]$$
$$= \text{the } \alpha\text{-superquantile radius of } X$$

$$\mathcal{V}(X) = EX + \min_B \Big\{ \tfrac{1}{2(1-\alpha)} E\Big[[B + X]_+ + [B - X]_+ \Big] - B \Big\}$$
$$= \alpha\text{-quantile-radius regret in } X$$

$$\mathcal{E}(X) = \tfrac{1}{2(1-\alpha)} \min_B E\Big[[B + X]_+ + [B - X]_+\Big]$$
$$= \alpha\text{-quantile-radius error in } X$$

This example will be justified and extended in Section 3 (through the Reverting Theorem).

What considerations have to be faced in constructing further quadrangle examples? For instance, is there a full quadrangle with $\mathcal{R}(X) = EX$, or with $\mathcal{R}(X) = \mathrm{VaR}_\alpha(X) = q_\alpha(X)$? The answer is yes in both cases, provided that $\mathrm{VaR}_\alpha(X)$ and $q_\alpha(X)$ (which can be intervals in our setting) are replaced by $\mathrm{VaR}_\alpha^-(X)$ and $q_\alpha^-(X)$, say, but the resulting quadrangles are "not interesting." For $\mathcal{R}(X) = EX$, we must have $\mathcal{D}(X) \equiv 0$ in accordance with Diagram 3. An associated measure of error would be

[12] This kind of sum, in which some of the terms could be intervals, is to be interpreted in general as referring to all results obtained by selecting particular values within those intervals.

$\mathcal{E}(X) = |EX|$, which is paired with $\mathcal{V}(X) = EX + |EX| = 2\max\{0, EX\}$. Then $\mathcal{S}(X) = EX$ and offers us nothing new.

For $\mathcal{R}(X) = \text{VaR}^-_\alpha(X)$, on the other hand, we have $\mathcal{D}(X) = \text{VaR}^-_\alpha(X - EX)$ and could take $\mathcal{E}(X) = \text{VaR}^-_\alpha(X - EX) + |EX|$ and correspondingly $\mathcal{V}(X) = \text{VaR}^-_\alpha(X) + 2\max\{0, EX\}$. However, then we merely have $\mathcal{S}(X) = EX$. Some different and more interesting $\mathcal{V}(X)$ might project onto $\mathcal{R}(X) = \text{VaR}^-_\alpha(X)$ through the formula in Diagram 3, but this remains to be seen.[13]

In a similar vein, it might be wondered whether the expression $\text{VaR}_\alpha(X) - \text{VaR}_{1-\alpha}(X)$ appearing as the statistic of Example 10 could serve as the deviation measure $\mathcal{D}(X)$ in some quadrangle, since it is nonnegative and vanishes for constant X. Again the answer is yes, but perhaps only trivially.

Anyway, the most important guideline for additional quadrangle examples is that the quantifiers must fit with the descriptions in Diagram 2, which have yet to be fleshed out with appeals to specific mathematical properties. That is our task in the coming section. Those properties have to make sense in applications and lead to a sturdy methodology, and the real trouble with $\mathcal{R}(X) = EX$ and $\mathcal{R}(X) = \text{VaR}^-_\alpha(X)$ as measures of risk is that they fall short of meeting such a standard. The Quadrangle Theorem of the coming Section 3, our central result, will therefore not apply to them.

3 The Main Properties and Relationships

In working with random variables we adopt the standard model in probability theory which interprets them as functions on a probability space. Specifically, we suppose there is an underlying space Ω with elements ω standing for future states, or scenarios, along with a measure which assigns probabilities to various subsets of Ω. There is no loss of generality in this, but technicalities come in which we wish to avoid getting too occupied with at present.[14] Random variables from now on are functions $X : \Omega \to \mathbb{R}$, but we restrict attention to those for which $E[X^2] < \infty$, indicating this by $X \in \mathcal{L}^2(\Omega)$. Here "$E$" is the expectation with respect to the background probability measure on Ω.

Any $X \in \mathcal{L}^2(\Omega)$ also has $E|X| < \infty$, so that $\mu(X) = EX$ is well defined and finite. Furthermore, the variance $\sigma^2(X) = E[X - EX]^2$ and its square root, the standard deviation $\sigma(X)$, are well defined and finite.[15] These expressions characterize the natural ("strong") convergence in $\mathcal{L}^2(\Omega)$ of a sequence of random variables X^k to a random variable X:

$$\begin{aligned} \mathcal{L}^2\text{-}\lim_{k\to\infty} X^k = X \quad &\Longleftrightarrow \quad \lim_{k\to\infty} ||X^k - X||_2 = 0 \\ &\Longleftrightarrow \quad \lim_{k\to\infty} \mu(X^k - X) = 0 \;\text{ and }\; \lim_{k\to\infty} \sigma(X^k - X) = 0. \end{aligned} \tag{3.1}$$

In many applications Ω may be finite, i.e., comprised of finitely many elements ω, each having a positive probability weight. The choice of norm makes no difference then because $\mathcal{L}^2(\omega)$ is finite-dimensional.

[13] No claim is made about there being a *unique* $\mathcal{E}$ projecting onto some $\mathcal{D}$, or a *unique* $\mathcal{V}$ projecting onto some $\mathcal{R}$, and indeed that must not be hoped for. The real issue instead is that of determining an "natural" antecedent with valuable characteristics. For instance, any risk measure $\mathcal{R}$ can be projected from $\mathcal{V}(X) = \mathcal{R}(X) + \lambda|EX|$ and any deviation measure from $\mathcal{E}(X) = \mathcal{D}(X) + \lambda|EX|$ for arbitrary $\lambda > 0$, with the pointless consequence that $\mathcal{S}(X) = EX$.

[14] More explanation is provided in Section 6, which also offers motivation and examples for readers who might not be so familiar with this way of thinking.

[15] It might be wondered why we insist on boundedness of second moments when requiring only $E|X| < \infty$ would cover a larger class of random variables. The main reason is that this leads to a simpler exposition in Section 6, when we come to the dualization of risk in terms of sets of probability densities Q (having $Q \geq 0$, $EQ = 1$). With the finiteness of $E|X|$ as the only requirement we would be limited there to bounded densities Q. It would be better really if we could draw on all possible densities Q, but that would force us to go to the opposite extreme of requiring X to be essentially bounded. The choice made here is a workable compromise.

The quantifiers $\mathcal{R}$, $\mathcal{D}$, $\mathcal{V}$ and $\mathcal{E}$, all of which assign numerical values, possibly including $+\infty$, to random variables X, are said to be "functionals" on $\mathcal{L}^2(\Omega)$. Some of the properties that come up may be shared, so it is expedient to state them in terms of a general functional $\mathcal{F} : \mathcal{L}^2(\Omega) \to (-\infty, \infty]$:

- $\mathcal{F}$ is *convex* if $\mathcal{F}((1-\tau)X + \tau X') \leq (1-\tau)\mathcal{F}(X) + \tau\mathcal{F}(X')$ for all X, X', and $\tau \in (0,1)$.[16]
- $\mathcal{F}$ is *positively homogeneous* if $\mathcal{F}(0) = 0$ and $\mathcal{F}(\lambda X) = \lambda\mathcal{F}(X)$ for all $\lambda \in (0, \infty)$.
- $\mathcal{F}$ is *subadditive* if $\mathcal{F}(X + X') \leq \mathcal{F}(X) + \mathcal{F}(X')$ for all X, X'.
- $\mathcal{F}$ is *monotonic* (nondecreasing, here) if $\mathcal{F}(X) \leq \mathcal{F}(X')$ when $X \leq X'$.[17]
- $\mathcal{F}$ is *closed* if, for all $C \in I\!R$, the set $\{\, X \mid \mathcal{F}(X) \leq C \,\}$ is closed.[18]

Convexity will be valuable for much of what we undertake. Positive homogeneity is a more special property which, in the study of risk, was emphasized more in the past than now. An elementary fact of convex analysis is that

$$\mathcal{F} \text{ convex} + \text{positively homogeneous} \iff \mathcal{F} \text{ subadditive} + \text{positively homogeneous}. \tag{3.2}$$

The combinations in (3.2) are equivalent to *sublinearity*: $\mathcal{F}(\sum_k \lambda_k X_k) \leq \sum_k \lambda_k \mathcal{F}(X_k)$ for $\lambda_k \geq 0$.[19]

Other important consequences of convexity emerge only in combination with closedness. One that will be applied in several ways is the following rule coming out of convex analysis.[20]

$$\begin{aligned}&\text{If } \mathcal{F} \text{ is closed convex, and } X_0,\, Y, \text{ are such that the function } f(t) = \mathcal{F}(X_0 + tY) \text{ is}\\ &\text{bounded above for } t \in [0,\infty), \text{ then } \mathcal{F}(X + tY) \leq \mathcal{F}(X) \text{ for all } X \text{ and } t \in [0,\infty).\end{aligned} \tag{3.3}$$

An immediate consequence, for instance, is that[21]

$$\text{for } \mathcal{F} \text{ closed convex:} \quad \text{if } \mathcal{F}(X) \leq 0 \text{ whenever } X \leq 0, \text{ then } \mathcal{F} \text{ is monotonic.} \tag{3.4}$$

To assist with closedness, it may help to note that this property of $\mathcal{F}$ holds when $\mathcal{F}$ is continuous,[22] and moreover, when $\mathcal{F}$ does not take on ∞, that stronger property is automatic in broad circumstances of interest to us. Namely,[23]

$$\mathcal{F} \text{ is continuous on } \mathcal{L}^2(\Omega) \text{ when } \begin{cases} \mathcal{F} \text{ is finite, convex, and closed, or} \\ \mathcal{F} \text{ is finite, convex, and monotonic, or} \\ \mathcal{F} \text{ is finite, convex, and } \Omega \text{ is finite.} \end{cases} \tag{3.5}$$

Measures of risk. The role of a measure of risk, $\mathcal{R}$, is to assign to a random variable X, standing for an uncertain "cost" or "loss," a numerical value $\mathcal{R}(X)$ that can serve as a surrogate for overall (net) cost or loss. However, the assignment must meet reasonable standards in order to make sense.

[16] In expressions like this, a sum of values in $(-\infty, \infty]$ is ∞ if any of them is ∞. Also, $\lambda\infty = \infty$ for $\lambda > 0$.

[17] This inequality is to be interpreted in the "almost sure" sense, meaning that the set of $\omega \in \Omega$ for which $X(\omega) \leq X'(\omega)$ has probability 1.

[18] This property is also called *lower semicontinuity.* A subset of $\mathcal{L}^2(\Omega)$ is closed when it contains all limits of its sequences in the sense of (3.1). For convex sets, weak limits give the same closure as those strong limits.

[19] Under the convention, if necessary, that $0\infty = 0$.

[20] Apply Theorem 8.6 of Rockafellar [1970] to the function $f(s,t) = \mathcal{F}(sX + tY)$.

[21] Interpret X' as $X - Y$ with $Y \leq 0$ and consider the case of (3.3) with $X_0 = 0$.

[22] Continuity of $\mathcal{F}$ means that $\mathcal{F}(X^k) \to \mathcal{F}(X)$ whenever $X^k \to X$ as in (3.1).

[23] For the first: [Rockafellar, 1974, Corollary 8B]. For the second: [Ruszczyński and Shapiro, 2006a, Proposition 3.1]. For the third: [Rockafellar, 1970, Theorem 10.1], recalling that $\mathcal{L}^2(\Omega)$ is finite-dimensional when Ω is finite.

The class of *coherent* measures of risk has attracted wide attention in finance in this regard. A functional $\mathcal{R}$ belongs to this class, as introduced in Artzner et al. [1999], if it is convex and positively homogeneous (or equivalently by (3.2) subadditive and positively homogeneous), as well as monotonic, and, in addition, satisfies[24]

$$\mathcal{R}(X + C) = \mathcal{R}(X) + C \text{ for all } X \text{ and constants } C. \tag{3.6}$$

Closedness of $\mathcal{R}$ was not mentioned in Artzner et al. [1999], but the context there supposed $\mathcal{R}$ to be finite (and actually Ω finite, too), so that closedness and even continuity of $\mathcal{R}$ were implied by coherency through (3.5). Subsequent researchers considered dropping the positive homogeneity, and with it the term "coherent," cf. Föllmer and Schied [2004], Ruszczyński and Shapiro [2006a].[25] However, without denying the importance of these ideas, we will organize things a bit differently here. The crucial role that EX has in the fundamental risk quadrangle is our guide.

By a *regular measure of risk* we will mean a functional $\mathcal{R}$ with values in $(-\infty, \infty]$ that is *closed convex* with

$$\mathcal{R}(C) = C \text{ for constants } C \tag{3.7}$$

and furthermore

$$\mathcal{R}(X) > EX \text{ for nonconstant } X. \tag{3.8}$$

Property (3.8) is *aversity* to risk.[26] Observe that (3.7) implies the seemingly stronger property (3.6) of Artzner et al. [1999] by the rule in (3.3) and therefore entails

$$\mathcal{R}(X - EX) = \mathcal{R}(X) - EX \text{ for all } X \tag{3.9}$$

in particular. An advantage of stipulating (3.7) in place of (3.6) lies in motivation. The surrogate cost value that a measure of risk should assign to a random variable that always comes out with the value C ought to be C itself.

In all of the Examples 1–11 above, $\mathcal{R}$ is a regular measure of risk. and in Examples 1–3, 5–6, 10–11, $\mathcal{R}$ is positively homogeneous. In Examples 2–3, 5–10, $\mathcal{R}$ is monotonic, but in Examples 1, 4 and 11 it is not. Only the risk measures in Examples 2–3, 5–6 and 10 are coherent in the sense of Artzner et al. [1999]. For $\mathcal{R} = \mathrm{CVaR}_\alpha$, this was perceived from several angles that eventually came together; see Pflug [2000], Acerbi and Tasche [2002] and Rockafellar and Uryasev [2002].

An example of a coherent measure of risk that is not regular is $\mathcal{R}(X) = EX$, which lacks aversity. On the other hand, $\mathcal{R}(X) = \mathrm{VaR}_\alpha^-(X)$ fails to be a regular measure of risk by lacking closedness, convexity and the aversity in (3.8), in general, although it does have positive homogeneity, satisfies (3.6) and is monotonic. It fails to be a coherent measure of risk through the absence of convexity.

Measures of deviation. The role of a measure of deviation, $\mathcal{D}$, is to quantify the nonconstancy (as the uncertainty) in a random variable X. By a *regular measure of deviation* we will mean a functional $\mathcal{D}$ with values in $[0, \infty]$ that is *closed convex* with

$$\mathcal{D}(C) = 0 \text{ for constants } C, \text{ but } \mathcal{D}(X) > 0 \text{ for nonconstant } X. \tag{3.10}$$

The measures of deviation in Examples 1–11 all fit this prescription. Note that symmetry is not required: perhaps $\mathcal{D}(-X) \neq \mathcal{D}(X)$.

[24] A slightly different, but ultimately equivalent property was originally formulated in Artzner et al. [1999].

[25] In our view, the idea behind "coherency" is tied mainly to monotonicity plus convexity. In Rockafellar [2007], risk measures satisfying the axioms of coherency except for this were called *coherent in the extended sense.*

[26] Risk measures satisfying this condition were introduced as *averse* measures of risk in Rockafellar et al. [2006a]. A constant random variable $X \equiv C$ has $\mathcal{R}(X) = EX$ by (3.7).

Measures of error. The role of a measure of error, $\mathcal{E}$, is to quantify the nonzeroness in a random variable X. By a *regular measure of error* we will mean a functional $\mathcal{E}$ with values in $[0, \infty]$ that is *closed convex* with

$$\mathcal{E}(0) = 0 \text{ but } \mathcal{E}(X) > 0 \text{ when } X \not\equiv 0 \tag{3.11}$$

and satisfies for sequences of random variables $\{X_k\}_{k=1}^{\infty}$ the condition that

$$\lim_{k\to\infty} \mathcal{E}(X^k) = 0 \implies \lim_{k\to\infty} EX^k = 0. \tag{3.12}$$

The latter enters in the projection from $\mathcal{E}$ to $\mathcal{D}$ that is featured on the right side of the quadrangle and is equivalent actually to the seemingly stronger property that $\mathcal{E}(X) \geq \psi(EX)$ for a convex function ψ on $(-\infty, \infty)$ having $\psi(0) = 0$ but $\psi(t) > 0$ for $t \neq 0$.[27] In common situations it holds automatically, as for instance when Ω is finite,[28] or in the expectation case with $\mathcal{E}(X) = E[e(X)]$ for a convex function e on $(-\infty, \infty)$ having $e(0) = 0$ but $e(x) > 0$ for $x \neq 0$.[29] In Examples 1–11 every measure of error is regular, but some cases can have $\mathcal{E}(-X) \neq \mathcal{E}(X)$.

Measures of regret and relative utility. The role of a measure of regret, $\mathcal{V}$, is to quantify the displeasure associated with the mixture of potential positive, zero and negative outcomes of a random variable X that stands for an uncertain cost or loss. Regret in this sense is close to the notion of an overall penalty, but it might sometimes come out negative and therefore act as a reward. As mentioned in the introduction, regret is the flip side of relative utility. Measures of regret $\mathcal{V}$ correspond to measures of relative utility $\mathcal{U}$ through

$$\mathcal{V}(X) = -\mathcal{U}(-X), \qquad \mathcal{U}(Y) = -\mathcal{V}(-Y), \tag{3.13}$$

where Y denotes a random variable oriented toward uncertain gain instead of loss. Everything said about regret could be conveyed instead in the language of utility, but that would trigger switches of orientation between loss and gain together with tedious minus signs coming from (3.13).

By a *regular measure of regret* we will mean a functional $\mathcal{V}$ with values in $(-\infty, \infty]$ that is *closed convex* with the aversity property that

$$\mathcal{V}(0) = 0 \text{ but } \mathcal{V}(X) > EX \text{ when } X \not\equiv 0 \tag{3.14}$$

and satisfies for sequences of random variables $\{X^k\}_{k=1}^{\infty}$ the condition that

$$\lim_{k\to\infty} [\mathcal{V}(X^k) - EX^k] = 0 \implies \lim_{k\to\infty} EX^k = 0. \tag{3.15}$$

The limit condition is associated with the similar one in (3.12) and likewise is automatic when Ω is finite or in the expectation case where $\mathcal{V}(X) = E[v(X)]$ for a convex function v on $(-\infty, \infty)$ having $v(0) = 0$ but $v(x) > x$ for $x \neq 0$. All the measures of regret in Examples 1–11 are regular.

As with measures of risk $\mathcal{R}$, there is strong incentive for asking $\mathcal{V}$ also to be monotonic. That property holds for the measures of regret in Examples 2–3, 5–10, but not in Examples 1, 4 and 11.[30]

By a *regular measure of relative utility* we will mean a functional $\mathcal{U}$ having the "flipped" properties that correspond to those of a regular measure of regret $\mathcal{V}$ through (3.13).[31]

---

[27] From (3.12) the function $\psi(t) = \inf \{ \mathcal{E}(X) \mid EX = t \}$ has these properties.

[28] The finite-dimensionality of $\mathcal{L}^2(\Omega)$ and the closed convexity $\mathcal{E}$ in combination with (3.11) ensure then that the lower level sets of $\mathcal{E}$ are compact.

[29] Then $E[e(X)] \geq e(EX)$ by Jensen's Inequality.

[30] There is potential motivation sometimes for working without such monotonicity, as will be explained in Section 5.

[31] More details on this will be provided in Section 4.

Quadrangle Theorem.

(a) *The relations* $\mathcal{D}(X) = \mathcal{R}(X) - EX$ *and* $\mathcal{R}(X) = EX + \mathcal{D}(X)$ *give a one-to-one correspondence between regular measures of risk* $\mathcal{R}$ *and regular measures of deviation* $\mathcal{D}$*. In this correspondence,* $\mathcal{R}$ *is positively homogeneous if and only if* $\mathcal{D}$ *is positively homogeneous. On the other hand,*

$$\mathcal{R} \text{ is monotonic if and only if } \mathcal{D}(X) \leq \sup X - EX \text{ for all } X. \tag{3.16}$$

(b) *The relations* $\mathcal{E}(X) = \mathcal{V}(X) - EX$ *and* $\mathcal{V}(X) = EX + \mathcal{E}(X)$ *give a one-to-one correspondence between regular measures of regret* $\mathcal{V}$ *and regular measures of error* $\mathcal{E}$*. In this correspondence,* $\mathcal{V}$ *is positively homogeneous if and only if* $\mathcal{E}$ *is positively homogeneous. On the other hand,*

$$\mathcal{V} \text{ is monotonic if and only if } \mathcal{E}(X) \leq |EX| \text{ for } X \leq 0. \tag{3.17}$$

(c) *For any regular measure of regret* $\mathcal{V}$*, a regular measure of error* $\mathcal{R}$ *is obtained by*

$$\mathcal{R}(X) = \min_C \Big\{ C + \mathcal{V}(X - C) \Big\}. \tag{3.18}$$

If $\mathcal{V}$ *is positively homogeneous, then* $\mathcal{R}$ *is positively homogeneous, and if* $\mathcal{V}$ *is monotonic, then* $\mathcal{R}$ *is monotonic.*

(d) *For any regular measure of error* $\mathcal{E}$*, a regular measure of deviation* $\mathcal{D}$ *is obtained by*

$$\mathcal{D}(X) = \min_C \Big\{ \mathcal{E}(X - C) \Big\}. \tag{3.19}$$

If $\mathcal{E}$ *is positively homogeneous, then* $\mathcal{D}$ *is positively homogeneous, and if* $\mathcal{E}$ *satisfies the condition in (3.17), then* $\mathcal{D}$ *satisfies the condition in (3.16).*

(e) *In both (c) and (d), as long as the expression being minimized is finite for some* C*, the set of* C *values for which the minimum is attained is a nonempty, closed, bounded interval.*[32] *Moreover when* $\mathcal{V}$ *and* $\mathcal{E}$ *are paired as in (b), the interval comes out the same and gives the associated statistic:*

$$\mathcal{S}(X) = \underset{C}{\text{argmin}} \{ \mathcal{E}(X - C) \} = \underset{C}{\text{argmin}} \{ C + \mathcal{V}(X - C) \}. \tag{3.20}$$

This theorem integrates, in a new and revealing way, various results or partial results that were separately developed elsewhere, and in some cases only for positively homogeneous quantifiers. The correspondence between $\mathcal{R}$ and $\mathcal{D}$ in part (a) was officially presented in Rockafellar et al. [2006a] after being laid out much earlier in the unpublished report Rockafellar et al. [2002].[33] The results in parts (d) and (e) about projecting from $\mathcal{E}$ to $\mathcal{D}$ come from Rockafellar et al. [2008], where they were employed in generalized linear regression.[34] The observation in part (b) immediately translates them to the results in parts (c) and (e) about projecting from $\mathcal{V}$ to $\mathcal{R}$.

---

[32] Typically this interval reduces to a single point.

[33] As in those works, even though they only looked at the positively homogeneous case, the justification of (3.17) follows by applying (3.4) to $\mathcal{F} = \mathcal{R}$. The justification of (3.13) works the same way with $\mathcal{F} = \mathcal{V}$ in (3.4).

[34] The only real effort in the proof of the projection claims is in showing that, when $\mathcal{D}$ comes from (3.19), the minimum over C is attained and $\mathcal{D}$ inherits the closedness of $\mathcal{E}$. This draws on (3.12). The argument in Rockafellar et al. [2008] utilized positive homogeneity, but it is readily generalized as follows through the existence under (3.12) of a convex function ψ with $\psi(0) = 0$, $\psi(t) > 0$ for $t \neq 0$, such that $\mathcal{E}(X) \geq \psi(EX)$. The level sets $\{ t \,|\, \psi(t) \leq c \}$ are then bounded.

Observe first that if a sequence of finite error values $\mathcal{E}(X - C^k)$ approaches the minimum with respect to C, it is a bounded sequence and therefore, since $\mathcal{E}(X - C^k) \geq \psi(EX - C^k)$, the sequence of expected values $E[X - C^k]$ is bounded. Then the sequence $\{ C^k \}_{k=1}^{\infty}$ is bounded, so a subsequence will converge to some C. That C gives the minimum, due to the closedness of $\mathcal{E}$.

Next fix a value $c \in I\!R$ and suppose that $X^k \to X$ with $\mathcal{D}(X^k) \leq c$ for $k = 1, 2, \ldots$. The issue is whether $\mathcal{D}(X) \leq c$. For each k there is a C^k with $\mathcal{D}(X^k) = \mathcal{E}(X^k - C^k)$, and those error values are bounded then by c. In consequence, the sequence of values $E[X^k - C^k]$ is bounded. Since $X^k \to X$, hence $EX^k \to EX$, it follows that a subsequence of $\{ C^k \}_{k=1}^{\infty}$ has to converge to some C, in which case the corresponding subsequence of $\{ X^k - C^k \}_{k=1}^{\infty}$ converges to $X - C$. The closedness of $\mathcal{E}$ ensures that $\mathcal{E}(X - C) \leq c$ and hence $\mathcal{D}(X) \leq c$, as required.

Although the parallel between $\mathcal{E} \to \mathcal{D}$ and $\mathcal{V} \to \mathcal{R}$, which ties the two sides of the quadrangle fully together, is mathematically elementary, it has not come into focus easily despite its conceptual significance. That, especially, is where the theorem innovates. What was absent in the past was the broad concept of a measure of regret and the realization it could anchor a fourth corner in the relationships, thereby serving as a conduit for bringing in "utility" and eventually "entropy."

Risk measure formulas of type (3.18) with accompaniment in (3.20) have gradually emerged without any thought that they might be connected somehow with generalized regression. The first such formula was presented in Rockafellar and Uryasev [2000] and its follow-up Rockafellar and Uryasev [2002],[35]

$$\begin{aligned} \text{CVaR}_\alpha(X) &= \min_C \Big\{ C + \frac{1}{1-\alpha} E[X-C]_+ \Big\}, \\ \text{VaR}_\alpha(X) &= \operatorname*{argmin}_C \Big\{ C + \frac{1}{1-\alpha} E[X-C]_+ \Big\}. \end{aligned} \tag{3.21}$$

We later learned that the "argmin" part of this was already known in the statistics of quantile regression, cf. Koenker and Bassett [1978], Koenker [2005], but with the minimization expression differing from ours by a positive factor; the associated "min" quantity got no attention in that subject. In those days we were mainly occupied with the numerical usefulness of (3.21) in solving problems of stochastic optimization involving VaR and CVaR and were looking no further in the direction of statistics.

Earlier, on a different frontier, the concept of "optimized certainty equivalent" was defined in Ben-Tal and Teboulle [1991] by a trade-off formula very much like the one for getting $\mathcal{S}$ from $\mathcal{V}$ but focused on expected utility ("normalized") and maximization, instead on general regret and minimization. It was applied to problems of optimization in Ben-Tal and Ben-Israel [1991] and subsequently Ben-Tal and Ben-Israel [1997]. Much later in Ben-Tal and Teboulle [2007], once the theory of risk measures had come into development, the "min" quantity in the trade-off received attention alongside of the "argmin," and (3.21) could be cast as a special case of their previous work with expected utility. An important feature of that work, brought out further in Ben-Tal and Teboulle [2007], was duality with notions of information and entropy.[36]

In Krokhmal [2007] a much wider class of trade-off formulas for risk measures was studied with the aim of generalizing (3.21) through $\mathcal{V}$-type expressions not restricted to the expectation case. In that research, as in Ben-Tal and Teboulle [2007], no connections with statistical theory were contemplated. In other words, the bottom of the quadrangle was still out of sight.

The idea that risk issues in optimization modeling might interact in special ways with statistics started in Rockafellar et al. [2008]. It was demonstrated there that the choice of a risk measure in a constraint or objective might need to dictate the choice of the error measure employed in the regression set-up for approximating that constraint or objective.

It is convenient to speak of the quantifiers at the corners of the fundamental quadrangle, under the relations in Diagram 3, as constituting a *quadrangle quartet* $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$. In the regular case portrayed in the Quadrangle Theorem, it is a *regular* quadrangle quartet. The most attractive case adds monotonicity to $\mathcal{R}$ and $\mathcal{V}$ along with the corresponding properties of $\mathcal{D}$ and $\mathcal{E}$ in (3.16) and (3.17); we will then call the quartet *monotonic*. On the other hand, in the case where the four quantifiers are positively homogeneous we will speak of a quartet with positive homogeneity.

Although good examples of regular quadrangle quartets with and without monotonicity have been provided in Section 2, the question arises of how additional examples might be constructed. We round out this section with three results which can assist in that direction.

[35] In some papers in this area the random variables X were taken as representing uncertain "gains" instead of "losses." The resulting formulas are of course equivalent in that case, but minus signs have to be juggled in the translation.

[36] Here, see the end of Section 6.

The first one is elementary but puts into the proper perspective of an entire quadrangle the operation of blending risk with expectation that is seen in the formula it gives for $\mathcal{R}(X)$. Such blending, for instance with $\mathcal{R}_0(X) = \mathrm{CVaR}_\alpha(X)$, has gained some attention in finance.

Scaling Theorem. *Let $(\mathcal{R}_0, \mathcal{D}_0, \mathcal{V}_0, \mathcal{E}_0)$ be a regular quadrangle quartet with statistic $\mathcal{S}_0$ and consider any $\lambda \in (0, \infty)$. Then a regular quadrangle quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$ is given by*

$$\begin{aligned} &\mathcal{S}(X) = \mathcal{S}_0(X), \\ &\mathcal{R}(X) = (1-\lambda)EX + \lambda\mathcal{R}_0(X), \\ &\mathcal{D}(X) = \lambda\mathcal{D}_0(X), \\ &\mathcal{V}(X) = (1-\lambda)EX + \lambda\mathcal{V}_0(X), \\ &\mathcal{E}(X) = \lambda\mathcal{E}_0(X). \end{aligned} \tag{3.23}$$

Monotonicity and positive homogeneity are preserved in this construction.

Such scaling is present in Examples 1 and 1′ and could be added in Examples 2, 3 and 4.

Mixing Theorem. *For $k = 1, \ldots, r$ let $(\mathcal{R}_k, \mathcal{D}_k, \mathcal{V}_k, \mathcal{E}_k)$ be a regular quadrangle quartet with statistic $\mathcal{S}_k$, and consider any weights $\lambda_k > 0$ with $\lambda_1 + \cdots + \lambda_r = 1$. A regular quadrangle quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$ is given then by*

$$\begin{aligned} &\mathcal{S}(X) = \lambda_1\mathcal{S}_1(X) + \cdots + \lambda_r\mathcal{S}_r(X), \\ &\mathcal{R}(X) = \lambda_1\mathcal{R}_1(X) + \cdots + \lambda_r\mathcal{R}_r(X), \\ &\mathcal{D}(X) = \lambda_1\mathcal{D}_1(X) + \cdots + \lambda_r\mathcal{D}_r(X), \\ &\mathcal{V}(X) = \min_{B_1,\ldots,B_r}\Big\{ \sum\nolimits_{k=1}^r \lambda_k\mathcal{V}_k(X - B_k) \,\Big|\, \sum\nolimits_{k=1}^r \lambda_k B_k = 0 \Big\}, \\ &\mathcal{E}(X) = \min_{B_1,\ldots,B_r}\Big\{ \sum\nolimits_{k=1}^r \lambda_k\mathcal{E}_k(X - B_k) \,\Big|\, \sum\nolimits_{k=1}^r \lambda_k B_k = 0 \Big\}. \end{aligned} \tag{3.22}$$

Moreover $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ is monotonic if every $(\mathcal{R}_k, \mathcal{D}_k, \mathcal{V}_k, \mathcal{E}_k)$ is monotonic, and $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ is positively homogeneous if every $(\mathcal{R}_k, \mathcal{D}_k, \mathcal{V}_k, \mathcal{E}_k)$ is positively homogeneous.

This generalizes a result in Rockafellar et al. [2008] which dealt only with positively homogeneous quantifiers.[37] The quadrangle in Example 10 illustrates it for a particular case.

Reverting Theorem. *For $i = 1, 2$, let $(\mathcal{R}_i, \mathcal{D}_i, \mathcal{V}_i, \mathcal{E}_i)$ be a regular quadrangle quartet with statistic $\mathcal{S}_i$. Then a regular quadrangle quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$ is given by*

$$\begin{aligned} &\mathcal{S}(X) = \tfrac{1}{2}[\mathcal{S}_1(X) - \mathcal{S}_2(-X)], \\ &\mathcal{R}(X) = EX + \tfrac{1}{2}[\mathcal{R}_1(X) + \mathcal{R}_2(-X)], \\ &\mathcal{D}(X) = \tfrac{1}{2}[\mathcal{D}_1(X) + \mathcal{D}_2(-X)] = \tfrac{1}{2}[\mathcal{R}_1(X) + \mathcal{R}_2(-X)], \\ &\mathcal{V}(X) = EX + \min_B\Big\{ \tfrac{1}{2}[\mathcal{V}_1(B + X) + \mathcal{V}_2(B - X)] - B \Big\}, \\ &\mathcal{E}(X) = \min_B\Big\{ \tfrac{1}{2}[\mathcal{E}_1(B + X) + \mathcal{E}_2(B - X)] \Big\}. \end{aligned} \tag{3.23}$$

Positive homogeneity is preserved in this construction, but not monotonicity.

Example 11 illustrates a case where $(\mathcal{R}_1, \mathcal{D}_1, \mathcal{V}_1, \mathcal{E}_1)$ and $(\mathcal{R}_2, \mathcal{D}_2, \mathcal{V}_2, \mathcal{E}_2)$ coincide. The proof of the Reverting Theorem takes advantage of bounds $\mathcal{E}_i(X) \geq \psi_i(EX)$ produced from (3.12).[38]

[37] The proof is essentially the same as in that case, the main task being to demonstrate that $\mathcal{R}$ and $\mathcal{D}$ are closed and the minimum over $B_1, \ldots, B_r$ is attained. The argument follows the pattern we have indicated above for the projection part of the Quadrangle Theorem, making use of inequalities $\mathcal{E}_k(X) \geq \psi_k(EX)$ coming from (3.12).

[38] It starts with a direct calculation of the minimum of $\mathcal{E}(X - C)$ over C with the $\min_B$ expression for $\mathcal{E}$ inserted. A change of variables $C_1 = C - B$, $C_2 = -C - B$, shows that this yields the claimed $\mathcal{S}$, and $\mathcal{D}$. The corresponding $\mathcal{R}$ and $\mathcal{V}$ are confirmed then from the quadrangle formulas.

4 Further Interpretations and Results

More can be said now about about regret versus utility and how this can affect the right side of the quadrangle. In relying on (3.13) for a one-to-one correspondance between regular measures of regret $\mathcal{V}$ and regular measures of relative utility $\mathcal{U}$, we are in particular replacing the convexity of $\mathcal{V}$ with the concavity of $\mathcal{U}$ and requiring, for a random variable Y oriented toward gain, that

$$\mathcal{U}(0) = 0 \text{ but } \mathcal{U}(Y) < EY \text{ when } Y \not\equiv 0. \tag{4.1}$$

This is where the term "relative" comes in. The gain in Y needs to be viewed as *gain relative to some benchmark*. That contrasts with the way utility theory is ordinarily articulated in terms of the "absolute" utility of an outcome. But practitioners appreciate nowadays that investors, for instance, are highly influenced by benchmarks in their attitudes toward gain or loss.

The case of expected utility, focused on $E[u(Y)]$ for a one-dimensional utility function u giving $u(y)$ for a sure gain y, serves well in explaining this. A large body of traditional theory in finance, laid out authoritatively in Föllmer and Schied [2004], looks toward maximizing such an expression under various side conditions in putting together a good portfolio. The utility function u captures the preferences of an investor, and the expectation deals with the uncertainty when the gain y turns into a random variable Y. Standard functions u have logarithmic forms and the like, and there is often nothing "relative" about them.

In order to have a functional $\mathcal{U}(Y) = E[u(Y)]$ satisfy (4.1) and be closed concave,[39] the natural specialization is to require u to be a function of y with

$$u \text{ closed concave and } u(0) = 0 \text{ but } u(y) < y \text{ when } y \neq 0. \tag{4.2}$$

Again, the sense in that would come from a benchmark interpretation, namely that y no longer stands for an amount of money received in the future but rather an increment (positive or negative) to some reference amount. A utility function satisfying (4.2), but with "$<$" weakened to "$\leq$," is a *normalized* utility in the terminology of Ben-Tal and Teboulle [2007]. Normalization to create these properties is always possible in the expectation case because, in theory, as far as generating a preference ordering for y values is concerned, a utility u is only determined up to translations and an arbitrary scaling factor.[40] For our quadrangle scheme, however, such normalization is not merely a convenience but essential. Expected utility depends not only on the ordering induced by u on $(-\infty, \infty)$, but also on the "curvature" aspects of u, and the choice of a benchmark can have a large impact on that, apart from some special cases.

A utility function u satisfying (4.2) is paired with a regret function v satisfying

$$v \text{ closed convex and } v(0) = 0 \text{ but } v(x) > x \text{ when } x \neq 0. \tag{4.3}$$

under the correspondence[41]

$$v(x) = -u(-x), \qquad u(y) = -v(-y). \tag{4.4}$$

The properties in (4.3) are needed for $\mathcal{V}(X) = E[v(X)]$ to be a *regular* measure of regret. They are crucial moreover in the correspondence between $\mathcal{V}$ and $\mathcal{E}$ at the bottom of the quadrangle in making

[39]Closed concavity requires the "upper" level sets of type $\geq c$ to be closed for all $c \in I\!R$, in contrast to closed convexity, which requires all "lower" level sets of type $\leq c$ to be closed.

[40]Outside of the expectation case, it is still possible to shift to $\mathcal{U}(0) = 0$ as a "normalization," but rescaling is insufficient to proceed to $\mathcal{U}(Y) \leq EY$.

[41]In this correspondence the graphs of v and u reflect to each other through the origin of $I\!R^2$.

$\mathcal{E}(X) = E[e(X)]$ be a regular measure of error paired with $\mathcal{V}(X) = E[v(X)]$ under the relations

$$e(x) = v(x) - x, \qquad v(x) = x + e(x), \tag{4.5}$$

which entail having

$$e \text{ closed convex and } e(0) = 0 \text{ but } e(x) > 0 \text{ when } x \neq 0. \tag{4.6}$$

The condition on the utility function u in (4.2) implies that $u'(0) = 1$ when u is differentiable at 0, but it is important to realize that u might not be differentiable at 0, and this could even be desirable. From concavity, u is sure at least to have right derivatives $u'_+(y)$ and left derivatives $u'_-(y)$ satisfying $u'_+(y) \leq u'_-(y)$, usually with equality, but still maybe with $u'_-(0) > u'_+(0)$. This would mean that, in terms of relative utility, *the pain of a marginal loss relative to the benchmark is greater than pleasure of a marginal gain relative to the benchmark*. Just such a disparity in reactions to gains and losses is seen in practice and reflects, at least in part, the observations in Kahneman and Tversky [1979].

In translating this from a concave utility function u to a convex regret function v as in (4.3), we have, of course, right derivatives $v'_+(x)$ and $v'_-(x)$ satisfying $v'_-(x) \leq v'_+(x)$, usually with equality, but perhaps with $v'_-(0) < v'_+(0)$. However, something more needs to be understood in connection with the ability of v to take on ∞ and how that affects the way derivatives are treated in the formulas of the theorem below.

The convexity of v implies that the effective domain $\operatorname{dom} v = \{\, x \mid v(x) < \infty \,\}$ is an interval in $(-\infty, \infty)$ (not necessarily closed or bounded). If x is the right endpoint of $\operatorname{dom} v$, the definition of the right derivative naturally gives $v'_+(x) = \infty$; but just in case of doubt in some formula, this is also the interpretation to give of $v'_+(x)$ when x is off to the right of $\operatorname{dom} v$.[42] Likewise, if x is the left endpoint of $\operatorname{dom} v$, or further to the left, then $v'_-(x) = -\infty$.

These are the patterns also for an error function e as in (4.6).

For the fundamental quadrangle of risk, the consequences of these facts in the expectation case are summarized as follows.

Expectation Theorem. *For functions v and e on $(-\infty, \infty)$ related by (4.5), the properties in (4.3) amount to those in (4.6) and ensure that the functionals*

$$\mathcal{V}(X) = E[v(X)], \qquad \mathcal{E}(X) = E[e(X)], \tag{4.7}$$

form a corresponding pair consisting of a regular measure of regret and a regular measure of error.[43] *For $X \in D = \operatorname{dom} \mathcal{V} = \operatorname{dom} \mathcal{E}$ let $C^+(X) = \sup \{\, C \mid X - C \in D \,\}$ and $C^-(X) = \inf \{\, C \mid X - C \in D \,\}$. The associated statistic $\mathcal{S}$ in the quadrangle generated from $\mathcal{V}$ and $\mathcal{E}$ is characterized then by*

$$\mathcal{S}(X) = \Big\{\, C \,\Big|\, E[e'_-(X - C)] \leq 0 \leq E[e'_+(X - C)] \,\Big\} = \Big\{\, C \,\Big|\, E[v'_-(X - C)] \leq 1 \leq E[v'_+(X - C)] \,\Big\} \tag{4.8}$$

subject to the modification that, in both cases, the right side is replaced by ∞ if $C \leq C^-(X)$ and the left side is replaced by $-\infty$ if $C \geq C^+(X)$. The quadrangle is completed then by setting

$$\mathcal{D}(X) = E[e(X - C)] \ \text{ and } \ \mathcal{R}(X) = C + E[v(X - C)] \ \text{ for any/all } C \in \mathcal{S}(X). \tag{4.9}$$

[42] The issue is that a random X might produce such an outcome with probability 0, and yet one still needs to know how to think of the formula.

[43] Also, $\mathcal{V}$ corresponds then to a regular measure of relative utility $\mathcal{U}$ given by $\mathcal{U}(Y) = E[u(Y)]$ under (4.4) via (4.2).

Having $\mathcal{V}$ and $\mathcal{R}$ be monotonic corresponds (in tandem with convexity) to having $v(x) \leq 0$ when $x < 0$, or equivalently $e(x) \leq |x|$ when $x < 0$. Positive homogeneity holds in the quadrangle if and only if v and e have graphs composed of two linear pieces kinked at 0.

Beyond the aspects of this theorem that are already evident, the key ingredient is establishing (4.8). This is carried out by calculating that the right and left derivatives of the convex function $\phi(C) = E[e(X - C)]$ from their definitions and noting that C belongs to $\operatorname{argmin} \phi$ if and only if $\phi'_-(C) \leq 0 \leq \phi'_+(C)$. In situations where v and e are differentiable, the double inequalities in (4.8) can be replaced simply by the equations $E[e'(X - C)] = 0$ and $E[v'(X - C)] = 1$.

We proceed now to illustrate the Expectation Theorem by applying it to justify the details of the examples in Section 2 that belong to the expectation case.

Quantile-based quadrangle, Example 2 (including Example 3):

$$e(x) = \frac{\alpha}{1-\alpha}\max\{0,x\} + \max\{0,-x\}, \quad v(x) = \frac{1}{1-\alpha}\max\{0,x\}, \quad u(y) = \frac{1}{1-\alpha}\min\{0,y\}.$$

We have $D = \mathcal{L}^2(\Omega)$, $C^+(X) = \infty$, $C^-(X) = -\infty$, and

$$v'_+(x) = \begin{cases} \frac{1}{1-\alpha} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases} \qquad v'_-(x) = \begin{cases} \frac{1}{1-\alpha} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

with a gap between left and right derivatives occuring only at $x = 0$. Then with $F_X^-(C)$ denoting the left limit of F_X at C (the right limit $F_X^+(C)$ being just $F_X(C)$), we get

$$\begin{aligned} E[v'_-(X - C)] &= \tfrac{1}{1-\alpha}\text{prob}\,\{X > C\} = 1 - F_X(C), \\ E[v'_+(X - C)] &= \tfrac{1}{1-\alpha}\text{prob}\,\{X \geq C\} = 1 - F_X^-(C). \end{aligned}$$

It follows thereby from (4.8) that $\mathcal{S}(X) = \{\, C \,|\, F_X^-(C) \leq \alpha \leq F_X(C) \,\}$ and therefore $\mathcal{S}(X) = q_\alpha(X)$. Applying (4.10) yields $\mathcal{R}(X) = C + \frac{1}{1-\alpha}E\max\{0, X - C\} = C + \frac{1}{1-\alpha}\int_{(C,\infty)}(x - C)dF_X(x)$. Since the probability of (C, ∞) is $1 - F_X(C)$, this equals $\frac{1}{1-\alpha}[(F_X(C) - \alpha)C + \int_{[C,\infty)} x dF_X(x)$, which is the expectation of X with respect to its "α-tail distribution" as defined in Rockafellar and Uryasev [2002] and used there to properly define $\overline{q}_\alpha(X)$ even under the possibility that $F_X(C) > \alpha$.

Worst-case-based quadrangle, Example 5:

$$e(x) = \begin{cases} |x| & \text{if } x \leq 0 \\ \infty & \text{if } x > 0, \end{cases} \quad v(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \infty & \text{if } x > 0, \end{cases} \quad u(y) = \begin{cases} -\infty & \text{if } y < 0 \\ 0 & \text{if } y \geq 0. \end{cases}$$

We have $D = \mathcal{L}^2_-(\Omega)$, $C^+(X) = \infty$, $C^-(X) = \sup X$. In the v part of (4.8) the left side equals 0 always and the right side equals 0 if $C < \sup X$ but (through the prescribed modification) equals ∞ if $C = \sup X$. Therefore, $C = \sup X$ is the unique element of $\mathcal{S}(X)$ (when that is finite).

The Truncated-Mean-Based Quadrangle, Example 7:

$$e(x) = \begin{cases} |x| - \frac{\beta}{2} & \text{if } |x| \geq \beta, \\ \frac{1}{2\beta}x^2 & \text{if } |x| \leq \beta, \end{cases} \quad v(x) = \begin{cases} -\frac{\beta}{2} & \text{if } x \leq -\beta, \\ x + \frac{1}{2\beta}x^2 & \text{if } |x| \leq \beta, \\ 2x - \frac{\beta}{2} & \text{if } x \geq \beta, \end{cases} \quad u(y) = \begin{cases} 2y + \frac{\beta}{2} & \text{if } y \leq -\beta, \\ y - \frac{1}{2\beta}y^2 & \text{if } |y| \leq \beta, \\ \frac{\beta}{2} & \text{if } y \geq \beta. \end{cases}$$

This time, $D = \mathcal{L}^2(X)$, so $C^+(X) = \infty$ and $C^-(X) = -\infty$. The statistic is determined by solving $E[e'(X - C)] = 0$ for C, and since

$$\beta e'(x) = T_\beta(x) = \begin{cases} \beta & \text{if } x \geq \beta, \\ x & \text{if } -\beta \leq x \leq \beta, \\ -\beta & \text{if } x \leq -\beta, \end{cases}$$

this gives the result described.

Log-exponential-based quadrangle, Example 8:

$$e(x) = \exp x - x - 1, \quad v(x) = \exp x - 1, \quad u(y) = 1 - \exp(-y).$$

Here $D = \{ X \mid E[\exp X] < \infty \}$. Because $E[\exp(X - C)] = \exp(-C)E[\exp X]$, we have $C^+(X) = \infty$ and $C^-(X) = -\infty$ for any $X \in D$, so the need for a modification of the bounds in (4.8) is avoided. Indeed, since $v'(x) = \exp x$, we just have an equation to solve for C, namely $E[\exp(X - C)] = 1$. This equation can be rewritten as $E[\exp X] = \exp C$, which yields $C = \log E[\exp X]$ as $\mathcal{S}(X)$. Substituting that into $C + \mathcal{V}(X - C)$, we get $\mathcal{R}(X) = \log E[\exp X]$ and the quadrangle is confirmed.

The Rate-Based Quadrangle, Example 9:

$$e(x) = \begin{cases} \log \frac{1}{1-x} - x & \text{if } x < 1, \\ \infty & \text{if } x \geq 1, \end{cases} \quad v(x) = \begin{cases} \log \frac{1}{1-x} & \text{if } x < 1, \\ \infty & \text{if } x \geq 1, \end{cases} \quad u(y) = \begin{cases} \log(1+y) & \text{if } y > -1, \\ -\infty & \text{if } y \leq -1. \end{cases}$$

Here $D = \{ X < 1 \mid E[\log \frac{1}{1-X}] < \infty \}$, so $C^+(X) = 1 - \sup X$ and $C^-(X) = -\infty$. Because v is differentiable (where finite), we have an equation to solve in (4.8): $E[\frac{1}{1-(X-C)}] = 1$. The solution is the statistic $\mathcal{S}(X)$.

Quadrangles from kinked utility and regret. More examples beyond the differentiable case of the Expectation Theorem can be produced by starting from an "absolute" utility function $u_0(y_0)$ that is differentiable, increasing and strictly concave, introducing a benchmark value B, and a "kink" parameter $\delta > 0$, and defining

$$u(y) = \frac{u_0(y + B) - u_0(B)}{u_0'(B)} + \delta \min\{0, y\}. \tag{4.10}$$

This will satisfy $u(0) = 0$ and $u(y) < y$ when $y \neq 0$, and it will be differentiable when $y \neq 0$, but have

$$u_+'(0) = 1 \text{ but } u_-'(0) = 1 + \delta. \tag{4.11}$$

The kink parameter δ models the extra pain experienced in falling short of the benchmark, in contrast to the milder pleasure experienced in exceeding it. From this u it is straightforward to pass to the corresponding v, e, and the full quadrangle associated with them by the theorem. In general, that quadrangle will depend on both B and δ, but in special situations like CARA or HARA utilities[44] the B dependence can drop out or reduce to simple rescaling.

The surprising fact is that all such manipulations are propagated by the quadrangle scheme into applications not just to risk management and optimization but also to statistical estimation. Those applications will be discussed further in Section 5.

General interpretations of the quadrangle "statistic." Returning finally to the general level of the correspondence $\mathcal{U} \leftrightarrow \mathcal{V}$ between relative utility and regret in (3.13) we look at ways of interpreting the trade-off formula $\mathcal{R}(X) = \min_C \{ C + \mathcal{V}(X - C) \}$. Through a change of variables $Y = -X$, $D = -C$, switching loss to gain, this corresponds to

$$-\mathcal{R}(-Y) = \max_D \{ D + \mathcal{U}(Y - D) \}. \tag{4.12}$$

Considerations were focused in Ben-Tal and Teboulle [2007] on the expectation case, but an interpretation suggested there works well for (4.14) in general. To begin with, note that in adding D to

[44]See [Föllmer and Schied, 2004, pages 68–69].

$\mathcal{U}(Y-D)$ it is essential that D be measured in the same units as $\mathcal{U}(Y-D)$, and moreover they have to be the same units as those of Y. A simple case where this makes perfect sense is the one in which the units are money units, like dollars. Then D represents an income that is certain, whereas $Y-D$ is residual income that is uncertain; $\mathcal{U}$ assigns to that uncertain income a equally desirable amount of certain income in something akin to a discount. This leads, in Ben-Tal and Teboulle [2007], a D giving the max in (4.12) being called an *optimized certainty equivalent* for Y.

Much the same can be said about the regret version of trade-off, $\mathcal{R}(X) = \min_C \{ C + \mathcal{V}(X-C) \}$. There, C is a loss that is certain, $X-C$ is a residual loss that is uncertain. The regret measure $\mathcal{V}$ assigns to $X-C$ an amount of money that could be deemed adequate as immediate compensation for taking on the burden of $X-C$. It is possible to elaborate this with ideas of insurance, insurance premium, "deductibles," and so forth. For some insurance interpretations in the utility context of (4.12), see Ben-Tal and Teboulle [2007].

Although these "min" formulas and interpretations and natural in their own right, the special insight from the risk quadrangle, namely that they have a parallel life in theoretical statistics, is new.

5 Roles in Optimization and Generalized Regression

Risk in the sense quantified by a risk measure $\mathcal{R}$ is central in the management and control of cost or loss. For a hazard variable X, the crucial issue is how to model a "soft" upper bound, i.e., a condition that the outcomes of X be "more or less" $\leq C$ for some threshold level C. As already explained in Section 1, the broad prescription for handling this is to pass to a numerical inequality $\mathcal{R}(X) \leq C$ through some choice of a risk measure $\mathcal{R}$, and many possibilities for $\mathcal{R}$ have been offered. Of course C can be taken to be 0 without any real loss of generality.

A choice of $\mathcal{R}$ corresponds to an expression of preferences toward risk, but it might not yet be clear why some measures of risk are better motivated or computationally more tractable than others. The key challenge is that most applications require more than just looking at $\mathcal{R}(X)$ for a single X, as far as optimization is concerned. Usually instead, there is a random variable that depends on parameters $x_1, \ldots, x_n$. We have $X(x_1, \ldots, x_n)$ and it becomes important to know how the numerical surrogate $\mathcal{R}(X(x_1, \ldots, x_n))$ depends on $x_1, \ldots, x_n$. This is where favorable conditions imposed on $\mathcal{R}$, like convexity and monotonicity, are indispensable.

Motivations in optimization modeling are important in particular. For insight, consider first a standard type of *deterministic* optimization problem, without any uncertainty, in which $x = (x_1, \ldots, x_n)$ is the decision vector, namely

$$(\mathcal{P}) \qquad \text{minimize } f_0(x) \text{ over all } x \in S \subset I\!R^n \text{ subject to } f_i(x) \leq 0 \text{ for } i = 1, \ldots, m.$$

A decision x selected from the set S results in *numerical values* $f_0(x)$, $f_1(x), \ldots, f_m(x)$, which can be subjected to the usual techniques of optimization methodology. Suppose next, though, that these cost-like expressions are uncertain through dependence on additional variables — random variables — whose realizations will not be known until later. A decision x merely results then in *random variables*[45]

$$X_0(x) = \underline{f}_0(x),\ X_1(x) = \underline{f}_1(x), \ldots,\ X_m(x) = \underline{f}_m(x), \tag{5.1}$$

which can only be *shaped in their distributions* through the choice of x, not pinned down to specific values. Now there is no longer a single, evident answer to how optimization should be viewed, but risk measures can come to the rescue.

[45]We employ underbars in this discussion to indicate uncertainty. The overbars appearing later emphasize that the random variable depending on x has been converted to a nonrandom numerical function of x.

As proposed in Rockafellar [2007], one can systematically pass to a *stochastic* optimization problem in the format[46]

$$(\underline{\mathcal{P}}) \qquad \text{minimize } \overline{f}_0(x) = \mathcal{R}_0(\underline{f}_0(x)) \text{ over } x \in S \text{ subject to } \overline{f}_i(x) = \mathcal{R}_i(\underline{f}_i(x)) \leq 0,\ i = 1, \ldots, m,$$

in which an individually selected "measure of risk" $\mathcal{R}_i$ has been combined with each $\underline{f}_i(x)$ to arrive at a numerical (nonrandom) function $\overline{f}_i$ of the decision vector x.[47]

An issue that must then be addressed is how the properties of $\overline{f}_i(x)$ with respect to x relate to those of $\underline{f}_i(x)$ through the choice of $\mathcal{R}_i$, and whether those properties are conducive to good optimization modeling and solvability. This is not to be taken for granted, because seemingly attractive examples like $\mathcal{R}_i(X) = EX + \lambda_i \sigma(X)$ with $\lambda_i > 0$ and $\mathcal{R}_i(X) = q_{\alpha_i}(X) = \mathrm{VaR}_{\alpha_i}(X)$ with $0 < \alpha_i < 1$ are known to suffer from troubles with "coherency" in the sense of Artzner et al. [1999].

Convexity-Preservation Theorem[48]. *In problem $(\underline{\mathcal{P}})$, the convexity of $\overline{f}_i(x)$ with respect to x is assured if $\underline{f}_i(x)$ is linear in x and $\mathcal{R}_i$ is a regular measure of risk, or if $\underline{f}_i(x)$ is convex in x and $\mathcal{R}_i$ is, in addition, a monotonic measure of risk.*

The huge advantage of having the functions $\overline{f}_i$ be convex is that then, with the set S also convex, $(\underline{\mathcal{P}})$ is an *optimization problem of convex type*. Such problems are vastly easier to solve in computation.

The use of $\mathcal{R}_i(X) = q_{\alpha_i}(X) = \mathrm{VaR}_{\alpha_i}(X)$ in this setting could destroy whatever underlying convexity with respect to $x = (x_1, \ldots, x_n)$ might be available in the problem data, because this measure of risk lacks convexity; it is not regular and not coherent. The shortcoming of $\mathcal{R}_i(X) = EX + \lambda_i \sigma(X)$ is different: it fails in general to be monotonic. The absence of monotonicity threatens the transmittal of convexity of $\underline{f}_i(x)$ to $\overline{f}_i(x)$. However, $\overline{f}_i(x)$ can still be convex in x, on the basis of the Convexity-Preservation Theorem, as long as $\underline{f}_i(x)$ is linear in x. This could be useful in applications to financial optimization, because linearity with respect to x, as a vector of "portfolio weights," is often encountered there.

Another example of a measure of risk that is regular without being monotonic is the reverted CVaR in Example 11: $\mathcal{R}_i(X) = EX + \frac{1}{2}[\mathrm{CVaR}_{\alpha_i}(X) + \mathrm{CVaR}_{\alpha_i}(-X)]$. Once more, although this choice would not preserve convexity in general, it would do so when $\underline{f}_i(x)$ is linear in x.

A question of modeling motivation must be confronted here. Why would one ever wish to use in a stochastic optimization problem $(\underline{\mathcal{P}})$ a regular risk measure that is not monotonic, even in applications with linearity in x, when so many choices do have that property? An interesting justification can actually be given, which could sometimes make sense in finance, at least. The rationale has to do with skepticism about the data in the model and especially a wish to not rely too much on data in the extreme lower tail of a cost distribution. Optimization with today's data will be succeeded by optimization with tomorrow's data, all data being imperfect. It would be wrong to swing very far in response to ephemeral changes, at least in connection with formulating the objective function $\overline{f}_0(x) = \mathcal{R}_0(\underline{f}_0(x))$.

---

[46]If taken too literally, this prescription could be simplistic. When uncertainty is present, much closer attention must be paid to whether the objective and constraint structure in the deterministic formulation itself was well chosen. The effects of possible recourse actions when constraints are violated may need to be brought in. Whether risk measures should be applied to the f_i's individually or to a combination passed through some joint expression must be considered as well.

[47]The constraint modeling in $(\underline{\mathcal{P}})$ follows the prescription that $\mathcal{R}_i(\underline{f}_i(x)) \leq 0$ provides a rigorous interpretation to the desire of having $\underline{f}_i(x)$ "more or less" ≤ 0, but the motivation for the treatment of the objective in $(\underline{\mathcal{P}})$ may be less clear. Actually, it follows the same prescription. Choosing x to minimize $\mathcal{R}_0(\underline{f}_0(x))$ can be identified with choosing a pair (x, C_0) subject to $\mathcal{R}_0(\underline{f}_0(x)) \leq C_0$ so as to get C_0 as low as possible, and the inequality $\mathcal{R}_0(\underline{f}_0(x)) \leq C_0$ models having $\underline{f}_0(x)$ "more or less" $\leq C_0$. This is valuable in handling the dangers of "cost overruns."

[48]This extends, in an elementary way, a principle in Rockafellar [2007].

The following idea comes up: replace this objective, in the case of a regular monotonic measure of risk $\mathcal{R}_0$, by a measure of risk having the form

$$\tilde{\mathcal{R}}_0(X) = \mathcal{R}_0(X) + \mathcal{D}(X) \text{ for some regular measure of deviation } \mathcal{D}. \tag{5.2}$$

This would be another regular measure of risk, even if not monotonic. The deviation term would be designed to have a "stabilizing" effect.

If a choice like $\mathcal{R}_i(X) = q_{\alpha_i}(X) = \mathrm{VaR}_{\alpha_i}(X)$ ought to be shunned when convexity in $(\underline{\mathcal{P}})$ is to be promoted, what might be the alternative? This is a serious issue because risk constraints involving this choice are very common, especially in reliability engineering,[49] because

$$q_{\alpha_i}(\underline{f}_i(x)) \leq 0 \iff \mathrm{prob}\{\, \underline{f}_i(x) \leq 0 \,\} \geq \alpha_i. \tag{5.3}$$

A strong argument can be made for passing from quantiles/VaR to superquantiles/CVaR by instead taking $\mathcal{R}_i(X) = \overline{q}_{\alpha_i}(X) = \mathrm{CVaR}_{\alpha_i}(X)$. This has the effect of replacing "probability of failure" by an alternative called "buffered probability of failure," which is safer and easier to work with computationally; see Rockafellar and Royset [2010] and Basova, Rockafellar and Royset [2011].

The claim that problem-solving may be easier with CVaR than with VaR could seem surprising from the angle that $\mathrm{CVaR}_\alpha(X)$ is defined as a conditional expectation in a "tail" which is dependent on $\mathrm{VaR}_\alpha(X)$, yet it rests on the characterization in (3.21). But we have explained in Rockafellar and Uryasev [2002][50] how, in the case of $(\underline{\mathcal{P}})$ with $\mathcal{R}_i = \mathrm{CVaR}_{\alpha_i}$ for each i, one can expand $\mathrm{CVaR}_{\alpha_i}(\underline{f}_i(x))$ through (3.21) into an expression involving a auxiliary parameter C_i and proceed to minimize not only with respect to x but also simultaneously with respect to the C_i's. This has the benefit not only of simplifying the overall minimization but also providing, along with the optimal solution $\bar{x}$ to $(\underline{\mathcal{P}})$, corresponding $\mathrm{VaR}_{\alpha_i}(\underline{f}_i(x))$ values as the optimal $\bar{C}_i$'s.

Now we are in position to point out, on the basis of the risk quadrangle, that this technique has a far-reaching extension.

Regret Theorem. *Consider a stochastic optimization problem $(\underline{\mathcal{P}})$ in which each $\mathcal{R}_i$ is a regular measure of risk coming from a regular measure of regret $\mathcal{V}_i$ with associated statistic $\mathcal{S}_i$ by the quadrangle formulas*

$$\mathcal{R}_i(X) = \min_C \{\, C + \mathcal{V}_i(X - C) \,\}, \qquad \mathcal{S}_i(X) = \operatorname*{argmin}_C \{\, C + \mathcal{V}_i(X - C) \,\}. \tag{5.4}$$

Solving $(\underline{\mathcal{P}})$ can be cast then as solving the expanded problem

$$(\underline{\mathcal{P}}') \qquad \begin{array}{l} \textit{choose } x = (x_1, \ldots, x_n) \textit{ and } C_0, C_1, \ldots, C_m \textit{ to} \\ \textit{minimize } C_0 + \mathcal{V}_0(\underline{f}_0(x) - C_0) \textit{ over } x \in S,\ C_i \in I\!R, \\ \textit{subject to } C_i + \mathcal{V}_i(\underline{f}_i(x) - C_i) \leq 0 \textit{ for } i = 1, \ldots, m. \end{array}$$

An optimal solution $(\bar{x}, \bar{C}_0, \bar{C}_1, \ldots, \bar{C}_m)$ to problem $(\underline{\mathcal{P}}')$ provides as $\bar{x}$ an optimal solution to problem $(\underline{\mathcal{P}})$ and as $\bar{C}_i$ a corresponding value of the statistic $\mathcal{S}_i(\underline{f}_i(\bar{x}))$ for $i = 0, 1, \ldots, m$.

The Mixing Theorem of Section 3 can be combined with Regret Theorem. When $\mathcal{V}_i$ is already itself expressed by a minimization formula involving additional parameters, these can be brought into $(\underline{\mathcal{P}})$ as well.

The topic of generalized regression is next on the agenda. As explained in Section 1, this concerns the approximation of a given random variable Y by a function $f(X_1, \ldots, X_n)$ of other random variables

[49] The article Samson et al. [2009] furnishes illuminating background.

[50] See also the tutorial paper Rockafellar [2007].

$X_1, \dots, X_n$. By the regression being "generalized" we mean that the difference $Z_f = Y - f(X_1, \dots, X_n)$ may be assessed for its nonzeroness by an error measure $\mathcal{E}$ different from the one in "least-squares" as in Example 1, or for that matter even from the kind in quantile regression, as in Example 2. The case of generalized *linear* regression, where the functions f in the approximation are limited to the form

$$f(x_1, \dots, x_n) = C_0 + C_1 x_1 + \cdots + C_n x_n \quad \text{(the linear case)}, \tag{5.5}$$

has already been studied in Rockafellar et al. [2008], but only for error measures $\mathcal{E}$ that are positively homogeneous. Here we go beyond those limitations and investigate the problem:

$$\text{minimize } \mathcal{E}(Z_f) \text{ over all } f \in \mathcal{C}, \text{ where } Z_f = Y - f(X_1, \dots, X_n), \tag{5.6}$$

for given random variables $X_1, \dots, X_n, Y$, and some given class $\mathcal{C}$ of functions f.

Taking $\mathcal{C}$ to be the class in (5.5) with respect to all possible coefficients $C_0, C_1, \dots, C_n$, would specialize to linear regression, pure and simple. Then $\mathcal{E}(Z_f)$ would be a function of these coefficients and we would be minimizing over $(C_0, C_1, \dots, C_n) \in I\!R^{n+1}$. However, even in the linear case there could be further specialization through placing conditions on some of the coefficients, such as perhaps nonnegativity. In fact, a broad example if the kinds of classes regression functions that can be brought into the picture is the following:[51]

$$\begin{array}{l} \mathcal{C} = \text{ all the functions } f(x_1, \dots, x_n) = C_0 + C_1 h_1(x_1, \dots, x_n) + \cdots + C_m h_m(x_1, \dots, x_n) \\ \text{for given } h_1, \dots, h_m \text{ on } I\!R^n \text{ and coefficient vectors } (C_1, \dots, C_m) \text{ in a given set } C \subset I\!R^m. \end{array} \tag{5.7}$$

Motivation for generalized regression comes from applications in which Y has the cost/loss orientation that we have been emphasizing in this paper. Underestimation might then be more dangerous than overestimation, and that could suggest using an asymmetric error measure $\mathcal{E}$, with $\mathcal{E}(Z_f) \neq \mathcal{E}(-Z_f)$.

Further motivation comes from "factor models" and other such regression techniques in finance and engineering, which might have unexpected consequences when utilized in stochastic optimization because of interactions with parameterization by the decision vector x. For instance, if one of the random "costs" $\underline{f}_i(x)$ in problem $(\underline{\mathcal{P}})$ is estimated by such a technique as $\underline{g}_i(x)$, it may be hard to determine the effects this could have on the optimal decision. We have argued in Rockafellar et al. [2008], and demonstrated with specific results, that it might be wise to "tune" the regression to the risk measure $\mathcal{R}_i$ applied to $\underline{f}_i(x)$ in $(\underline{\mathcal{P}})$. This would mean passing around the fundamental quadrangle from $\mathcal{R}_i$ to an error measure $\mathcal{E}_i$ in the same quartet.

Regression Theorem. *Consider problem (5.6) for random variables $X_1, \dots, X_n$ and Y in the case of $\mathcal{E}$ being a regular measure of error and $\mathcal{C}$ being a class of functions $f : I\!R^n \to I\!R$ such that*

$$f \in \mathcal{C} \implies f + C \in \mathcal{C} \text{ for all } C \in I\!R. \tag{5.8}$$

Let $\mathcal{D}$ and $\mathcal{S}$ correspond to $\mathcal{E}$ as in the Quadrangle Theorem. Problem (5.6) is equivalent then to the following:

$$\text{minimize } \mathcal{D}(Z_f) \text{ over all } f \in \mathcal{C} \text{ such that } 0 \in \mathcal{S}(Z_f). \tag{5.9}$$

Moreover if $\mathcal{E}$ is of expectation type and $\mathcal{C}$ includes a function f satisfying

$$\begin{array}{c} f(x_1, \dots, x_n) \in \mathcal{S}(Y(x_1, \dots, x_n)) \text{ almost surely for } (x_1, \dots, x_n) \in D, \\ \text{where } Y(x_1, \dots, x_n) = Y_{X_1 = x_1, \dots, X_n = x_n} \text{ (conditional distribution)}, \end{array} \tag{5.10}$$

[51] It should also be kept in mind that a possibly nonlinear change of scale in the variables, such as passing to logarithms, could be executed prior to this depiction.

with D being the support of the distribution in $\mathbb{R}^n$ induced by $X_1, \ldots, X_n$, then that f solves the regression problem and tracks this conditional statistic in the sense that

$$f(X_1, \ldots, X_n) = \mathcal{S}(Y \mid X_1, \ldots, X_n) \text{ almost surely.} \tag{5.11}$$

The first part of this result generalizes [Rockafellar et al., 2008, Theorem 3.2] on linear regression through elementary extension of the same proof. The second part is new. It comes from the observation that, in the expectation case, if f satisfies (5.10), then for any other $g \in \mathcal{C}$ one has

$$\mathcal{E}(Y - f(x_1, \ldots, x_n)) \leq \mathcal{E}(Y - g(x_1, \ldots, x_n)) \text{ almost surely for } (x_1, \ldots, x_n) \in D.$$

When $\mathcal{E}$ is of expectation type, this inequality can be "integrated" over the distribution of $(X_1, \ldots, X_n)$ to obtain $\mathcal{E}(Y - f(X_1, \ldots, X_n)) \leq \mathcal{E}(Y - g(X_1, \ldots, X_n))$.

Apart from that special circumstance, the question of the existence of an optimal regression function $f \in \mathcal{C}$ has not been addressed in the theorem, because we are reluctant here to delve deeply into the possible structure of the class $\mathcal{C}$. But existence in the case of linear regression has been covered in [Rockafellar et al., 2008, Theorem 3.1], and similar considerations would apply to the broader class in (5.7), with the coefficient set C taken to be closed.[52]

There could be many applications of these ideas, and much remains to be explored and developed. Some related research in special cases, largely concerned with quantile regression, can be seen in Trindade and Uryasev [2005], Trindade and Uryasev [2006] and Golodnikov et al. [2007]; see also Samson et al. [2009] for further motivation.

The measure of error in quantile regression is indeed of expectation type, so that the second part of our Regression Theorem can be applied if the class $\mathcal{C}$ of functions f is rich enough. The class of linear functions of $X_1, \ldots, X_n$ would very likely not meet that standard, but the class in (5.7) may offer hope through judicious choice of $h_1, \ldots, h_m$.

It is interesting to note that regression in the framework developed here includes one-sided approximations not usually associated with that term. For example, for the error measure of Example 5 the problem is to

$$\text{minimize } E|f(X_1, \ldots, X_n) - Y| \text{ over all } f \in \mathcal{C} \text{ satisfying } f(X_1, \ldots, X_n) \geq Y. \tag{5.12}$$

6 Probability Modeling and the Dualization of Risk

More explanation about the view of uncertainty that we take here may be helpful, especially for the sake of those who would like to make use of the ideas without having to go too far into the technical mathematics of probability theory. In modeling uncertain quantities as random variables, we tacitly regard them as having probability distributions, but this does not mean we assume those distributions are directly known. Sampling, for instance, might be required to learn more, and even then, only approximations might be available.

The characteristics of a random variable X, by itself, are embodied in its cumulative distribution function F_X, with $F_X(x) = \text{prob}\{X \leq x\}$. This induces a probability measure on the real numbers $\mathbb{R}$ which may or may not be expressible by a density function f with respect to ordinary integration, i.e., as $dF_X(x) = f(x)dx$. The lack of a density function is paramount when X is a discrete random variable with only finitely many possible outcomes. Then F_X is a step function.

[52] Work with the class in (5.7), which does of course satisfy (5.9), can actually be reduced to the linear case, so that generalized linear theory can be applied. To do this we can introduce new random variables $W_i = h_i(X_1, \ldots, X_n)$ with distributions inherited from the X_j's and carry out linear regression of Y with respect to $W_1, \ldots, W_m$.

Sometimes the underlying uncertainty being addressed revolves around observations of several random variables $V_1, \ldots, V_m$, and their *joint* distribution. The corresponding probability measure on $I\!R^m$ is induced then by the multivariate distribution function

$$F_{V_1,\ldots,V_m}(v_1, \ldots, v_m) = \text{prob}\Big\{ (V_1, \ldots, V_m) \leq (v_1, \ldots, v_m) \Big\}. \tag{6.1}$$

Functions $x = g(v_1, \ldots, v_m)$ give rise to random variables $X = g(V_1, \ldots, V_m)$ having $F_X(x) = \text{prob}\,\{g(V_1, \ldots, V_m) \leq x\}$. Again, the distribution of $(V_1, \ldots, V_m)$ need not be describable by a density function $f(v_1, \ldots, v_m)$. We might be dealing with a discrete distribution of $(V_1, \ldots, V_m)$ corresponding to an m-dimensional "scatter plot."

The standard framework of a *probability space* serves for handling all these aspects of randomness easily and systematically. It consists of a set Ω supplied with a probability measure P_0 and a field $\mathcal{A}$ of its subsets.[53] We think of the elements $\omega \in \Omega$ as "future states" (of information), or "scenarios." Having a subset A of Ω belong to $\mathcal{A}$ means that the probability of ω being in A is regarded as known in the present: $\text{prob}\{A\} = P_0(A)$. In that way, the field $\mathcal{A}$ is a model for present information about the future. There could be multistage approaches to such information, in which $\mathcal{A}$ is just the first in a chain of ever-larger collections of subsets of Ω, but we are not looking at that. A scenario ω could, in our setting, nonetheless involve multiple time periods, but we are not going to consider, here, how additional observations, as the scenario unfolds, might be put to use in optimization.

Random variables in this framework are functions $X : \Omega \to I\!R$, with future outcomes $X(\omega)$, such that, for every $x \in I\!R$, the set $A = \{\, \omega \,|\, X(\omega) \leq x \,\}$ belongs to $\mathcal{A}$.[54] The expected value of a random variable X is the integral $EX = \int_\Omega X(\omega) dP_0(\omega)$. As a special case, the probability space $(\Omega, \mathcal{A}, P_0)$ could be generated by future observations of some variables $V_1, \ldots, V_m$, as above, in which case Ω would be a subset of $I\!R^m$ with elements $\omega = (v_1, \ldots, v_m)$ and P_0 would be the probability measure induced by the joint distribution function $F_{V_1,\ldots,V_m}$. If P_0 has a density function $f(v_1, \ldots, v_m)$ with respect to ordinary integration, then for $X = g(V_1, \ldots, V_m)$ one has $EX = \int g(v_1, \ldots, v_m) f(v_1, \ldots, v_m) dv_1 \ldots dv_m$, but without such a density, it is not possible to rely this way on $dv_1 \ldots dv_m$. That is why, in achieving adequate generality, it is crucial to refer to a background probability measure P_0 as the source of all the distributions that come up.

Despite that focus, a means is provided for considering alternatives P to P_0, and indeed this will be very important in subsequent discussions of risk and its dualization. Other probability measures P can enter the picture as long as the expected value $E_P(X) = \int X(\omega) dP(\omega)$ can be expressed by $E[XQ] = \int X(\omega) Q(\omega) dP_0(\omega)$ for some random variable Q, which is then called *the density of P with respect to P_0* with notation $Q = dP/dP_0$.[55] For instance, in the case where Ω has finitely many elements ω_k, $k = 1, \ldots, N$, if P_0 gives them equal weight $1/N$ but P assigns probability p_k to ω_k, then $Q(\omega_k) = p_k N$.

The need to deal securely with expectations of random variables and certain products of random variables forces some restrictions. For any random variable X, the expressions $||X||_p$ introduced earlier are well defined but could be ∞. It is common practice to work with the spaces[56]

$$\mathcal{L}^p(\Omega) = \mathcal{L}^p(\Omega, \mathcal{A}, P_0) = \Big\{ X \,\Big|\, ||X||_p < \infty \Big\}, \text{ where } \mathcal{L}^1(\Omega) \supset \cdots \supset \mathcal{L}^p(\Omega) \supset \cdots \supset \mathcal{L}^\infty(\Omega). \tag{6.2}$$

For any X in these spaces, EX is well defined and finite, but the situation for products of random variables, like XQ above, is more delicate. While there are options with X in one space and Q in another, no choice is perfect.

[53] We write P_0 for this underlying probability measure in order to reserve P for general purposes below.

[54] The sets $A \in \mathcal{A}$ are called the "measurable" sets and the functions X in question the "measurable" functions.

[55] Such measures P are said to be "absolutely continuous" with respect to P_0.

[56] When Ω is a discrete set of N elements, these spaces coincide and can be identified with $I\!R^N$.

For our purposes here, $\mathcal{L}^2(\Omega)$ has been taken as the platform. That has the simplifying advantage that $E|XQ| < \infty$ for any $X \in \mathcal{L}^2(\Omega)$ and $Q \in \mathcal{L}^2(\Omega)$. However, it does mean that, in considering alternative probability measures P with densities $Q = dP/dP_0$ the restriction must be made to the cases where $\int_\Omega (dP/dP_0)^2(\omega)dP_0 < \infty$. Actually, though, this restriction makes little difference in the end, because other probability measures can adequately be mimicked by these (and for finite Ω is no restriction at all).

Dualization concerns the development of "dual representations" of various functionals, also called "envlope representations," which can yield major insights and provide tools for characterizing optimality. The functionals $\mathcal{F}$ may in general take on ∞ as a value (although usually $-\infty$ is excluded), and some notation for handling that is needed. The *effective domain* of $\mathcal{F}$ is the set

$$\operatorname{dom}\mathcal{F} = \{\, X \in \mathcal{L}^2(\Omega) \,|\, \mathcal{F}(X) < \infty \,\}. \tag{6.3}$$

When $\mathcal{F}$ is convex, this set is convex, but $\mathcal{F}$ closed convex does not necessarily entail $\operatorname{dom}\mathcal{F}$ also being closed. The platform for dualization is a correspondence among closed convex functionals $\mathcal{F}$:[57]

$$\begin{array}{l}\mathcal{F}: \mathcal{L}^2 \to (-\infty,\infty] \text{ closed convex, } \mathcal{F} \not\equiv \infty, \iff \text{ there is a } \mathcal{G}: \mathcal{L}^2 \to (-\infty,\infty],\ \mathcal{G} \not\equiv \infty, \text{ with} \\ \qquad\qquad \mathcal{F}(X) = \sup_{Q \in \mathcal{L}^2(\Omega)} \{\, E[XQ] - \mathcal{G}(Q)\,\} \textit{ for all } X. \\ \text{Moreover the lowest such } \mathcal{G} \text{ is } \mathcal{G} = \mathcal{F}^*, \text{ where } \mathcal{F}^* \text{ is closed convex and given by} \\ \qquad\qquad \mathcal{F}^*(Q) = \sup_{X \in \mathcal{L}^2(\Omega)} \{\, E[XQ] - \mathcal{F}(X)\,\} \text{ for all } Q. \end{array} \tag{6.4}$$

The functional $\mathcal{F}^*$ is said to be conjugate to $\mathcal{F}$, which in turn is conjugate to $\mathcal{F}^*$ through the first formula in (6.4) in the case of $\mathcal{G} = \mathcal{F}^*$, namely

$$\mathcal{F}(X) = \sup_{Q \in \mathcal{L}^2(\Omega)} \{\, E[XQ] - \mathcal{F}^*(Q)\,\} \text{ for all } X. \tag{6.5}$$

The nonempty convex set $\operatorname{dom}\mathcal{F}^* = \{\, Q \,|\, \mathcal{F}^*(Q) < \infty \,\}$ can replace $\mathcal{L}^2(\Omega)$ in this formula, and similarly $\operatorname{dom}\mathcal{F}$ can replace $\mathcal{L}^2(\Omega)$ in the first formula of (6,4).

Here are some cases that will be especially important to us. (The "1" in the second line refers to the constant r.v. with value 1.)

$$\text{for } \mathcal{F} \text{ closed convex } \not\equiv \infty \begin{cases} \mathcal{F}(0) = 0 \iff \inf \mathcal{F}^* = 0, \\ \mathcal{F}(X) \geq EX \iff \mathcal{F}^*(1) \leq 0, \\ \mathcal{F} \text{ is monotonic } \iff Q \geq 0 \text{ when } Q \in \operatorname{dom}\mathcal{F}^*, \\ \mathcal{F} \text{ is pos. homog. } \iff \mathcal{F}^*(Q) = 0 \text{ when } Q \in \operatorname{dom}\mathcal{F}^*. \end{cases} \tag{6.6}$$

In positive homogeneity, this comes down to the following characterization:

$$\begin{array}{l}\text{there is a one-to-one correspondence between nonempty, closed, convex sets } \mathcal{Q} \subset \mathcal{L}^2(\Omega) \\ \text{and closed convex pos. homogeneous functionals } \mathcal{F}: \mathcal{L}^2 \to (-\infty,\infty], \text{ given by} \\ \qquad \mathcal{F}(X) = \sup_{Q \in \mathcal{Q}} E[XQ] \text{ for all } X, \qquad \mathcal{Q} = \{\, Q \,|\, E[XQ] \leq \mathcal{F}(X) \text{ for all } X \,\}. \end{array} \tag{6.7}$$

The second formula identifies $\mathcal{Q}$ with $\operatorname{dom}\mathcal{F}^*$. Any $\mathcal{Q}$ for which the first formula holds must moreover have $\operatorname{dom}\mathcal{F}^*$ as its closed, convex hull.

[57]See Theorem 5 of Rockafellar [1974]. The operation $\mathcal{F} \to \mathcal{F}^*$ is called the Legendre-Fenchel transform.

Envelope Theorem[58]**.** *The functionals $\mathcal{J}$ that are the conjugates $\mathcal{R}^*$ of the regular measures of risk $\mathcal{R}$ on $\mathcal{L}^2(\Omega)$ are the closed convex functionals $\mathcal{J}$ with effective domains $\mathcal{Q} = \operatorname{dom} \mathcal{J}$ such that*

(a) *$EQ = 1$ for all $Q \in \mathcal{Q}$,*

(b) *$0 = \mathcal{J}(1) \leq \mathcal{J}(Q)$ for all $Q \in \mathcal{Q}$,*

(c) *for each nonconstant $X \in \mathcal{L}^2(\Omega)$ there exists $Q \in \mathcal{Q}$ such that $E[XQ] - EX > \mathcal{J}(Q)$.*

The dual representation of $\mathcal{R}$ corresponding to $\mathcal{J} = \mathcal{R}^*$ is

$$\mathcal{R}(X) = \sup_{Q \in \mathcal{Q}} \Big\{ E[XQ] - \mathcal{J}(Q) \Big\}. \tag{6.8}$$

Here $\mathcal{R}$ is positively homogeneous if and only if $\mathcal{J}(Q) = 0$ for all $Q \in \mathcal{Q}$, whereas $\mathcal{R}$ is monotonic if and only if $Q \geq 0$ for all $Q \in \mathcal{Q}$.

If $\mathcal{V}$ is a regular measure of regret that projects to $\mathcal{R}$, then $\mathcal{Q} = \{ Q \in \operatorname{dom} \mathcal{V}^* \mid EQ = 1 \}$ and the conjugate $\mathcal{J} = \mathcal{R}^*$ has $\mathcal{J}(Q) = \mathcal{V}^*(Q)$ for $Q \in \mathcal{Q}$.

The error measure $\mathcal{E}$ paired with the regret measure $\mathcal{V}$ has $\mathcal{E}^*(X) = \mathcal{V}^*(X+1)$. Likewise, the deviation measure $\mathcal{D}$ paired with the risk measure $\mathcal{R}$ has $\mathcal{D}^*(X) = \mathcal{R}^*(X+1)$.

Risk envelopes and identifiers. The convex set $\mathcal{Q}$ in this theorem is called the *risk envelope* associated with $\mathcal{R}$, and a Q furnishing the maximum in (6.8) is a *risk identifier* for X.

The monotonic case in the theorem combines $EQ = 1$ with $Q \geq 0$ and thereby allows us to interpret each $Q \in \mathcal{Q}$ as a probability density dP/dP_0 describing an alternative probability measure P on Ω. For positively homogeneous $\mathcal{R}$, the $\mathcal{J}(Q)$ term drops out of the representation in (6.8) (by being 0). The formula then characterizes $\mathcal{R}(X)$ as giving the worst "cost" that might result from considering the expected values $E[XQ] = E_P[X]$ over all those alternative probability measures P having densities Q in the risk envelope $\mathcal{Q}$.

The nonhomogeneous case has a similar interpretation, but distinguishes within $\mathcal{Q}$ a subset $\mathcal{Q}_0$ consisting of the densities Q for which $\mathcal{J}(Q) = 0$, which always includes $Q \equiv 1$ (the density of P_0 with respect to itself). Densities Q that belong to $\mathcal{Q}$ but not $\mathcal{Q}_0$ have $\mathcal{J}(Q) \in (0, \infty)$. In (6.8) that term then drags the expectation down. In a sense, $\mathcal{J}(Q)$ downgrades the importance of such densities.

The conjugates $\mathcal{V}^*$ of regular measures of regret $\mathcal{V}$ have virtually the same characterization as the conjugates $\mathcal{R}^*$ in the theorem. Property (a) is omitted, but on the other hand there is a provision to enforce the property in (3.15) (in the cases when it is not guaranteed to hold automatically). This provision is that $\mathcal{V}^*(C) < \infty$ for C near enough to 1.

Some examples of risk envelopes in the positively homogeneous case, where (6.8) holds with $J(Q)$ omitted, are the following:

$$\begin{aligned}
&\mathcal{R}(X) = EX + \lambda\sigma(X) \;\longleftrightarrow\; \mathcal{Q} = \Big\{ 1 + \lambda Y \;\Big|\; \|Y\|_2 \leq 1,\ EY = 0 \Big\} \\
&\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X) \;\longleftrightarrow\; \mathcal{Q} = \Big\{ Q \;\Big|\; 0 \leq Q \leq \tfrac{1}{1-\alpha},\ EQ = 1 \Big\} \\
&\mathcal{R}(X) = \sup X \;\longleftrightarrow\; \mathcal{Q} = \Big\{ Q \,|\, Q \geq 0,\ EQ = 1 \Big\} \\
&\mathcal{R}(X) = \textstyle\sum_{k=1}^r \lambda_k \mathcal{R}_k(X) \;\longleftrightarrow\; \mathcal{Q} = \Big\{ \textstyle\sum_{k=1}^r \lambda_k Q_k \;\Big|\; Q_k \in \mathcal{Q}_k \Big\}, \text{ where } \mathcal{R}_k \;\longleftrightarrow\; \mathcal{Q}_k.
\end{aligned} \tag{6.9}$$

Another illustration comes out of Example 6, which can now be formalized via (2.5) in terms of a partition of Ω into disjoint subsets Ω_k of probability $p_k > 0$ with $\sup_k X$ being the essential supremum of X on Ω_k and $E_k X$ being the conditional expectation $E[X|\Omega_k]$:

$$\mathcal{R}(X) = \sum_{k=1}^r p_k \sup\nolimits_k X \;\longleftrightarrow\; \mathcal{Q} = \Big\{ Q \geq 0 \,|\, E[Q|\Omega_k] = p_k \Big\}. \tag{6.10}$$

[58] Most of the facts in this compilation, which follow from the general properties of conjugacy as above, are already well understood and have been covered, for instance, in Föllmer and Schied [2004]. The new aspects are the dualization of aversity in condition (c) and the final assertion, connecting with the dualization of regret.

Examples beyond positive homogeneity, where nonzero values of $\mathcal{J}$ may enter, are simple to work out in the expectation case:

> For quadrangles in the Expectation Theorem, with regret $\mathcal{V}(X) = E[v(X)]$, the conjugate $\mathcal{J} = \mathcal{R}^*$ of the risk measure $\mathcal{R}$ projected from $\mathcal{V}$ is given by
>
> $$\mathcal{J}(Q) = \begin{cases} E[v^*(Q)] & \text{if } EQ = 1 \\ \infty & \text{if } EQ \neq 1 \end{cases} \quad \text{for the function } v^* \text{ conjugate to } v, \tag{6.11}$$
>
> given by $v^*(q) = \sup_x \{ xq - v(x) \}$. The properties of v^* corresponding to those of v in (4.3), are that v^* is closed convex with $v^*(1) = 0,\ v^{*\prime}(1) = 0$.

This holds from the description in Envelope Theorem of the $\mathcal{J}$ in projection from $\mathcal{V}$ because the functional conjugate to $\mathcal{V}(X) = E[v(X)]$ is $\mathcal{V}^*(Q) = E[v^*(Q)]$.[59] The dualization of the properties of v to those of v^* comes from one-dimensional convex analysis; see Rockafellar [1970].

An especially interesting illustration is furnished by Example 8, where one has

$$v(x) = \exp x - 1, \qquad v^*(q) = \begin{cases} q \log q - q & \text{if } q > 0, \\ 0 & \text{if } q = 0, \\ \infty & \text{if } q < 0. \end{cases} \tag{6.12}$$

Through (6.11) this yields

$$\mathcal{R}(X) = \log E[\exp X] \ \longleftrightarrow \ \mathcal{J}(Q) = \begin{cases} E[Q \log Q] & \text{if } Q \geq 0,\ EQ = 1, \\ \infty & \text{otherwise.} \end{cases} \tag{6.13}$$

Here $-\mathcal{J}(Q)$ is a well known expression for the *entropy* with respect to the probability measure P_0 of the probability measure P having $Q = dP/dP_0$.[60]

The idea that the functional $-\mathcal{J} = -\mathcal{R}^*$ stands, in general, for a sort of *generalized entropy* associated with the risk measure $\mathcal{R}$ is tantalizing and can be carried quite far. The focus needs to be on the case of monotonic $\mathcal{R}$, of course, in order that the elements of $\mathcal{Q} = \operatorname{dom} \mathcal{J}$ may be interpreted as probability densities. We can experiment in that case with calling $-\mathcal{J}$ the *$\mathcal{R}$-entropy functional* with respect to the underlying P_0, noting that Q will be a convex set containing the density 1, and that $\mathcal{J}(1) = 0 = \min \mathcal{J}$.

To see where this can lead, consider the following "generalized entropy problem" which seeks, through Q and subject to side conditions on expectations, a probability measure P that is "closest" to P_0 in the sense of maximizing $\mathcal{R}$-entropy:

$$\text{maximize } \ -\mathcal{J}(Q) \ \text{ over } Q \in \mathcal{Q} \text{ subject to } \ E[X_i Q] = C_i \ \text{ for } i = 1, \ldots, m. \tag{6.14}$$

For the entropy in (6.13) this is a classical "moment problem."

Conjugate duality theory, cf. Rockafellar [1974], can associate this with problem a dual problem of minimization by way of the Lagrangian

$$L(Q, \lambda_1, \ldots, \lambda_m) = -\mathcal{J}(Q) + \sum\nolimits_{i=1}^m \lambda_k E[X_k - C_k] \text{ with multipliers } \lambda_k. \tag{6.15}$$

Through the conjugacy correspondence between $\mathcal{J}$ and $\mathcal{R}$ this dual problem comes out to be

$$\text{minimize } \mathcal{R}\Big(\sum\nolimits_{i=1}^m \lambda_i [X_i - C_i] \Big) \ \text{ over } (\lambda_1, \ldots, \lambda_m) \in I\!R^m. \tag{6.16}$$

[59] This follows a general rule of convex analysis in [Rockafellar, 1974, Theorem 21]. The "inner product" in the function space $\mathcal{L}^2(\Omega)$ is $\langle X, Q \rangle = E[XQ]$.

[60] See Ben-Tal and Teboulle [2007] for more background.

If $(\lambda_1, \ldots, \lambda_m)$ solves (6.16) and Q is a risk identifier for $\sum_{i=1}^m \lambda_i [X_i - C_i]$, or in other words, Q maximizes $L(Q, \lambda_1, \ldots, \lambda_m)$ in (6.15), then Q solves (6.14). Results in Rockafellar [1974] can carry the details of this further. Even more powerful developments of optimization duality, tailored to the fine points of financial mathematics, have recently been contributed by Pennanen [2011]. For more insights on entropic modeling versus risk, see Grechuk et al. [2008], which emphasizes the role of deviation measures.

Also coming out of the Envelope Theorem is further insight into the degree of nonuniqueness of the error measures $\mathcal{E}$ that project to a specified deviation measure $\mathcal{D}$, or the regret measures $\mathcal{V}$ that project to a specified risk measure $\mathcal{R}$. In the positively homogeneous case of $\mathcal{V}$, for instance, the conjugate $\mathcal{V}^*$ has by (6.6), (6.7), the simple form that it is 0 on a certain closed, convex set $\mathcal{K}$ but ∞ outside of $\mathcal{K}$; then $\operatorname{dom} \mathcal{V}^* = \mathcal{K}$. The theorem says the risk envelope $\mathcal{Q}$ determining the risk measure $\mathcal{R}$ projected from $\mathcal{V}$ has $\mathcal{Q}$ equal to the intersection of $\mathcal{K}$ with the hyperplane $\{ Q \mid EQ = 1 \}$. That intersection only uses one "slice" of $\mathcal{K}$. Different $\mathcal{K}$'s that agree for this "slice" will give different $\mathcal{V}$'s yielding the same $\mathcal{R}$. Discovering a "natural" antecedant $\mathcal{V}$ for $\mathcal{R}$ therefore amounts geometrically to discovering a "natural" extension $\mathcal{K}$ of $\mathcal{Q}$ beyond the hyperplane $\{ Q \mid EQ = 1 \}$.

References

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

Acerbi, C. and Tasche, D. (2002) "On the coherence of expected shortfall," *Journal of Banking and Finance* **26**, 1487–1503.

Optimization and Risk Management Case Studies with Portfolio SafeGuard (PSG), AORDA–American Optimal Decisions, 2010; see Mortgage Pipeline Hedging.

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

Basova, H. G., Rockafellar, R. T., and Royset, J. O. (2011) "A computational study of buffered failure probability in reliability-based design optimization," *Proceedings of the 11th Conference on Application of Statistics and Probability in Civil Engineering, Zürich, Switzerland, 2011.*

Ben Tal, A., and Ben-Israel, A. (1991) "A recourse certainty equivalent for decisions under uncertainty," *Annals of Operations Research* **30**, 3–44.

Ben Tal, A., and Ben-Israel, A. (1997) "Duality and equilibrium prices in economics of uncertainty," *Mathematical Methods of Operations Research* **46**, 51–85.

Ben Tal, A., and Teboulle, M. (1991) "Portfolio theory for the recourse certainty equivalent maximizing investor," *Annals of Operations Research* **31**, 479–499.

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

Dembo, R. S. and King, A. J. (1992) "Tracking models and the optimal regret distribution in asset allocation," *Applied Stochastic Models* **8**, 151–157.

Föllmer, H. and Schied, A. (2004) *Stochastic Finance*, 2nd edition, De Gruyter, New York.

Golodnikov, A., Macheret, Y, Trindate, A., Uryasev, S., and Zrazhevski, G. (2007) "Statistical modelling of composition and processing parameters for alloy development: a statistical model-based approach," *J. Industrial and Management Optimization* **3** (view online).

Gneiting, T. (2011) "Making and evaluating point forecasts," *Journal of the American Statistical Society* **106**, 746–762.

Grechuk, B., Molyboha, A., Zabarankin, M. (2008) "Maximum entropy principle with general deviation measures," *Mathematics of Operations Research* **34** (2009), 445–467.

Kahneman, D., and Tversky, A. (1979) "Prospect theory: an analysis of decision under risk," *Econometrica* **57**, 263–291.

Koenker, R. and Bassett, G. (1978) "Regression quantiles," *Econometrica* **46**, 33–50.

Koenker, R. (2005) *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press.

Krokhmal, P. A. (2007) "Higher moment risk measures," *Quantitative Finance* **7**, 373–387.

Ogryczak, W. and Ruszczyński, A. (1997) "From stochastic dominance to mean-risk models: semideviations as risk measures," *Technical Report IR-97-027*, International Institute for Applied Systems Analysis.

Pennanen, T. (2011) "Convex duality in stochastic programming and mathematical finance," *Mathematics of Operations Research* **36**, 340–362.

Pflug, G. (2000) "Some remarks on the value-at-risk and the conditional value-at-risk," *Probabilistic Constrained Optimization: Methodology and Applications,* S. Uryasev, ed., Kluwer Academic Publishers, Norwell, MA.

Rockafellar, R. T. (1970) *Convex Analysis*, Princeton University Press, Princeton, NJ.

Rockafellar, R. T. (1974) *Conjugate Duality and Optimization*, No. 16 in the Conference Board of Math. Sciences Series, SIAM, Philadelphia.

Rockafellar, R. T. (2007) "Coherent approaches to risk in optimzation under uncertainty," *Tutorials in Operations Research INFORMS 2007*, 38–61.

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

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

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

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2002) "Deviation measures in risk analysis and optimization," *Technical Report 2002-7*, Department of Industrial and Systems Engineering, University of Florida.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006a) “Generalized deviations in risk analysis,” *Finance and Stochastics* **10**, 51–74.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006b) “Master funds in portfolio analysis with general deviation measures,” *Journal of Banking and Finance* **30** (2), 743–778.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006c) “Optimality conditions in portfolio analysis with general deviation measures,” *Mathematical Programming, Series B* **108**, 515–540.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2008) “Risk tuning with generalized linear regression,” *Math. of Operations Research* **33** (3), 712–729.

Rockafellar, R. T., and Wets, R. J-B (2007) *Variational Analysis,* Springer-Verlag, New York.

Roell, A. (1987) “Risk aversion in Quiggin and Yaari’s rank-order model of choice under uncertainty,” *The Economic Journal, Issue Supplement: Conference papers* **97**, 143–159.

Ruszczyński, A., and Shapiro, A. (2006a) “Optimization of convex risk functions,” *Math. of Operations Research* **31**, 433–452.

Ruszczyński, A., and Shapiro, A. (2006b) “Conditional risk mappings,” *Math. of Operations Research* **31**, 544–561.

Samson,S., Thoomu, S., Fadel, G., and Reneke, J. (2009), “Reliable design optimization under aleatory and epistemic uncertainties,” *Proceedings of IDETC/DAC 2009, ASME 2009 International Design Engineering Technical Conferences & 36th Design Automation Conference*, San Diego, California, August 30–September 2, 2009.

Trindade, A. and Uryasev, S. (2005) “Improved tolerance limits by combining analytical and experimental data: an information integration methodology”.

Trindade, A. and Uryasev, S. (2006) “Optimal determination of percentiles and allowables: CVaR regression approach,” in: “Robust Optimization-Directed Design” (A.J. Kurdila et al., eds.), 179–247, Springer Publishers.

Trindade A., Uryasev, S., Shapiro, A., and Zrazhevsky, G. (2007) “Financial prediction with constrained tail risk,” *Journal of Banking and Finance* **31**, 3524–3538.

Yaari, M. E. (1987) “The dual theory of choice under risk,” *Econometrica* **55**, 95–115.