THE RISK QUADRANGLE IN RISK MANAGEMENT, OPTIMIZATION, STATISTICAL ESTIMATION, AND MACHINE LEARNING II

Bogdan Grechuk* Anton Malandii† Terry Rockafellar‡ Stan Uryasev†

June 26, 2025

Abstract

This paper extends the 2013 development by Rockafellar and Uryasev of the Risk Quadrangle (RQ) as a unified scheme for integrating risk management, optimization, and statistical estimation. The RQ features four stochastic functionals — risk, deviation, regret, and error, along with an associated "stastic," and articulates their revealing and in some ways surprising interrelationships and dualizations. Quadrangle additions and facts that have come to light since 2013 are reviewed in a synthesis focused on both theoretical advancements and practical applications. New quadrangles — superquantile, superquantile norm, expectile, biased mean, quantile symmetric average union, and φ-divergence-based quadrangles — offer novel approaches to risk-sensitive decision-making across various fields such as machine learning, statistics, finance, and PDE-constrained optimization.

The theoretical contribution comes in axioms for "subregularity" which relax the "regularity" of the quadrangle functionals where what has turned out to be too restrictive for some budding applications. The main RQ theorems and connections are revisited and rigorously extended to this more ample framework. Examples are provided in portfolio optimization, regression and classification, that demonstrate the advantages and the role played by duality, especially in ties to robust optimization and generalized stochastic divergences.

Keywords: risk quadrangle, risk, deviation, error, regret, quantile, superquantile, value-at-risk, conditional value-at-risk, expected shortfall, coherency, convexity, duality, stochastic optimization, regression, stochastic divergences, robust optimization.

Contents

*School of Computing and Mathematic Sciences, University of Leicester, Leicester LE1 7RH, UK;
email: bg83@leicester.ac.uk
†Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, NY 11794, USA;
emails: anton.malandii@stonybrook.edu, stanislav.uryasev@stonybrook.edu
‡University of Washington, Department of Applied Mathematics, 408 L Guggenheim Hall, PS Box 352420, Seattle, WA 98195, USA;
email: rtr@math.washington.edu

1 Introduction

This paper extends the work of Rockafellar and Uryasev [2013] and reviews the new results published since 2013. Although the material in the paper is self-contained and can be read independently, readers are strongly advised to familiarize themselves with the content of Rockafellar and Uryasev [2013].

Background. The Risk Quadrangle (RQ) is a unified framework that integrates risk management, optimization, and statistical estimation by grouping four axiomatically defined stochastic functionals: risk, deviation, regret and error connected by a so-called statistic. For a random variable X, *risk* provides a numerical surrogate for the overall hazard in X, *deviation* measures the "nonconstancy" in X, *regret* assesses the distaste in facing the mix of good-and-bad outcomes of X, and *error* quantifies the "nonzeroness" in X. The *statistic* identifies the constant closest to X with respect to a particular form of error functional, but also gives the level of trade-off between future loss and immediate acceptance that is appropriate for a particular form of regret functional. The fruitful interplay between these functionals supports a vast spectrum of applications across all engineering areas where probabilistically modeled uncertainty is involved. Among them, to name a few, are quantitative finance, reliability engineering, mechanical engineering, medical imaging, and machine learning.

The Risk Quadrangle

	risk $\mathcal{R} \longleftrightarrow \mathcal{D}$ deviation	
optimization	$\uparrow\downarrow \quad \mathcal{S} \quad \downarrow\uparrow$	*estimation*
	regret $\mathcal{V} \longleftrightarrow \mathcal{E}$ error	

General Relationships

$$\mathcal{R}(X) = \mathcal{D}(X) + \mathbb{E}[X] \qquad \mathcal{V}(X) = \mathcal{E}(X) + \mathbb{E}[X]$$
$$\mathcal{D}(X) = \min_C \{ \mathcal{E}(X - C) \} \qquad \mathcal{R}(X) = \min_C \{ C + \mathcal{V}(X - C) \}$$
$$\underset{C}{\text{argmin}} \{ \mathcal{E}(X - C) \} = \mathcal{S}(X) = \underset{C}{\text{argmin}} \{ C + \mathcal{V}(X - C) \}$$

The relationship between risk and regret allows: 1) constructing new risk measures and quadrangles; 2) building efficient algorithms for risk optimization (see e.g., Rockafellar and Uryasev [2000, 2002]; Ben-Tal and Teboulle [2007]; Kuzmenko [2020]; Malandii et al. [2024], and more). The relation between deviation and error links statistical estimation and risk management, resulting in concepts such as risk-tuned regression and risk tracking (see Rockafellar et al. [2008]; Rockafellar and Royset [2015]). This relation shows equivalences among various types of regressions and provides robust and efficient estimation techniques (see Rockafellar et al. [2014]; Golodnikov et al. [2019]; Malandii et al. [2024]; Malandii and Uryasev [2024]). For more details, see Rockafellar and Uryasev [2013]; Royset [2022].

Motivation. The motivation of this paper is twofold: on the *practical side*, it aims to offer engineers and practitioners robust, implementable methods for regression and risk estimation under real-world constraints; on the *theoretical side*, it seeks to advance foundational principles in stochastic optimization, distributionally robust optimization, convex analysis, and statistical estimation, targeting pure mathematicians who explore deep relationships within these fields.

Practical aspects. In their original work, Rockafellar and Uryasev [2013] established the foundation of the RQ, formulating key theorems and relationships that unify the RQ elements within a single mathematical structure. The innovations were largely theoretical, without extensive demonstrations of practical applications or empirical benefits, but in the last decade, a series of other works has begun to fill the gap. The RQ scheme has been leveraged to answer application-driven questions with practical results in risk management, statistical estimation, machine learning, fairness-aware machine learning, and beyond. This has encompassed new additions such as the superquantile quadrangle (Rockafellar et al. [2014]), the superquantile norm quadrangle (Mafusalov and Uryasev [2016]), the expectile quadrangles (Malandii et al. [2024]), the biased mean quadrangle (Malandii and Uryasev [2024]), and the quantile symmetric average union quadrangle (Malandii and Uryasev [2022]).

In particular, these recent contributions illustrate how the RQ provides a powerful framework for constructing new regression estimators. For example, the expectile and least squares regressions (known to be quadratic programs) have been reduced to linear programs, enabling fast algorithms that also accommodate mixed-integer constraints–an essential feature in many real-world applications. The biased mean quadrangle introduced a novel estimator for the so-called biased mean, broadening the scope of classical methods in robust statistics. Likewise, the superquantile norm quadrangle has uncovered fundamental connections between support vector regression

(SVR) and distributionally robust optimization, while the quantile symmetric average union quadrangle has shown that SVR estimates the average of two symmetric quantiles, thus clarifying its theoretical and practical strengths and limitations.

Another active area of research—fairness-aware machine learning—has also significantly benefited from the development of the axiomatic theory of risk and deviation measures. In particular, Williamson and Menon [2019] defined fairness as a deviation in subgroup risks, which enabled the reduction of fairness-aware classification to a convex optimization problem. This problem can be efficiently solved by leveraging the relationship between risk and regret.

Despite this progress, the emerging body of work remains scattered across multiple publications, each focusing on a specific quadrangle. Researchers and practitioners seeking to exploit these new developments lack a single, cohesive resource detailing how the RQ can be specialized, extended, and integrated into broad classes of optimization and estimation problems.

This paper synthesizes recent advancements in the RQ framework, highlighting how these new quadrangles were derived and the benefits they offer for risk-sensitive decision-making. The following Map of Applications illustrates several representative examples from machine learning, statistics, quantitative finance, and risk-averse PDE-constrained optimization (see, for example, Antil et al. [2018]), showcasing the breadth of the framework. The following Portfolio Optimization example for risk and deviation is based on general relationships between elements of quadrangle, see Theorem 3.1. The Expectile Minimization example is based on the Expectile Quadrangle considered in Example 7. The Linear Regression example is based on Regression Theorem 4.1.

Map of Applications

Unsupervised learning

Given: $\mathbf{X} = (X_1, \dots, X_d)$ = features

Model: $\ell(\mathbf{w}; \mathbf{X})$ = loss, $\mathbf{w} \in \mathcal{W}$

Problem: $\min_{\mathbf{w}} \mathcal{R}(\ell(\mathbf{w}; \mathbf{X}))$ or $\min_{\mathbf{w}} \mathcal{D}(\ell(\mathbf{w}; \mathbf{X}))$

Supervised learning

$\mathbf{X} = (X_1, \dots, X_d)$ = features, Y = target

$\ell(\mathbf{w}; \mathbf{X}, Y)$ = loss, $\mathbf{w} \in \mathcal{W}$

$\min_{\mathbf{w}} \mathcal{R}(\ell(\mathbf{w}; \mathbf{X}, Y))$ or $\min_{\mathbf{w}} \mathcal{E}(\ell(\mathbf{w}; \mathbf{X}, Y))$

Examples

Portfolio Optimization

$$\ell(\mathbf{w}; \mathbf{X}) = -\mathbf{w}^\top \mathbf{X}$$

$$\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^d : \mathbf{w}^\top \mathbf{1} = 1,\ \mathbb{E}[\mathbf{w}^\top \mathbf{X}] = \mu \geq 0\}$$

$$\min_{\mathbf{w} \in \mathcal{W}} \mathcal{R}(l(\mathbf{w}; \mathbf{X})) = \min_{\mathbf{w} \in \mathcal{W}, C} \{C + \mathcal{V}(l(\mathbf{w}; \mathbf{X}) - C)\}$$

$$\text{or } \min_{\mathbf{w} \in \mathcal{W}} \mathcal{D}(l(\mathbf{w}; \mathbf{X})) = \min_{\mathbf{w} \in \mathcal{W}, C} \mathcal{E}(l(\mathbf{w}; \mathbf{X}) - C) =$$

$$= \min_{\mathbf{w} \in \mathcal{W}} \{\mathcal{R}(l(\mathbf{w}; \mathbf{X})) - \mathbb{E}[l(\mathbf{w}; \mathbf{X})]\}$$

Linear Regression

$$\ell(\mathbf{w}; \mathbf{X}, Y) = Y - \bar{\mathbf{w}}^\top \mathbf{X} - w_0$$

$$\mathcal{W} = \mathbb{R}^{d+1}, \quad \mathbf{w} = (\bar{\mathbf{w}}, w_0)$$

$$\min_{\mathbf{w} \in \mathcal{W}} \mathcal{E}(\ell(\mathbf{w}; \mathbf{X}, Y)) =$$

$$= \min_{\mathbf{w} \in \mathcal{W}} \mathcal{D}(\ell(\mathbf{w}; \mathbf{X}, Y)) \text{ s.t. } 0 \in \mathcal{S}(\ell(\mathbf{w}; \mathbf{X}, Y))$$

PDE-constrained Expectile Minimization

$$\ell(\mathbf{w}; \mathbf{X}) = L(u(\mathbf{w}; \mathbf{X}), \mathbf{X}),\ \mathcal{B} = \text{Banach space}$$

$$\mathcal{W} = \{\mathbf{w} \in \mathcal{B} : \underbrace{f(u(\mathbf{w}; \mathbf{X}), \mathbf{w}, \mathbf{X})}_{\text{PDE in weak form}} = 0\}$$

$$\min_{\mathbf{w} \in \mathcal{W}} e_K(\ell(\mathbf{w}; \mathbf{X})) =$$

$$= \min_{\mathbf{w} \in \mathcal{W}, C} C + \{\mathbb{E}[\ell(\mathbf{w}; \mathbf{X}) - C + \frac{1}{K}(\ell(\mathbf{w}; \mathbf{X}) - C)_+]\}_+$$

Support Vector Classification

$$\ell(\mathbf{w}; \mathbf{X}, Y) = -Y(\bar{\mathbf{w}}^\top \mathbf{X} + w_0),\ Y = \{-1, +1\}$$

$$\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^{d+1} : \|\bar{\mathbf{w}}\|_2 \leq 1,\ w_0 \in \mathbb{R}\}$$

$$\min_{\mathbf{w} \in \mathcal{W}} \text{CVaR}_\alpha(\ell(\mathbf{w}; \mathbf{X}, Y)) =$$

$$= \min_{\mathbf{w} \in \mathcal{W}, C} \{C + \frac{1}{1-\alpha} \mathbb{E}[l(\mathbf{w}; \mathbf{X}) - C]_+\}$$

Theoretical aspects. Practical challenges and advances have in turn stimulated the theoretical development of the framework — especially its axiomatic foundations. The basic theorems and relationships have accordingly needed updating beyond their original forms in Rockafellar and Uryasev [2013]. Interest in generalized stochastic divergences in connection with distributional robustness, for example, has led to the discovery that such divergence functionals are actually dual to risk functionals, but of a sort not covered by the original scheme. That has pointed to the need for some relaxation in the original axioms.

The stochastic functionals that comprise the quadrangle are axiomatically defined (see The List of Potential Axioms), with their axioms deeply rooted in convex analysis and classical statistics. The axiomatic nature of the framework elevates the optimization problem settings beyond the expectation-type objectives that have long been common in statistics and more recently in machine learning. The significance of objectives other than expected loss in machine learning became evident with the development of ν–Support Vector Machines by Schölkopf et al. [2000].

The rapid development and global influence of machine learning have revealed the need for axiomatic revisions in the RQ setting. Specifically, several useful error quantifiers such as the Vapnik error (or expected ε–insensitive loss), $\mathcal{E}(X) = \mathbb{E}[|X| - \varepsilon]$, $\varepsilon \geq 0$ and the superexpectation error, $\mathcal{E}(X) = \{\mathbb{E}[X_-] - x_+, \mathbb{E}[X_+] - x_-\}$, $x \in \mathbb{R}$, do not satisfy the strict positivity axiom (A9) (see The List of Potential Axioms), which stimulates an appropriate relaxation of this particular axiom, followed by axioms (A6)–(A8). The strict positivity axiom states that an error of a random variable X should always be positive whenever X is nonzero, which holds for most classical errors, such as mean squared error, mean absolute error, etc. However, in practice, there may exist so-called "error-insensitive" regions, where errors below a certain threshold are ignored. The most famous example is in ε–Support Vector Regression (cf. Drucker et al. [1996]; Vapnik et al. [1996]), where the Vapnik error is minimized. To address the aforementioned issues, this paper introduces *subregularity axioms*, where the concept of subaversity is used to relax the (A6)–(A9). The proofs of the fundamental theorems have to be adjusted to incorporate this extension.

List of Potential Axioms

For a functional $\mathcal{F} : \mathcal{L}^p \to (-\infty, \infty]$ consider the following set of axioms:

(A1) *Convexity:* $\mathcal{F}(\lambda X + (1-\lambda)Y) \leq \lambda \mathcal{F}(X) + (1-\lambda)\mathcal{F}(Y)$ for all $X, Y \in \mathcal{L}^p$ and every $\lambda \in [0, 1]$.

(A2) *Lower semicontinuity:* $\{\mathcal{F} \leq C\} := \{X \in \mathcal{L}^p : \mathcal{F}(X) \leq C\}$ is closed for all $C \in \mathbb{R}$.

(A3) *Constant fidelity:* $\mathcal{F}(X) = C \in \mathbb{R}$ if $\mathbb{P}(X = C) = 1$.

(A4) *Constant neutrality:* $\mathcal{F}(X) = 0 \in \mathbb{R}$ if $\mathbb{P}(X = C) = 1$.

(A5) *Zero fidelity:* $\mathcal{F}(X) = 0 \in \mathbb{R}$ if $\mathbb{P}(X = 0) = 1$.

(A6) *Aversity:* $\mathcal{F}(X) > \mathbb{E}[X]$ for all $X \neq C$.

(A7) *Zero aversity:* $\mathcal{F}(X) > \mathbb{E}[X]$ for all $X \neq 0$.

(A8) *Nonconstant positivity:* $\mathcal{F}(X) > 0$ for all $X \neq C$.

(A9) *Strict positivity:* $\mathcal{F}(X) > 0$ for all $X \neq 0$.

(A10) *Monotonicity:* $\mathcal{F}(X) \leq \mathcal{F}(Y)$ when $\mathbb{P}(X \leq Y) = 1$.

(A11) *Positive homogeneity:* $\mathcal{F}(\lambda X) = \lambda \mathcal{F}(X)$ when $\lambda > 0$.

Regular Quadrangle Axioms

($\mathcal{R}$) *Regular risk*: (A1), (A2), (A3), (A6).

($\mathcal{V}$) *Regular regret*: (A1), (A2), (A5), (A7).

($\mathcal{D}$) *Regular deviation*: (A1), (A2), (A4), (A8).

($\mathcal{E}$) *Regular error*: (A1), (A2), (A5), (A9).

Originally in Rockafellar and Uryasev [2013], the definitions of regular error and regret also included limit conditions. For $\mathcal{E}$, this required that $\lim_{k\to\infty} \mathbb{E}[X_k] = 0$ for any sequence of r.v.s $\{X_k\}_{k=1}^{\infty}$ satisfying $\lim_{k\to\infty} \mathcal{E}[X_k] = 0$, and there was something analogous for $\mathcal{V}$. Those conditions were needed back then for the proofs of some theorems, but in the meantime they have been found to be superfluous — other proofs succeed without invoking them. Presenting the RQ framework in that simpler form is one of the contributions of this paper.

For the orientation of readers who may be familiar with risk measures $\mathcal{R}$ that are *coherent*, but perhaps new to risk measures that are *regular*, an explanation of the relationship may be helpful. As introduced in the pioneering days of risk theory, a coherent measure of risk had to satisfy axioms that, although stated differently, were equivalent to the combination of (A1), (A2), (A3), (A10) and (A11) for $\mathcal{R}$ in place of $\mathcal{F}$ (although (A2) did not appear from the start, because the original context was such that the convexity of the functionals entailed their continuity). Later there was incentive on many fronts to drop (A11) from this list, and then $\mathcal{R}$ was called, by some, simply a convex measure of risk. The trouble with that there are very prominent instances of risk measure $\mathcal{R}$, such as the one in the Standard Quadrangle at the beginning of Section 2, that are convex but lack the monotonicity in (A10). That monotonicity, however, is a watershed requirement for dualization that reflects in terms of alternative probability distributions compared to the nominal one. Rockafellar suggested it would make better sense to tie coherency to that monotonicity and adjust the terminology by speaking of risk measures satisfying (A1), (A2), (A3), and (A10) as coherent in the general sense. But if mononicity is so important, why is (A10) absent from the definition of a regular risk measure $\mathcal{R}$ in Rockafellar and Uryasev [2013]? That is because the Standard Quadrangle, with its long-running usage and deep connection to classical statistics, had to be admitted to the RQ picture despite its lack of (A10), in particular for bringing out that deficiency.

Incentives coming from stochastic divergences. The stochastic modeling of uncertainty entails an underlying probability space $(\Omega, \mathcal{M}, \mathbb{P}_0)$, which is used to define the random losses (or costs), $\ell(\mathbf{w}, \omega)$, $\omega \in \Omega$, as real-valued random variables. Specifically, the probability of a random loss being less than a certain numerical threshold is measured by the cumulative distribution function defined by the underlying probability measure, i.e., $F(x) = \mathbb{P}_0(\{\omega \in \Omega : \ell(\mathbf{w}, \omega) \leq x\})$. When the goal is to find a decision $\mathbf{w}$ that "minimizes" the loss function, the stochastic nature of this function raises the question: "minimizes in which sense?" The answer to this question significantly relies on the information regarding the underlying probability $\mathbb{P}_0$, which is usually either unknown or only partially available.

When the natural designation of $\mathbb{P}_0$ is at hand, the classical approach suggests minimizing the average loss, i.e.,

$$\min_{\mathbf{w}} \ \mathbb{E}_{\mathbb{P}_0}[\ell(\mathbf{w}, \omega)].$$

The average loss minimization is often referred to as *risk-neutral* optimization. On the other hand, if there is a lack of information regarding the underlying probability, the *robust* approach suggests minimizing the loss associated with the worst outcome $\omega \in \Omega$, i.e.,

$$\min_{\mathbf{w}} \max_{\omega \in \Omega} \ \ell(\mathbf{w}, \omega).$$

While this approach has its merits and successes, it can also be overly conservative and, consequently, costly.

The *distributionally robust* approach offers a sort of compromise between neutrality and conservativeness. Rather than relying solely on a single distribution $\mathbb{P}_0$ or avoiding probabilities entirely, we can consider sets of alternative distributions $\mathbb{P} \in \mathcal{P}^{\mathcal{R}}$ and minimize the expected value with respect to the worst one, i.e.,

$$\min_{\mathbf{w}} \max_{\mathbb{P} \in \mathcal{P}^{\mathcal{R}}} \ \mathbb{E}_{\mathbb{P}}[\ell(\mathbf{w}, \omega)].$$

Consequently, it raises the question of how to construct the set of probability alternatives $\mathcal{P}^{\mathcal{R}}$. It turns out that the answer is intimately connected to the duality theory for coherent measures of risk $\mathcal{R}$ in the basic sense, satisfying (A1), (A2), (A3), (A10) and (A11), and the concept of distributional robustness in that way is an integral part of the RQ picture. A key there to constructing sets $\mathcal{P}^{\mathcal{R}}$ consisting of probability alternatives is to take those sets to be "neighborhoods" of nominal probability measure $\mathbb{P}_0$ with respect to some sort of "distance" concept. Such distances are furnished by general *stochastic divergence* functionals. That much has been understood by practitioners in machine learning and elsewhere, but here there will be something more. Stochastic divergence functionals will be given an axiomatic definition and seen as duals to special risk measures that are not quite regular, but fit with a relaxation of regularity in addition to being coherent in the general sense.

Outline. The rest of the paper is organized as follows. Section 2 provides a list of examples of risk quadrangles highlighting the framework's scope and motivating its further extensions.

Section 3 reviews and extends the main properties and relationships of the subregular RQ. Specifically, subregularity axioms and definitions are given in Subsection 3.1. The discussion on the fundamental theorems– cornerstones of the framework– is a subject of Subsection 3.2. The primal aspects of quadrangle construction are covered in Subsection 3.2.1. Dual representation and conjugate duality of the subregular risk, deviation, regret, and error are discussed in Subsection 3.2.2, and the concept of generating families of positive homogeneous functionals is presented in Subsection 3.3. Section 4 is devoted to discussing generalized regression and statistical estimation in the subregular RQ framework. Finally, Section 5 overviews the ideas of robust and distributionally robust optimization in the subregular RQ framework.

2 Examples of Risk Quadrangles

Following Rockafellar and Uryasev [2013], we start with the examples highlighting the framework's scope and motivating its further extension. We begin with a few main examples from Rockafellar and Uryasev [2013], and then continue with new examples.

Example 1 (Standard Quadrangle, $\lambda > 0$)**.** This is Example 1 in Rockafellar and Uryasev [2013], where it is called the Mean-based Quadrangle.

Standard Quadrangle
$\mathcal{S}(X) = \mathbb{E}[X] = \text{mean}$ $\mathcal{R}(X) = \mathbb{E}[X] + \lambda\sigma(X) = \text{safety margin tail risk}$ $\mathcal{D}(X) = \lambda\sigma(X) = \text{standard deviation, scaled}$ $\mathcal{V}(X) = \mathbb{E}[X] + \lambda\|\|X\|\|_2 = \mathcal{L}^2\text{-regret, scaled}$ $\mathcal{E}(X) = \lambda\|\|X\|\|_2 = \mathcal{L}^2\text{-error, scaled}$

Example 2 (Quantile-based Quadrangle, $\alpha \in (0,1)$)**.** This is Example 2 in Rockafellar and Uryasev [2013]. Recall that for any r.v. X its cumulative distribution function is $F_X(x) = \text{prob}\,(X \le x)$, and its α-quantile is $q_\alpha(X) = [q_\alpha^-(X), q_\alpha^+(X)]$, where $q_\alpha^-(X) = \sup\{x \,|\, F_X(x) < \alpha\}$ and $q_\alpha^+(X) = \inf\{x \,|\, F_X(x) > \alpha\}$. Recall also notation $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$.

Quantile-based Quadrangle

$$
\begin{aligned}
\mathcal{S}(X) &= \text{VaR}_\alpha(X) = q_\alpha(X) = \text{quantile}\\
\mathcal{R}(X) &= \text{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \text{VaR}_\beta(X)d\beta = \text{CVaR}\\
\mathcal{D}(X) &= \text{CVaR}_\alpha(X - \mathbb{E}[X]) = \text{CVaR-deviation}\\
\mathcal{V}(X) &= \frac{1}{1-\alpha}\mathbb{E}[X_+] = \text{average absolute loss, scaled}\\
\mathcal{E}(X) &= \mathbb{E}\left[\frac{\alpha}{1-\alpha}X_+ + X_-\right] = \text{normalized Koenker-Bassett error}
\end{aligned}
$$

We refer to Rockafellar and Uryasev [2013] for many more examples. We next present quadrangles not listed in Rockafellar and Uryasev [2013].

Example 3 (CVaR-based Quadrangle, $\alpha \in (0,1)$)**.** Estimating conditional value-at-risk (CVaR, also called superquantile) via regression has long been challenged by CVaR's known non-elicitability property Chun et al. [2012]; Gneiting [2011], which often forces indirect, quantile regression-based approximations. Overcoming this obstacle is crucial for developing more direct and reliable algorithms for superquantile estimation.

A breakthrough came with the CVaR Quadrangle developed in Rockafellar and Royset [2014]; Rockafellar et al. [2014], which underpins new methodologies for CVaR estimation. In particular, CVaR regression functions emerge naturally as the optimal solutions to the CVaR2 error minimization problem. This fundamental result circumvents the non-elicitability issue without relying on indirect approaches.

The CVaR-based quadrangle is regular.

CVaR Quadrangle

$$
\begin{aligned}
\mathcal{S}(X) &= \text{CVaR}_\alpha(X) = \text{CVaR}\\
\mathcal{R}(X) &= \frac{1}{1-\alpha}\int_\alpha^1 \text{CVaR}_\beta(X)d\beta = \text{CVaR2 risk}\\
\mathcal{D}(X) &= \frac{1}{1-\alpha}\int_\alpha^1 \text{CVaR}_\beta(X)d\beta - \mathbb{E}[X] = \text{CVaR2 deviation}\\
\mathcal{V}(X) &= \frac{1}{1-\alpha}\int_0^1 [\text{CVaR}_\beta(X)]_+ d\beta = \text{CVaR2 regret}\\
\mathcal{E}(X) &= \frac{1}{1-\alpha}\int_0^1 [\text{CVaR}_\beta(X)]_+ d\beta - \mathbb{E}[X] = \text{CVaR2 error}
\end{aligned}
$$

Example 4 (Quantile Symmetric Average Quadrangle (CVaR Norm Quadrangle)[1], $\alpha \in (0,1)$)**.** Addressing the limitations of classical $\mathcal{L}^p$-norms in functional analysis, statistical estimation, and machine learning has driven interest in alternative norms capable of quantifying the risk of rare events. One promising candidate is the CVaR norm, which provides robust performance in various regression scenarios.

Building on this motivation, the quantile symmetric average quadrangle–developed and studied by Mafusalov and Uryasev [2016]—introduces the CVaR norm as its error function, earlier explored by Pavlikov and Uryasev

[1] Originally, this quadrangle was named "CVaR Norm Quadrangle" in Mafusalov and Uryasev [2016]. However this paper uses the term "Quantile Symmetric Average Quadrangle," following the original logic of Rockafellar and Uryasev [2013].

[2014]; Bertsimas et al. [2011]. This framework has shown noteworthy success in linear regression applications, where Malandii and Uryasev [2022] demonstrated that regression with respect to the CVaR norm is equivalent to the well-known ν-Support Vector Regression (SVR) of Schölkopf et al. [2000] and Stable Regression of Bertsimas and Paskov [2020] which is essentially a well-kown dual formulation of CVaR.

The quantile symmetric average quadrangle is regular.

Quantile Symmetric Average Quadrangle (CVaR Norm Quadrangle)

$$\begin{aligned}
\mathcal{S}(X) &= \frac{1}{2}\left(\mathrm{VaR}_{(1-\alpha)/2}(X) + \mathrm{VaR}_{(1+\alpha)/2}(X)\right)\\
\mathcal{R}(X) &= \frac{1}{2}\left((1+\alpha)\mathrm{CVaR}_{(1-\alpha)/2}(X) + (1-\alpha)\mathrm{CVaR}_{(1+\alpha)/2}(X)\right)\\
\mathcal{D}(X) &= \frac{1}{2}\left((1+\alpha)\mathrm{CVaR}_{(1-\alpha)/2}(X-\mathbb{E}[X]) + (1-\alpha)\mathrm{CVaR}_{(1+\alpha)/2}(X-\mathbb{E}[X])\right)\\
\mathcal{V}(X) &= \langle\!\langle X \rangle\!\rangle_\alpha + \mathbb{E}[X] = \text{CVaR norm regret}\\
\mathcal{E}(X) &= \mathrm{CVaR}_\alpha(|X|) = \langle\!\langle X \rangle\!\rangle_\alpha = \text{CVaR norm}
\end{aligned}$$

Example 5 (Quantile Symmetric Average Union Quadrangle, $0 \le x < \frac{1}{2}(\operatorname{ess\,sup}(X) - \operatorname{ess\,inf}(X))$)**.** Integrating machine learning tools with classical statistics and risk management has long been of interest for creating robust and versatile estimation frameworks. A notable step in this direction is the effort to embed well-known methods like ε-SVR Drucker et al. [1996]; Vapnik et al. [1996] into the Risk Quadrangle framework.

As constructed in Malandii and Uryasev [2022], the quantile symmetric average union quadrangle accomplishes this integration: the optimal solution to the regression problem with the Vapnik error emerges as an estimate of the average of two symmetric quantiles, effectively bridging the gap between traditional statistical tools and modern machine learning methods.

The quantile symmetric average union quadrangle is not regular. For instance, axiom (A9) fails for its error measure. Nonetheless, it is a subregular risk quadrangle (see Definition 3.11).

Define

$$\mathcal{A}_x(X) = \left\{ \alpha \in [0,1) \,\middle|\, x \in \frac{1}{2}\Big(\mathrm{VaR}_{(1+\alpha)/2}(X) - \mathrm{VaR}_{(1-\alpha)/2}(X)\Big) \right\}.$$

Then

Quantile Symmetric Average Union Quadrangle

$$\begin{aligned}
\mathcal{S}(X) &= \bigcup_{\alpha\in\mathcal{A}_x(X)} \frac{1}{2}\left(\mathrm{VaR}_{(1-\alpha)/2}(X) + \mathrm{VaR}_{(1+\alpha)/2}(X)\right)\\
\mathcal{R}(X) &= \frac{1}{2}\left((1+\alpha)\mathrm{CVaR}_{(1-\alpha)/2}(X) + (1-\alpha)\mathrm{CVaR}_{(1+\alpha)/2}(X)\right) - (1-\alpha)x,\ \forall\, \alpha \in \mathcal{A}_x(X)\\
\mathcal{D}(X) &= \frac{1}{2}\left((1+\alpha)\mathrm{CVaR}_{(1-\alpha)/2}(X) + (1-\alpha)\mathrm{CVaR}_{(1+\alpha)/2}(X)\right) - \mathbb{E}[X] - (1-\alpha)x,\ \forall\, \alpha \in \mathcal{A}_x(X)\\
\mathcal{V}(X) &= \mathbb{E}[|X|-x]_+ + \mathbb{E}[X] = \text{Vapnik regret}\\
\mathcal{E}(X) &= \mathbb{E}[|X|-x]_+ = \text{Vapnik error}
\end{aligned}$$

Example 6 (Expectile-based Quadrangle (Asymmetric Mean Squared Error Version), $q \in (0,1)$)**.** The *expectile* of a random variable X at confidence level $q \in (0,1)$ is defined as (cf. Newey and Powell [1987])

$$e_q(X) = \operatorname*{argmin}_{C\in\mathbb{R}} \left\{ \mathcal{E}(X-C) \right\}, \tag{1}$$

where $\mathcal{E}(X) = \mathbb{E}[qX_+^2 + (1-q)X_-^2]$ is the *asymmetric mean squared error*. The case $q = 0.5$ gives the usual mean value.

The expectile-based quadrangle (asymmetric mean squared error version) was introduced in Kuzmenko [2020] and studied in Malandii et al. [2024]. This quadrangle generalizes the mean quadrangle introduced in Rockafellar and Uryasev [2013]. *The expectile-based quadrangle (asymmetric mean squared error version) is regular.*

Expectile-based Quadrangle (Asymmetric Mean Squared Error Version)
$\mathcal{S}(X) = e_q(X) = \text{expectile}$ $\mathcal{R}(X) = \mathcal{D}(X) + \mathbb{E}[X] = \text{asymmetric risk}$ $\mathcal{D}(X) = q\mathbb{E}\left[((X - e_q(X))_+)^2\right] + (1-q)\mathbb{E}\left[((X - e_q(X))_-)^2\right] = \text{asymmetric variance}$ $\mathcal{V}(X) = \mathcal{E}(X) + \mathbb{E}[X] = \text{asymmetric regret}$ $\mathcal{E}(X) = \mathbb{E}[qX_+^2 + (1-q)X_-^2] = \text{asymmetric mean squared error}$

Example 7 (Expectile-based Quadrangle (Piecewise Linear Version), $K > 0$)**.** Despite growing interest in expectiles for risk management and statistical estimation, there has been no unified framework for both efficient optimization and statistical estimation.

Expectile can also be defined by the necessary condition of extremum for (1) as a solution to the equation

$$q\mathbb{E}[(X - C)_+] = (1-q)\mathbb{E}[(X - C)_-]. \tag{2}$$

With formula $\mathbb{E}[X - C] = \mathbb{E}[(X - C)_+] - \mathbb{E}[(X - C)_-]$ equation (2) is equivalently transformed to

$$C - \mathbb{E}[X] = \frac{1}{K}\mathbb{E}[(X - C)_+], \tag{3}$$

where $K = \frac{1-q}{2q-1}$. A one-to-one correspondence exists between values $K > 0$ and values q in the interval $1/2 < q < 1$. Also, there is a one-to-one correspondence between values K in the interval $K < -1$ and values q in the interval $0 < q < 1/2$.

The equation (3) should be considered separately for $K > 0$ $(1/2 < q < 1)$ and for $K < -1$ $(0 < q < 1/2)$ because properties of expectile as a function of the parameter q change at the point $q = 1/2$.

For the range $K > 0$ $(1/2 < q < 1)$, expectile is a coherent risk measure in the basic sense (see Shapiro [2012]), i.e., it is translation invariant, positively homogeneous, monotonic, and subadditive.

The expectile-based quadrangle (piecewise linear version) introduced in Kuzmenko [2020] and studied in Malandii et al. [2024] has the expectile, e_K, as a risk measure as well as the statistic. Having the expectile as a risk allows to leverage the [Rockafellar and Uryasev, 2013, Regret Theorem] for efficient expectile optimization. Having a piecewise linear error allows for the reduction of the expectile estimation with linear regression to a linear programming problem.

The expectile-based quadrangle (piecewise linear version) is regular.

Expectile-based Quadrangle (Piecewise Linear Version)
$\mathcal{S}(X) = e_K(X) = \text{expectile}$ $\mathcal{R}(X) = \min_C\{C + \mathcal{V}(X - C)\} = e_K(X) = \text{expectile risk}$ $\mathcal{D}(X) = \min_C\{\mathcal{E}(X - C)\} = e_K(X - \mathbb{E}[X]) = \text{expectile deviation}$ $\mathcal{V}(X) = \left(\mathbb{E}[X] + \frac{1}{K}\mathbb{E}[X_+]\right)_+ = \left(\mathbb{E}\left[X + \frac{1}{K}X_+\right]\right)_+ = \text{piecewise linear regret}$ $\mathcal{E}(X) = \mathcal{V}(X) - \mathbb{E}[X] = \max\left\{-\mathbb{E}[X], \frac{1}{K}\mathbb{E}[X_+]\right\} = \text{piecewise linear error}$

Example 8 (Mean-based Quadrangle (Piecewise Linear Version))**.** Modern risk management and statistical estimation often require alternatives to the classical mean squared error. A piecewise linear approach can offer greater flexibility and potentially more efficient computational methods, motivating the development of new frameworks that incorporate mean–upper–semideviation risk measures within standard regression tasks.

The mean-based quadrangle (piecewise linear version), introduced and studied by Malandii and Uryasev [2024], provides exactly such a framework. By pairing a mean–upper–semideviation risk measure with the statistical estimation of mean value via regression, this quadrangle replaces the classical mean squared error with a piecewise linear alternative. Crucially, the resulting linear regression can be efficiently reformulated as a linear programming problem, offering computational advantages and more robust modeling possibilities.

The mean-based quadrangle (piecewise linear version) is regular.

Mean-based Quadrangle (Piecewise Linear Version)
$\mathcal{S}(X) = \mathbb{E}[X] = \text{mean}$ $\mathcal{R}(X) = \mathbb{E}[X - \mathbb{E}[X]]_+ + \mathbb{E}[X] = \text{mean–upper–semideviation risk}$ $\mathcal{D}(X) = \mathbb{E}[X - \mathbb{E}[X]]_+ = \text{upper–semideviation}$ $\mathcal{V}(X) = \max\{\mathbb{E}[X_-], \mathbb{E}[X_+]\} + \mathbb{E}[X] = \text{mean–upper–semideviation regret}$ $\mathcal{E}(X) = \max\{\mathbb{E}[X_-], \mathbb{E}[X_+]\} = \text{mean–upper–semideviation error}$

Example 9 (Biased Mean Quadrangle, $x \in \mathbb{R}$)**.** A growing need exists for alternative regression frameworks that go beyond classical mean-based approaches, particularly in engineering and finance. The concept of superexpectation Rockafellar and Royset [2014] offers a way to capture biased mean estimates, which can be more relevant for certain practical applications. Leveraging these ideas, the biased mean quadrangle developed in Malandii and Uryasev [2024] enables conditional biased mean estimation.

Within this biased mean quadrangle, the statistic is the biased mean $\mathcal{S}(X) = x + \mathbb{E}[X]$, for any $x \in \mathbb{R}$. The superexpectation error plays a central role in estimating the conditional biased mean via regression, creating a wide range of potential engineering and financial applications. Notably, this approach to regression is equivalent to quantile regression; rather than specifying a confidence level $\alpha \in (0,1)$, one determines the desired quantile by choosing a distance $x \in \mathbb{R}$ from the mean. Furthermore, the mean-based quadrangle (piecewise linear version) emerges as a special case of this more general quadrangle.

The biased mean quadrangle is not regular. However, it is a subregular risk quadrangle, see Definition 3.11.

Biased Mean Quadrangle
$\mathcal{S}(X) = x + \mathbb{E}[X] =$ biased mean $\mathcal{R}(X) = \mathbb{E}[X - \mathbb{E}[X] - x]_+ - x_- + \mathbb{E}[X] =$ superexpectation risk $\mathcal{D}(X) = \mathbb{E}[X - \mathbb{E}[X] - x]_+ - x_- =$ superexpectation deviation $\mathcal{V}(X) = \max\{\mathbb{E}[X_-] - x_+, \mathbb{E}[X_+] - x_-\} + \mathbb{E}[X] =$ superexpectation regret $\mathcal{E}(X) = \max\{\mathbb{E}[X_-] - x_+, \mathbb{E}[X_+] - x_-\} =$ superexpectation error

Example 10 (φ–Divergence-based Quadrangle, $\beta > 0$)**.** The φ–divergence-based quadrangle introduced and studied in Peng et al. [2024] is based upon the concept of distributionally robust risk measures studied by Shapiro [2017]; Dommel and Pichler [2020]. The function φ here is a so-called extended divergence function, i.e., a convex lower semi-continuous function $\varphi : \mathbb{R} \to (\infty, \infty]$ satisfying[2]

$$\varphi(1) = 0, \quad 1 \in \operatorname{int}(\{x : \varphi(x) < +\infty\})\,. \tag{4}$$

In what follows φ^* denotes the convex conjugate function of φ. This quadrangle provides an interpretation of portfolio optimization, classification, and regression as robust (or distributionally robust) optimization. *The φ–divergence-based quadrangle is regular.*

φ–Divergence-based Quadrangle
$\mathcal{R}_{\varphi,\beta}(X) = \inf\limits_{C\in\mathbb{R},\, t>0} t\left\{C + \beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right)\right]\right\}$ $\mathcal{D}_{\varphi,\beta}(X) = \inf\limits_{C\in\mathbb{R},\, t>0} t\left\{C + \beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right) - \frac{X}{t}\right]\right\}$ $\mathcal{V}_{\varphi,\beta}(X) = \inf\limits_{t>0} t\left\{\beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right)\right]\right\}$ $\mathcal{E}_{\varphi,\beta}(X) = \inf\limits_{t>0} t\left\{\beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right) - \frac{X}{t}\right]\right\}$ $\mathcal{S}_{\varphi,\beta}(X) = \operatorname{argmin}\limits_{C\in\mathbb{R}} \inf\limits_{t>0} t\left\{\frac{C}{t} + \beta + \mathbb{E}\left[\varphi^*\left(\frac{X-C}{t}\right) - \frac{X}{t}\right]\right\}$

The following examples are specific instances of the φ-divergence-based quadrangle for different divergence functions φ.

Example 10.1 (Kullback–Leibler Divergence-based Quadrangle, $\alpha \in (0,1)$). The Kullback–Leibler divergence-based quadrangle is built upon the entropic value-at-risk (EVaR) introduced and studied in Ahmadi-Javid [2012]. The divergence function and its convex conjugate are

$$\varphi(x) = x\ln(x) - x + 1, \quad \varphi^*(z) = \exp(z) - 1.$$

Let $\beta = \ln\left(\frac{1}{1-\alpha}\right)$. The complete quadrangle is as follows:

[2]We call φ a divergence function if it additionally satisfies $\varphi(x) = +\infty$ for $x < 0$.

Kullback–Liebler Divergence-based Quadrangle

$$\mathcal{R}_{\varphi,\alpha}(X) = \text{EVaR}_\alpha(X) = \inf_{t>0} t\left\{ \ln \mathbb{E}\Big[\frac{e^{\frac{X}{t}}}{1-\alpha}\Big]\right\}$$

$$\mathcal{D}_{\varphi,\alpha}(X) = \text{EVaR}_\alpha(X) - \mathbb{E}[X] = \inf_{t>0} t\left\{ \ln \mathbb{E}\Big[\frac{e^{\frac{X-\mathbb{E}[X]}{t}}}{1-\alpha}\Big]\right\}$$

$$\mathcal{V}_{\varphi,\alpha}(X) = \inf_{t>0} t\left\{ \ln\Big(\frac{1}{1-\alpha}\Big) + \mathbb{E}\big[e^{\frac{X}{t}} - 1\big]\right\}$$

$$\mathcal{E}_{\varphi,\alpha}(X) = \inf_{t>0} t\left\{ \ln\Big(\frac{1}{1-\alpha}\Big) + \mathbb{E}\Big[e^{\frac{X}{t}} - \frac{X}{t} - 1\Big]\right\}$$

$$\mathcal{S}_{\varphi,\alpha}(X) = t^* \ln \mathbb{E}\big[e^{\frac{X}{t^*}}\big]$$

In the quadrangle, $t^* = t^*(X)$ is a solution of the following equation:

$$t^* \ln\Big(\frac{1}{1-\alpha}\Big) + t^* \ln \mathbb{E}\big[e^{\frac{X}{t^*}}\big] - \frac{\mathbb{E}\big[Xe^{\frac{X}{t^*}}\big]}{\mathbb{E}\big[e^{\frac{X}{t^*}}\big]} = 0.$$

Example 10.2 (Total Variation Divergence-based Quadrangle). The total variation divergence-based quadrangle relies on Example 3.10 of Shapiro [2017], where the derivation of the risk measure was carried out. Consider the following divergence function and its convex conjugate

$$\varphi(x) = \begin{cases} |x-1|, & x \geq 0 \\ +\infty, & x < 0 \end{cases} \quad \text{and} \quad \varphi^*(z) = \begin{cases} -1 + [z+1]_+, & z \leq 1 \\ +\infty, & z > 1 \end{cases} .$$

The complete quadrangle is as follows:

Total Variation Divergence-based Quadrangle

$$\mathcal{R}_{\varphi,\beta}(X) = \frac{\beta}{2}\text{ess sup}(X) + (1-\frac{\beta}{2})\text{CVaR}_{\frac{\beta}{2}}(X)$$

$$\mathcal{V}_{\varphi,\beta}(X) = \inf_{t>0, t\geq \text{ess sup} X} \left\{ t(\beta-1) + \mathbb{E}\Big[X+t\Big]_+ \right\}$$

$$\mathcal{D}_{\varphi,\beta}(X) = \frac{\beta}{2}\text{ess sup}(X) + (1-\frac{\beta}{2})\text{CVaR}_{\frac{\beta}{2}}(X) - \mathbb{E}[X]$$

$$\mathcal{E}_{\varphi,\beta}(X) = \inf_{t>0} \left\{ t(\beta-1) + \mathbb{E}\Big[\Big[X+t\Big]_+ - X\Big] \right\}$$

$$\mathcal{S}_{\varphi,\beta}(X) = \text{ess sup}(X) - 2\text{VaR}_{1-\frac{\beta}{2}}(X)$$

Example 10.3 (Pearson Divergence-based Quadrangle). The Pearson divergence-based quadrangle is a special case of the higher-order quantile-based quadrangle in Example 12 of Rockafellar and Uryasev [2013]. The second-order superquantile risk measure from this quadrangle was introduced and studied by Krokhmal [2007]. The divergence function and its convex conjugate are

$$\varphi(x) = \begin{cases} (x-1)^2, & x \geq 0 \\ +\infty, & x < 0 \end{cases} \quad \text{and} \quad \varphi^*(z) = \begin{cases} \frac{(z+2)^2}{4} - 1, & z+2 \geq 0 \\ -1, & z+2 < 0 \end{cases} = \frac{1}{4}\left[z+2\right]_+^2 - 1 .$$

The complete quadrangle is

Pearson Divergence-based Quadrangle

$$
\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \min_{C\in\mathbb{R}}\left(\sqrt{(\beta+1)\mathbb{E}\left[(X-C)_+^2\right]}+C\right) = \text{ second-order superquantile}\\
\mathcal{V}_{\varphi,\beta}(X) &= \sqrt{(\beta+1)\mathbb{E}\left[X_+^2\right]} = \mathcal{L}^2\text{–normed absolute loss, scaled}\\
\mathcal{D}_{\varphi,\beta}(X) &= \min_{C\in\mathbb{R}}\left(\sqrt{(\beta+1)\mathbb{E}\left[(X-C)_+^2\right]}-\mathbb{E}[X-C]\right) = \text{ second–order superquantile deviation}\\
\mathcal{E}_{\varphi,\beta}(X) &= \sqrt{(\beta+1)\mathbb{E}\left[X_+^2\right]}-\mathbb{E}[X] = \text{ second–order quantile error}\\
\mathcal{S}_{\varphi,\beta}(X) &= \underset{C\in\mathbb{R}}{\text{argmin}}\left(\sqrt{(\beta+1)\mathbb{E}\left[(X-C)_+^2\right]}-\mathbb{E}[X-C]\right) = \text{ second–order quantile}
\end{aligned}
$$

Example 10.4 (Extended Pearson Divergence-based Quadrangle). Consider the following extended divergence function and its convex conjugate

$$
\varphi(x)=(x-1)^2 \quad \text{and} \quad \varphi^*(z)=\frac{z^2}{4}+z\,.
$$

Then, the extended Pearson χ^2-divergence risk measure is given by

$$
\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \inf_{t>0,C\in\mathbb{R}} t\left\{C+\beta+\frac{1}{4t^2}\mathbb{E}[(X-C)^2]+\mathbb{E}\left[\frac{X-C}{t}\right]\right\}\\
&= \inf_{t>0,C\in\mathbb{R}}\left\{t\beta+\frac{1}{4t}\mathbb{E}[(X-C)^2]+\mathbb{E}[X]\right\}\\
&= \mathbb{E}[X]+\sqrt{\beta\mathbb{V}[X]},
\end{aligned}
$$

where $\mathbb{V}[X]=\mathbb{E}[(X-\mathbb{E}[X])^2]$ is the variance of X and (t^*,C^*), which furnish the minimum are

$$
t^*=\sqrt{\frac{\mathbb{V}[X]}{4\beta}}, \qquad C^*=\mathbb{E}[X].
$$

The corresponding regret is given by

$$
\begin{aligned}
\mathcal{V}_{\varphi,\beta}(X) &= \mathbb{E}[X]+\sqrt{\beta\mathbb{E}[X^2]}\\
&= \mathbb{E}[X]+\sqrt{\beta}\,\|X\|_2\,.
\end{aligned}
$$

Let $\tau=\sqrt{\beta}$ and $\sigma(X)=\sqrt{\mathbb{V}[X]}$, then the complete quadrangle is

Extended Pearson Divergence-based Quarangle

$$
\begin{aligned}
\mathcal{R}_{\varphi,\tau}(X) &= \mathbb{E}[X]+\tau\sigma(X) = \text{ safety margin tail risk}\\
\mathcal{V}_{\varphi,\tau}(X) &= \mathbb{E}[X]+\tau\,\|X\|_2 = \mathcal{L}^2\text{–regret, scaled}\\
\mathcal{D}_{\varphi,\tau}(X) &= \tau\sigma(X) = \text{ standard deviation, scaled}\\
\mathcal{E}_{\varphi,\tau}(X) &= \tau\,\|X\|_2 = \mathcal{L}^2\text{–error, scaled}\\
\mathcal{S}_{\varphi,\tau}(X) &= \mathbb{E}[X] = \text{ mean}
\end{aligned}
$$

exactly the mean quadrangle in Example 1. The divergence function is the extended version of the divergence function of χ^2–divergence. It is worth noting that the radius β of the uncertainty set does not impact the regression result, since it only impacts the scale of the error function.

Example 10.5 (Generalized Extended Pearson Divergence-based Quadrangle). Let $0<q<1$. Consider the following extended divergence function and its convex conjugate

$$
\varphi(x)=\begin{cases}\frac{1}{q}(x-1)^2, & x>1\\ \frac{1}{1-q}(x-1)^2, & x\le 1\end{cases} \quad \text{and} \quad \varphi^*(z)=\begin{cases}(\frac{qz}{2}+1)z-\frac{1}{q}(\frac{qz}{2})^2=\frac{qz^2}{4}+z, & z>0\\ (\frac{(1-q)z}{2}+1)z-\frac{1}{1-q}(\frac{(1-q)z}{2})^2=\frac{(1-q)z^2}{4}+z, & z\le 0\end{cases}.
$$

The error measure is given by

$$\begin{aligned}
\mathcal{E}_{\varphi,\beta}(X) &= \inf_{t>0}\ t\beta + \mathbb{E}\left[t\varphi^*\left(\frac{X}{t}\right) - X\right] \\
&= \inf_{t>0}\ t\beta + \frac{1}{4t}\mathbb{E}\left[qX_-^2 + (1-q)X_+^2\right] \\
&= t\beta + \frac{1}{4t}\mathbb{E}\left[qX_-^2 + (1-q)X_+^2\right]\Big|_{t=\sqrt{\frac{\mathbb{E}\left[qX_-^2+(1-q)X_+^2\right]}{4\beta}}} \\
&= \sqrt{\beta\mathbb{E}\left[qX_-^2 + (1-q)X_+^2\right]}.
\end{aligned}$$

Thus, the complete quadrangle is as follows

Generalized Extended Pearson Divergence-based Quadrangle
$\mathcal{R}_{\varphi,\beta}(X) = \mathcal{D}_{\varphi,\beta}(X) + \mathbb{E}[X]$
$\mathcal{V}_{\varphi,\beta}(X) = \mathcal{E}_{\varphi,\beta}(X) + \mathbb{E}[X]$
$\mathcal{D}_{\varphi,\beta}(X) = \sqrt{\beta\left(q\mathbb{E}[(X-e_q(X))_+^2] + (1-q)\mathbb{E}[(X-e_q(X))_-^2]\right)}$
$\mathcal{E}_{\varphi,\beta}(X) = \sqrt{\beta\mathbb{E}\left[qX_-^2 + (1-q)X_+^2\right]} =$ asymmetric $\mathcal{L}^2$–error, scaled
$\mathcal{S}_{\varphi,\beta}(X) = e_q(X) =$ expectile

Therefore, we recover the Expectile-based Quadrangle (Asymmetric $\mathcal{L}^2$–error version) introduced in Kuzmenko [2020]; Malandii et al. [2024]. The divergence function $\varphi(x)$ gives rise to a generalized Pearson χ^2–divergence.

3 Theoretical Framework

This section provides a theoretical foundation for defining, constructing, and utilizing risk quadrangles.

3.1 Definitions and Axioms

This subsection introduces the central definitions and axioms required for further development. We first fix the functional space setting and define regular measures of risk, deviation, regret, and error introduced in Rockafellar and Uryasev [2013] and refined in Rockafellar and Royset [2015].

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space, where Ω is a set of elementary outcomes, $\mathcal{M} \subseteq 2^\Omega$ is a σ-algebra of subsets of Ω, and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{M})$. A random variable (r.v.) $X : \Omega \to \mathbb{R}$ is a measurable function defined on Ω taking values in $\mathbb{R}$, defined up to the set of measure 0, that is, r.v. X and Y such that $\mathbb{P}(X = Y) = 1$ will be identified. The mathematical expectation of an r.v. X is defined by $\mathbb{E}[X] = \int_\Omega X d\mathbb{P}$.

For $p = [1, \infty]$, let $\mathcal{L}^p(\Omega) = \mathcal{L}^p(\Omega, \mathcal{M}, \mathbb{P})$ be a normed space of all r.v.'s X with $\|X\|_p < \infty$, where $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$ for $p < \infty$, and $\|X\|_\infty = \operatorname{ess\,sup}|X|$. A functional $\mathcal{F} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called *convex* if

$$\mathcal{F}(\lambda X + (1-\lambda)Y) \le \lambda\mathcal{F}(X) + (1-\lambda)\mathcal{F}(Y) \text{ for all } X, Y \in \mathcal{L}^p(\Omega) \text{ and every } \lambda \in [0, 1]$$

and *closed (or lower-semicontinuous)* if

$$\{X : \mathcal{F}(X) \le C\} \text{ is a closed set in } \mathcal{L}^p(\Omega) \text{ for every constant } C.$$

Definition 3.1 (Regular risk)**.** A closed convex functional $\mathcal{R} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called a *regular measure of risk* if

(R0) $\mathcal{R}(C) = C$ for constants C and $\mathcal{R}(X) > \mathbb{E}X$ for nonconstant X.

Definition 3.2 (Coherent risk)**.** A closed convex functional $\mathcal{R} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called a *coherent measure of risk* in the basic sense if

(C0) $\mathcal{R}(C) = C$ for constants C;

(C1) $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ for all $\lambda > 0$;

(C2) $\mathcal{R}(X) \le \mathcal{R}(Y)$ for all X and Y such that $X \le Y$ almost surely,

and coherent in the general sense if (C1) is left out.

Definition 3.3 (Regular deviation)**.** A closed convex functional $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$ is called a *regular measure of deviation* if

(D0) $\mathcal{D}(C) = 0$ for constants C and $\mathcal{D}(X) > 0$ for nonconstant X.

Definition 3.4 (Regular regret)**.** A closed convex functional $\mathcal{V} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called a *regular measure of regret* if

(V0) $\mathcal{V}(0) = 0$ and $\mathcal{V}(X) > \mathbb{E}X$ for nonconstant X.

Definition 3.5 (Regular error)**.** A closed convex functional $\mathcal{E} : \mathcal{L}^p(\Omega) \to [0, \infty]$ is called a *regular measure of error* if

(E0) $\mathcal{E}(0) = 0$ and $\mathcal{E}(X) > 0$ for $X \not\equiv 0$.

Definition 3.6 (Regular quadrangle)**.** A quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ of regular risk, deviation, regret, and error is called a *regular risk quadrangle* if it satisfies the following relationship formulae:

(Q1) **error projection:** $\mathcal{D}(X) = \min_C \{ \mathcal{E}(X - C) \}$;

(Q2) **regret formula:** $\mathcal{R}(X) = \min_C \{ C + \mathcal{V}(X - C) \}$;

(Q3) **mean-centering:**

$$\mathcal{R}(X) = \mathcal{D}(X) + \mathbb{E}[X] \quad \text{and} \quad \mathcal{D}(X) = \mathcal{R}(X) - \mathbb{E}[X] \tag{5}$$

$$\mathcal{V}(X) = \mathcal{E}(X) + \mathbb{E}[X] \quad \text{and} \quad \mathcal{E}(X) = \mathcal{V}(X) - \mathbb{E}[X]; \tag{6}$$

where the argmin in (Q1) and the argmin in (Q2) coincide, i.e.,

(Q4) **statistic:** $\mathcal{S}(X) = \text{argmin}_C \{ \mathcal{E}(X - C) \} = \text{argmin}_C \{ C + \mathcal{V}(X - C) \}$

and called a *statistic.*

Remark 3.1. (Coherent quadrangle) A quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ satisfying (Q1)–(Q4), where $\mathcal{R}$ is a coherent measure of risk is called a *coherent risk quadrangle.*

The axioms listed above are too restrictive to accommodate certain examples commonly used in the literature. One key example is the Vapnik error, defined by

$$\mathcal{E}(X) = \mathbb{E}[|X| - x]_+, \tag{7}$$

where $[X]_+ = \max(X, 0)$, and $x \geq 0$ is a parameter. The idea behind the Vapnik error is to disregard errors whose absolute values are below x while penalizing all errors exceeding that threshold. For any $x > 0$, Vapnik error (7) does not satisfy (E0), because $\mathcal{E}(X) = 0$ for r.v. X such that $X = x$ almost surely.

This observation motivates a relaxation of the axioms listed above and leads to the following definitions.

Definition 3.7 (Subregular risk)**.** A closed convex functional $\mathcal{R} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called a *subregular risk measure* if

(R1) $\mathcal{R}(C) = C$ for constants C and $\mathcal{R}(X) \geq \mathbb{E}[X]$ for all X;

(R2) for all non-constant $X \in \mathcal{L}^p(\Omega)$ there exists $\lambda > 0$ such that $\mathcal{R}(\lambda X) > \mathbb{E}(\lambda X)$.

Definition 3.8 (Subregular deviation)**.** A closed convex functional $\mathcal{D} : \mathcal{L}^p(\Omega) \to [0, \infty]$ is called a *subregular deviation measure* if

(D1) $\mathcal{D}(C) = 0$ for constants C and $\mathcal{D}(X) \geq 0$ for all X;

(D2) for all non-constant $X \in \mathcal{L}^p(\Omega)$ there exists $\lambda > 0$ such that $\mathcal{D}(\lambda X) > 0$.

Definition 3.9 (Subregular regret)**.** A closed convex functional $\mathcal{V} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called a *subregular regret measure* if

(V1) $V(0) = 0$ and $\mathcal{V}(X) \geq \mathbb{E}[X]$ for all X;

(V2) for all non-zero $X \in \mathcal{L}^p(\Omega)$ there exists $\lambda > 0$ such that $\mathcal{V}(\lambda X) > \mathbb{E}(\lambda X)$.

Definition 3.10 (Subregular error)**.** A closed convex functional $\mathcal{E} : \mathcal{L}^p(\Omega) \to [0, \infty]$ is called a *subregular error measure* if

(E1) $\mathcal{E}(0) = 0$ and $\mathcal{E}(X) \geq 0$ for all X;

(E2) for all non-zero $X \in \mathcal{L}^p(\Omega)$ there exists $\lambda > 0$ such that $\mathcal{E}(\lambda X) > 0$.

Definition 3.11 (Subregular quadrangle)**.** A quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ of subregular risk, deviation, regret, and error satisfying (Q1)–(Q4) is called a *subregular risk quadrangle.*

Axioms (E1)–(E2) relax axiom (E0) by allowing $\mathcal{E}(X) = 0$ for some nonzero but "not-too-large" random variables X. For example, the Vapnik error (7) satisfies (E1)–(E2) and thus qualifies as a subregular error measure. Similarly, axioms (V1)–(V2), (D1)–(D2), and (R1)–(R2) relax axioms (V0), (D0), and (R0), respectively.

In the following section, we will extend the theory developed in Rockafellar and Uryasev [2013]; Rockafellar and Royset [2015] for measures of error, regret, deviation, and risk to the more general subregular measures introduced in Definitions 3.7–3.10. In particular, we will show that relations (Q1)–(Q4) still hold in this broader framework. In fact, relations (Q3) are straightforward to verify, while (Q2) is equivalent to (Q1) under (Q3), so we only need to prove (Q1).

3.2 Fundamental Theorems and Construction of Quadrangles

This section states and proves the main theorems of the RQ framework. These theorems establish the relationships among the quadrangle elements and provide a foundation for constructing RQs.

Subsection 3.2.1 discusses the theory in the primal space, i.e., the space where the stochastic functionals are defined. Meanwhile, Subsection 3.2.2 focuses on the dual space and the aspects of conjugate duality.

3.2.1 Primal Representation

Rockafellar and Uryasev [2013] presented seven key theorems – cornerstones of the RQ framework: *Quadrangle Theorem, Scaling Theorem, Mixing Theorem, Reverting Theorem, Expectation Theorem, Regret Theorem, and Convexity Theorem.* These theorems hold for regular quadrangles and can be found in Rockafellar and Uryasev [2013].

This section revisits the aforementioned theorems for subregular quadrangles and provides proofs where the extension to subregularity is non-trivial.

We begin with a central theorem of the RQ framework:

Theorem 3.1 (Quadrangle Theorem). *Let $(\mathcal{R}, \mathcal{V}, \mathcal{D}, \mathcal{E})$ be a subregular risk quadrangle with a statistic $\mathcal{S}$. Then*

(a) *The relations $\mathcal{D}(X) = \mathcal{R}(X) - \mathbb{E}X$ and $\mathcal{R}(X) = \mathbb{E}X + \mathcal{D}(X)$ give a one-to-one correspondence between subregular measures of risk $\mathcal{R}$ and subregular 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,*

$$\mathcal{R} \text{ is monotonic if and only if } \mathcal{D}(X) \le \sup X - \mathbb{E}X \text{ for all } X. \tag{8}$$

(b) *The relations $\mathcal{E}(X) = \mathcal{V}(X) - \mathbb{E}X$ and $\mathcal{V}(X) = \mathbb{E}X + \mathcal{E}(X)$ give a one-to-one correspondence between subregular measures of regret $\mathcal{V}$ and subregular 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,*

$$\mathcal{V} \text{ is monotonic if and only if } \mathcal{E}(X) \le |\mathbb{E}X| \text{ for } X \le 0. \tag{9}$$

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

$$\mathcal{R}(X) = \inf_C \Big\{ C + \mathcal{V}(X - C) \Big\}. \tag{10}$$

If $\mathcal{V}$ is positively homogeneous, $\mathcal{R}$ is positively homogeneous. If $\mathcal{V}$ is monotonic, $\mathcal{R}$ is monotonic.

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

$$\mathcal{D}(X) = \inf_C \Big\{ \mathcal{E}(X - C) \Big\}. \tag{11}$$

If $\mathcal{E}$ is positively homogeneous, $\mathcal{D}$ is positively homogeneous. If $\mathcal{E}$ satisfies the condition in (9), then $\mathcal{D}$ satisfies the condition in (8).

(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.*[3] *Moreover, when $\mathcal{V}$ and $\mathcal{E}$ are paired as in (b), the interval comes out the same and gives the associated statistic:*

$$\operatorname*{argmin}_C \{ C + \mathcal{V}(X - C) \} = \mathcal{S}(X) = \operatorname*{argmin}_C \{ \mathcal{E}(X - C) \}, \quad \text{with } \mathcal{S}(X + C) = \mathcal{S}(X) + C. \tag{12}$$

Proof. Parts (a) and (b) are easy to check, while part (c) follows easily from (d). Hence, we only need to prove parts (d) and (e). To do this, we first note that Lemma 2.1 in Rockafellar and Royset [2015] remains valid for subregular error measures.

Proposition 3.1. For a subregular error measure $\mathcal{E} : \mathcal{L}^p(\Omega) \to [0, \infty]$ and a sequence $\{C_n\}_{n=1}^{\infty}$ of scalars, the following holds: If sequences $X_n \in \mathcal{L}^p(\Omega)$ and $b_n \in \mathbb{R}$ converge to $X \in \mathcal{L}^p(\Omega)$ and $b \in \mathbb{R}$ respectively, and $\mathcal{E}(X_n - C_n) \le b_n$ for all n, then sequence $\{C_n\}_{n=1}^{\infty}$ is bounded, and any accumulation point C_0 satisfies $\mathcal{E}(X - C_0) \le b$.

[3] Typically, this interval reduces to a single point.

Proof. By applying property (E2) to $X \equiv -1$ and $X \equiv 1$, we conclude that

(E2') there exist constants $K_1 < 0 < K_2$ such that $\mathcal{E}(K_1) > 0$ and $\mathcal{E}(K_2) > 0$.

We will prove the statement of the Proposition for any closed convex functional $\mathcal{E} : \mathcal{L}^p(\Omega) \to [0, \infty]$ satisfying (E1) and (E2'). By contradiction, assume that sequence $\{C_n\}_{n=1}^\infty$ is unbounded. By passing to a subsequence if necessary, we may assume that all C_n have the same sign (say, positive). Then we may also assume that $C_n \geq |K_1|$ for all n, and $\lim\limits_{n\to\infty} C_n = \infty$. Then sequence $\lambda_n = -K_1/C_n, n = 1, 2, \dots$ is contained in $[0, 1]$ and converges to 0. Because $\mathcal{E}$ is convex and $\mathcal{E}(0) = 0$, we have $\mathcal{E}(\lambda Y) \leq \lambda\mathcal{E}(Y)$ for all $Y \in \mathcal{L}^p(\Omega)$ and all $\lambda \in [0, 1]$. Hence,

$$\lambda_n b_n \geq \lambda_n \mathcal{E}(X_n - C_n) \geq \mathcal{E}(\lambda_n X_n + K_1) \geq 0.$$

Because $\lim\limits_{n\to\infty} \lambda_n b_n = \lim\limits_{n\to\infty} \lambda_n \lim\limits_{n\to\infty} b_n = 0 \cdot b = 0$ and $\lim\limits_{n\to\infty}(\lambda_n X_n + K_1) = 0 \cdot X + K_1 = K_1$, the closedness of $\mathcal{E}$ implies that $\mathcal{E}(K_1) = 0$, but this is a contradiction with (E2').

Hence, the sequence $\{C_n\}_{n=1}^\infty$ is bounded. This implies that it must have accumulation points. If C_0 is any such point, then, by passing to a subsequence if necessary, we may assume that $\lim\limits_{n\to\infty} C_n = C_0$. Then the closedness of $\mathcal{E}$ implies that

$$\mathcal{E}(X - C_0) = \mathcal{E}\left(\lim_{n\to\infty}(X_n - C_n)\right) \leq \limsup_{n\to\infty} \mathcal{E}(X_n - C_n) \leq \lim_{n\to\infty} b_n = b.$$

□

Now, we prove the statements (d) and (e).

The proof is similar to the proof of Theorem 2.2 in Rockafellar and Royset [2015]. Let us first prove that "inf" in (11) is always attained. If $\mathcal{D}(X) = \infty$, it is attained for all C. If $\mathcal{D}(X) < \infty$, then there exists a sequence $\{C_n\}_{n=1}^\infty$ such that $\lim\limits_{n\to\infty} \mathcal{E}(X - C_n) = D(X)$. Applying Proposition 3.1 with $X_n = X$, $b_n = \mathcal{E}(X - C_n)$ and $b = \mathcal{D}(X)$, we obtain that sequence C_n is bounded and has an accumulation point C_0 for which we have $\mathcal{E}(X - C_0) \leq b$. On the other hand, $\mathcal{E}(X - C_0) \geq \inf_C \mathcal{E}(X - C) = \mathcal{D}(X) = b$, hence equality holds. Thus, "inf" in (11) is always attained, and the set of minimizers $S(X)$ is a non-empty set. Because $\mathcal{E}(X - C)$ is a convex closed function of C, $S(X)$ is a convex and closed subset of $\mathbb{R}$, and therefore is a non-empty closed interval. For any sequence $\{C_n\}_{n=1}^\infty \subseteq S(X)$ we have $\mathcal{E}(X - C_n) = \mathcal{D}(X)$, hence, by Proposition 3.1 with $X_n = X$ and $b_n = \mathcal{D}(X)$ we conclude that $\{C_n\}_{n=1}^\infty$ is bounded. Thus, $S(X)$ is bounded.

We next prove that $\mathcal{D}$ in (11) is a subregular deviation measure. Because "inf" in (11) is always attained, for any $X, Y \in \mathcal{L}^p(\Omega)$ we have $\mathcal{D}(X) = \mathcal{E}(X - C_X)$ and $\mathcal{D}(Y) = \mathcal{E}(X - C_Y)$ for some constants C_X, C_Y. Then, for any $\lambda \in [0, 1]$,

$$\lambda\mathcal{D}(X) + (1-\lambda)\mathcal{D}(Y) = \lambda\mathcal{E}(X - C_X) + (1-\lambda)\mathcal{E}(Y - C_Y) \geq \mathcal{E}(\lambda X + (1-\lambda)Y - (\lambda C_X + (1-\lambda)C_Y))$$

$$\geq \inf_C \mathcal{E}(\lambda X + (1-\lambda)Y - C) = \mathcal{D}(\lambda X + (1-\lambda)Y),$$

which implies that $\mathcal{D}$ is a convex function. We next prove its closedness. Assume that X_n is a sequence converging to X and $\mathcal{D}(X_n) \leq b < \infty$ for all n. Because "inf" in (11) is always attained, there exists constants C_n such that $\mathcal{E}(X_n - C_n) = \mathcal{D}(X_n) \leq b$. Then by Proposition 3.1 with $b_n = b$, there exists a constant C_0 such that $\mathcal{E}(X - C_0) \leq b$. Then $\mathcal{D}(X) = \inf_C \mathcal{E}(X - C) \leq \mathcal{E}(X - C_0) \leq b$, which proves the closedness of $\mathcal{D}$.

Property (D1) of $\mathcal{D}$ follows immediately from property (E1) of $\mathcal{E}$, so it is left to prove (D2). For any non-zero $Y \in \mathcal{L}^p(\Omega)$, let

$$f(Y) = \sup\{\lambda \geq 0 \,|\, \mathcal{E}(\lambda Y) = 0\}.$$

Properties (E1) and (E2) imply that $0 \leq f(Y) < \infty$. Further, convexity and closedness of $\mathcal{E}$ in combination with (E1) implies that $\mathcal{E}(\lambda Y) = 0$ for all $\lambda \in [0, f(Y)]$. Next, if a sequence $\{Y_n\}_{n=1}^\infty$ converges to Y and $f(Y_n) \geq b$ for all n and some constant b, then $\mathcal{E}(bY_n) = 0$ for all n, which implies that $\mathcal{E}(bY) = 0$, hence $f(Y) \geq b$. Thus, the function f is upper semi-continuous.

Now, fix any non-constant r.v. $X \in \mathcal{L}^p(\Omega)$. Let U be the set of unit vectors in $\mathbb{R}^2$. For any $u = (u_1, u_2) \in U$, let $g_X(u) := f(u_1 X - u_2)$. Because $u_1 X - u_2 \not\equiv 0$ for any $u \in U$, properties of f imply that function $g_X(u)$ is non-negative, finite, and upper-semicontinuous on U. Because U is a compact set, this implies that $B_X := \sup_{u\in U} g_X(u) < \infty$. Hence, for any constant $\mu > B_X$ and every $u = (u_1, u_2) \in U$, we have $\mu > g_X(u) = f(u_1 X - u_2)$, or equivalently $\mathcal{E}(\mu(u_1 X - u_2)) > 0$.

Now, choose any $\lambda > B_X$, any $C \in \mathbb{R}$, and consider unit vector $u = (\lambda/\mu, C/\mu)$, where $\mu = \sqrt{\lambda^2 + C^2} > B_X$. Then

$$0 < \mathcal{E}(\mu(u_1 X - u_2)) = \mathcal{E}(\mu((\lambda/\mu)X - C/\mu)) = \mathcal{E}(\lambda X - C).$$

Because $C \in \mathbb{R}$ was arbitrary, and the infimum in (11) is always attained, this implies that $\mathcal{D}(\lambda X) > 0$, and proves (D2), which completes the proof of Theorem 3.1. □

Next, we establish the following property of the set of minimizers $\mathcal{S}(X)$.

Proposition 3.2. Let $\mathcal{E}$ be a subregular error measure. Let $\{X_n\}_{n=1}^{\infty}$ be the sequence or r.v.s $X_n \in \mathcal{L}^p(\Omega)$ converging to $X \in \mathcal{L}^p(\Omega)$ such that $\mathcal{D}(X_n) \le b$ for all $n = 1, 2, \dots$ and some $b \in \mathbb{R}$, where $\mathcal{D}$ is defined in (11). Then the union

$$U = \bigcup_{n=1}^{\infty} \mathcal{S}(X_n)$$

is a bounded set.

Proof. Because $\mathcal{S}(X_n)$ is bounded for every n by item (e) of Theorem 3.1, unboundedness of U would imply the existence of an unbounded sequence $\{C_n\}_{n=1}^{\infty}$ such that $C_n \in \mathcal{S}(X_n)$ for all n. But then $\mathcal{E}(X_n - C_n) = \mathcal{D}(X_n) \le b$. Applying Proposition 3.1 with $b_n = b$, we obtain that sequence $\{C_n\}_{n=1}^{\infty}$ is bounded, which is a contradiction. $\square$

Given a set of regular quadrangles $(\mathcal{E}_i, \mathcal{V}_i, \mathcal{D}_i, \mathcal{R}_i)$, $i = 1, \dots, r$, one can construct a new quadrangle by mixing, scaling, or reverting operations (see Rockafellar and Uryasev [2013]). The following theorem justifies the application of a mixing operation to subregular quadrangles.

Theorem 3.2 (Mixing Theorem)**.** *For $k = 1, \dots, r$ let $(\mathcal{R}_k, \mathcal{D}_k, \mathcal{V}_k, \mathcal{E}_k)$ be a subregular quadrangle quartet with statistic $\mathcal{S}_k$, and consider any weights $\lambda_k > 0$ with $\lambda_1 + \dots + \lambda_r = 1$. A subregular quadrangle quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$ is given then by*

$$\begin{aligned}
\mathcal{S}(X) &= \lambda_1 \mathcal{S}_1(X) + \dots + \lambda_r \mathcal{S}_r(X),\\
\mathcal{R}(X) &= \lambda_1 \mathcal{R}_1(X) + \dots + \lambda_r \mathcal{R}_r(X),\\
\mathcal{D}(X) &= \lambda_1 \mathcal{D}_1(X) + \dots + \lambda_r \mathcal{D}_r(X),\\
\mathcal{V}(X) &= \min_{C_1,\dots,C_r} \left\{ \sum\nolimits_{k=1}^{r} \lambda_k \mathcal{V}_k(X - C_k) \,\middle|\, \sum\nolimits_{k=1}^{r} \lambda_k C_k = 0 \right\},\\
\mathcal{E}(X) &= \min_{C_1,\dots,C_r} \left\{ \sum\nolimits_{k=1}^{r} \lambda_k \mathcal{E}_k(X - C_k) \,\middle|\, \sum\nolimits_{k=1}^{r} \lambda_k C_k = 0 \right\}.
\end{aligned}$$

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.

Proof. To prove the theorem, it suffices to verify that the equation

$$\mathcal{E}(X) = \inf_{\substack{C_1,\dots C_r \\ \lambda_1 C_1 + \dots + \lambda_r C_r = 0}} \lambda_1 \mathcal{E}_1(X - C_1) + \dots + \lambda_r \mathcal{E}_r(X - C_r), \tag{13}$$

defines a subregular measure of error, and that if $\mathcal{E}(X) < \infty$, then the set of minimizers in (13) is a non-empty bounded closed convex subset of $\mathbb{R}^r$. The rest of the proof follows from the original Mixing Theorem in Rockafellar and Uryasev [2013] in combination with Theorem 3.1.

Fix any X with $\mathcal{E}(X) < \infty$. Then Theorem 3.1 implies that for every $k = 1, \dots, r$ set $\mathcal{S}_k(X) = \arg\min_C \mathcal{E}_k(X - C)$ is a non-empty closed bounded interval in $\mathbb{R}$, which we will write as $[J_k^-(X), J_k^+(X)]$. Note that

$$f_k(C) = \mathcal{E}_k(X - C)$$

are convex functions of C with minimum on $\mathcal{S}_k(X)$, hence they are strictly decreasing on $(-\infty, J_k^-(X))$ and strictly increasing on $(J_k^+(X), +\infty)$. Let

$$M^-(X) = \left\{ (C_1, \dots, C_r) \in \mathbb{R}^r \ : \ \sum_{k=1}^{r} \lambda_k C_k = 0,\ C_k \ge J_k^-(X),\ k = 1, \dots, r \right\},$$

$$M^+(X) = \left\{ (C_1, \dots, C_r) \in \mathbb{R}^r \ : \ \sum_{k=1}^{r} \lambda_k C_k = 0,\ C_k \le J_k^+(X),\ k = 1, \dots, r \right\},$$

and $M(X) = M^-(X) \cup M^+(X)$. We claim that for every $(C_1, \dots C_r) \notin M(X)$ satisfying $\sum_{k=1}^{r} \lambda_k C_k = 0$ there exists $(C'_1, \dots C'_r) \in M(X)$ such that

$$\sum_{k=1}^{r} \mathcal{E}_k(X - C_k) > \sum_{k=1}^{r} \mathcal{E}_k(X - C'_k). \tag{14}$$

Indeed, $(C_1, \dots C_r) \notin M(X)$ implies that there exists indices i and j such that $C_i < J_i^-(X)$ and $C_j > J_j^+(X)$. Let $\delta = \min\left(\frac{J_i^-(X) - C_i}{\lambda_j}, \frac{C_j - J_j^+(X)}{\lambda_i} \right)$, and let $C'_i = C_i + \delta\lambda_j$, $C'_j = C_j - \delta\lambda_i$, and $C'_k = C_k$ for $k \notin \{i, j\}$. Then

$$\sum_{k=1}^{r} \lambda_k C'_k = \lambda_i(\delta\lambda_j) + \lambda_j(-\delta\lambda_i) + \sum_{k=1}^{r} \lambda_k C_k = 0$$

and

$$\sum_{k=1}^r \mathcal{E}_k(X-C'_k) - \sum_{k=1}^r \mathcal{E}_k(X-C_k) = (f_i(C'_i)-f_i(C_i)) + (f_j(C'_j)-f_j(C_j)).$$

Because $J_i^-(X) \geq C'_i > C_i$ and f_i is strictly decreasing on $(-\infty, J_i^-(X))$, we have $f_i(C'_i) < f_i(C_i)$. Similarly, $J_j^+(X) \leq C'_j < C_j$ implies that $f_j(C'_j) < f_j(C_j)$. This proves (14). If $(C'_1, \dots C'_r) \in M(X)$, the claim is proved. Otherwise, we can repeat this process. At every step, the number of coordinates C_k lying outside of the interval $\mathcal{S}_k(X)$ decreases. Hence, after a finite number of steps, we obtain $(C'_1, \dots C'_r) \in M(X)$, and the claim follows.

The claim implies that (13) is equivalent to

$$\mathcal{E}(X) = \inf_{(C_1,\dots,C_r)\in M(X)} \lambda_1\mathcal{E}_1(X-C_1) + \dots + \lambda_r\mathcal{E}_r(X-C_r) \tag{15}$$

Because $M(X)$ is a compact subset of $\mathbb{R}^r$, and the objective function is convex and lower-semicontinuous, it follows that the infimum in (15) is attained, and the set of minimizers is a non-empty, bounded, closed convex set.

We next prove that $\mathcal{E}(.)$ is a subregular error measure. Its convexity and property (E1) are obvious corollaries from the corresponding properties of $\mathcal{E}_k$, so we only need to prove (E2) and lower-semicontinuity. We start with (E2). Fix any non-zero $X \in \mathcal{L}^p(\Omega)$. First, assume that X is non-constant. By Theorem 3.1 functionals

$$\mathcal{D}_k(X) = \inf_C \mathcal{E}_k(X-C)$$

are subregular deviation measures, hence, by (D2), there exist positive constants μ_k, $k=1,\dots,r$, such that $\mathcal{D}_k(\mu_k X) > 0$. Then convexity of $\mathcal{D}_k$ together with $\mathcal{D}_k(0)=0$ imply that $\mathcal{D}_k(\mu X) > 0$ for $\mu = \max_{1\leq k\leq r}\mu_k$, or equivalently, $\mathcal{E}_k(\mu X - C_k) > 0$ for all k and all constants C_k. But this implies that $\sum_{k=1}^r \lambda_k \mathcal{E}_k(\mu X - C_k) > 0$ for any positive constants λ_k. Because infimum in (13) is always attained, this implies that $\mathcal{E}(\mu X) > 0$ and proves (E2) for non-constant X.

If X is a constant, it is sufficient to consider cases $X = \pm 1$, that is, prove property (E2') formulated in the proof of Proposition 3.1. By property (E2') for each $\mathcal{E}_k$, there are exist constants $K_k^- < 0 < K_k^+$ such that $\mathcal{E}_k(K_k^-) > 0$ and $\mathcal{E}_k(K_k^+) > 0$. If $K \in \mathbb{R}$ is any constant such that $\mathcal{E}(K) = 0$, then there exist constants $(C_1, \dots, C_r)$ such that

$$\sum_{k=1}^r \lambda_k C_k = 0 \quad \text{and} \quad \sum_{k=1}^r \lambda_k \mathcal{E}_k(K-C_k) = 0.$$

Because all $\lambda_k > 0$, this is possible only if $\mathcal{E}_k(K-C_k) = 0$ for all k. But then

$$K_k^- < K - C_k < K_k^+, \quad k=1,\dots,r.$$

Multiplying these inequalities by λ_k and adding, we obtain

$$\sum_{k=1}^r \lambda_k K_k^- < K\sum_{k=1}^r \lambda_k - \sum_{k=1}^r \lambda_k C_k = K < \sum_{k=1}^r \lambda_k K_k^+$$

This implies that (E2') for $\mathcal{E}(.)$ holds with constants $K^- = \sum_{k=1}^r \lambda_k K_k^- < 0$ and $K^+ = \sum_{k=1}^r \lambda_k K_k^+ > 0$.

Now let us prove lower-semicontinuity of $\mathcal{E}(.)$. We need to prove that for any sequence X_n converging to X, we have $\liminf_{n\to\infty} \mathcal{E}(X_n) \geq \mathcal{E}(X)$. By passing to a subsequence if necessary, we may assume that $L = \lim_{n\to\infty} \mathcal{E}(X_n)$ exists. If $L = \infty$, the inequality $L \geq \mathcal{E}(X)$ is trivial, so we may assume that $L < \infty$. Then there is a constant c such that $\mathcal{E}(X_n) \leq c$ for all n.

Because the infimum in (13) is attained, there exist vectors $\mathbf{C}_n = (C_{n1}, \dots, C_{nk})$ such that

$$\sum_{k=1}^r \lambda_k C_{nk} = 0 \quad \text{and} \quad \mathcal{E}(X_n) = \sum_{k=1}^r \lambda_k \mathcal{E}_k(X_n - C_{nk}), \quad n = 1, 2, \dots$$

We have proved above that $\mathbf{C}_n \in M(X_n)$ for all n.

Inequality $c \geq \mathcal{E}(X_n)$ implies that

$$b := \frac{c}{\min\{\lambda_1, \dots, \lambda_k\}} \geq \mathcal{E}_k(X_n - C_{nk}) \geq \min_C \mathcal{E}_k(X_n - C), \quad k=1,\dots,r.$$

Hence, by Proposition 3.2,

$$U_k = \bigcup_{n=1}^\infty \mathcal{S}_k(X_n), \quad k=1,\dots,r$$

are bounded sets. This implies that set

$$M^* = \bigcup_{n=1}^\infty M(X_n)$$

is a compact subset of $\mathbb{R}^r$. Because $\mathbf{C}_n \in M^*$ for all n, we may assume, after passing to a subsequence if necessary, that sequence $\mathbf{C}_n$ converges component-wise to a vector $\mathbf{C} = (C_1, \ldots, C_r)$. Then $\sum_{k=1}^r \lambda_k C_k = 0$. Also, for each k, the sequence $X_n - C_{nk}$ converges to $X - C_k$ in $\mathcal{L}^p(\Omega)$. By lower-semicontinuity of $\mathcal{E}_k$, this implies that

$$\mathcal{E}_k(X - C_k) \leq \liminf_{n\to\infty} \mathcal{E}_k(X_n - C_{nk}).$$

Then

$$\begin{aligned}\mathcal{E}(X) &\leq \sum_{k=1}^r \lambda_k \mathcal{E}_k(X - C_k) \leq \sum_{k=1}^r \lambda_k \liminf_{n\to\infty} \mathcal{E}_k(X_n - C_{nk}) \\ &\leq \liminf_{n\to\infty} \sum_{k=1}^r \lambda_k \mathcal{E}_k(X_n - C_{nk}) = \liminf_{n\to\infty} \mathcal{E}(X_n) = L.\end{aligned}$$

□

Note that the scaling and reverting theorems from Rockafellar and Uryasev [2013] hold for the subregular risk quadrangles. The proofs are similar to that of the mixing theorem, with an appropriate change of variables. Below, we provide the formulations of these theorems for completeness.

Theorem 3.3 (Reverting Theorem)**.** *For $i = 1, 2$, let $(\mathcal{R}_i, \mathcal{D}_i, \mathcal{V}_i, \mathcal{E}_i)$ be a subregular quadrangle quartet with statistic $\mathcal{S}_i$. Then a subregular quadrangle quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$ is given by*

$$\begin{aligned}&\mathcal{S}(X) = \tfrac{1}{2}[\mathcal{S}_1(X) - \mathcal{S}_2(-X)],\\ &\mathcal{R}(X) = EX + \tfrac{1}{2}[\mathcal{R}_1(X) + \mathcal{R}_2(-X)],\\ &\mathcal{D}(X) = \tfrac{1}{2}[\mathcal{D}_1(X) + \mathcal{D}_2(-X)] = \tfrac{1}{2}[\mathcal{R}_1(X) + \mathcal{R}_2(-X)],\\ &\mathcal{V}(X) = EX + \min_C \left\{ \tfrac{1}{2}[\mathcal{V}_1(C + X) + \mathcal{V}_2(C - X)] - C \right\},\\ &\mathcal{E}(X) = \min_C \left\{ \tfrac{1}{2}[\mathcal{E}_1(C + X) + \mathcal{E}_2(C - X)] \right\}.\end{aligned}$$

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

Theorem 3.4 (Scaling Theorem)**.** *Let $(\mathcal{R}_0, \mathcal{D}_0, \mathcal{V}_0, \mathcal{E}_0)$ be a subregular quadrangle quartet with statistic $\mathcal{S}_0$ and consider any $\lambda \in (0, \infty)$. Then a subregular quadrangle quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ with statistic $\mathcal{S}$ is given by*

$$\begin{aligned}&\mathcal{S}(X) = \mathcal{S}_0(X),\\ &\mathcal{R}(X) = (1-\lambda)EX + \lambda\mathcal{R}_0(X), \quad \mathcal{D}(X) = \lambda\mathcal{D}_0(X),\\ &\mathcal{V}(X) = (1-\lambda)EX + \lambda\mathcal{V}_0(X), \quad \mathcal{E}(X) = \lambda\mathcal{E}_0(X),\end{aligned} \tag{16}$$

or alternatively by

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

Monotonicity and positive homogeneity are preserved in these constructions, except that monotonicity requires $\lambda \geq 1$ in (16).

In some examples in Section 2, the error and regret measures are given by

$$\mathcal{E}(X) = \mathbb{E}[e(X)] \tag{17}$$

and

$$\mathcal{V}(X) = \mathbb{E}[v(X)] \tag{18}$$

for some functions $e : \mathbb{R} \to [0, \infty]$ and $v : \mathbb{R} \to (-\infty, \infty]$. Then $\mathcal{E}$ and $\mathcal{V}$ are related by (Q3) if and only if e and v are related by

$$e(x) = v(x) - x, \qquad v(x) = x + e(x). \tag{19}$$

Theorem 3.5 (Expectation Theorem)**.** *For functions $e : \mathbb{R} \to [0, \infty]$ and $v : \mathbb{R} \to (-\infty, \infty]$ related by (19), the properties*

$$e \;\; \textit{is closed convex}, \quad e(0) = 0, \quad e(x) \geq 0 \textit{ for all } \; x, \textit{ and } e(a) > 0, e(b) > 0 \textit{ for some } \; a < 0 < b \tag{20}$$

amount to

$$v \;\; \textit{is closed convex}, \quad v(0) = 0, \quad v(x) \geq x \textit{ for all } \; x, \textit{ and } v(a) > a, v(b) > b \textit{ for some } \; a < 0 < b \tag{21}$$

and ensure that the functionals

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

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

$$\mathcal{S}(X) = \left\{ C \,\middle|\, \mathbb{E}[e'_-(X - C)] \le 0 \le \mathbb{E}[e'_+(X - C)] \right\} = \left\{ C \,\middle|\, \mathbb{E}[v'_-(X - C)] \le 1 \le \mathbb{E}[v'_+(X - C)] \right\}$$

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

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

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

Proof. Assume that (17) is a subregular error measure. For any constant x, (17) implies that $\mathcal{E}(x) = \mathbb{E}[e(x)] = e(x)$. Hence, convexity and closedness of $\mathcal{E}$ imply the corresponding properties of e, axiom (E1) in Definition 3.10 implies that $e(0) = 0$ and $e(x) \ge 0$ for all $x \in \mathbb{R}$, while axiom (E2) implies that for any $x \neq 0$ there exists $\lambda > 0$ such that $e(\lambda x) > 0$. Applying this property to $x = -1$ and $x = 1$, we deduce that $e(a) > 0$ and $e(b) > 0$ for some $a < 0 < b$.

Conversely, assume that functional $\mathcal{E}(X)$ is given by (17) for some function $e : \mathbb{R} \to [0, \infty]$ satisfying (20). Then convexity and closedness of $\mathcal{E}$ follow from the corresponding properties of e and the linearity of expectation. Further, $\mathcal{E}(0) = \mathbb{E}[e(0)] = e(0) = 0$, and, for any r.v. X, $\mathcal{E}(X) = \mathbb{E}[e(X)] \ge 0$. It is left to prove (E2). For any non-zero $X \in \mathcal{L}^p(\Omega)$ there exists $\epsilon > 0$ such that $\mathbb{P}(A) > 0$, where $A = \{w \in \Omega : |X(\omega)| \ge \epsilon\}$. Let $\lambda = \frac{\max(|a|,|b|)}{\epsilon}$. Then for every $\omega \in A$ we have either $\lambda X(\omega) \le a$ or $\lambda X(\omega) \ge b$, which by (20) implies that $e(\lambda X(\omega)) > 0$. Then

$$\mathcal{E}(\lambda X) = \mathbb{E}[e(\lambda X)] = \int_\Omega e(\lambda X(\omega)) d\mathbb{P} \ge \int_A e(\lambda X(\omega)) d\mathbb{P} > 0.$$

The regret part follows from the error part together with relations (19). □

Any quadrangle generated from an error measure of the form (17) will be called an *expectation quadrangle*. The Quantile Symmetric Average Union Quadrangle in Example 5 is an example of the expectation quadrangle that is not regular.

Construction of quadrangles. Any subregular error measure uniquely defines a quadrangle: the corresponding deviation measure is uniquely defined via (Q1), and then the risk and regret are uniquely defined via (Q3). The same is true for any subregular regret measure. However, any given subregular deviation measure $\mathcal{D}$ belongs to infinitely many subregular quadrangles. For example, for any $\alpha > 0$ functional $\mathcal{E}(X) = \mathcal{D}(X) + \alpha|E[X]|$ is a subregular error measure whose projected deviation measure is $\mathcal{D}$. Similarly, any subregular risk measure $\mathcal{R}$ belongs to infinitely many subregular quadrangles.

If risk measure $\mathcal{R}$ is subredular and coherent (in the general sense), then it is easy to check that

$$\mathcal{E}(X) := \mathcal{R}(|X|) \tag{22}$$

is a sub-regular error measure, which can then be used to define a quadrangle. This is the way we constructed the Quantile Symmetric Average Quadrangle in Example 4 starting from risk measure $\mathcal{R}(X) = \mathrm{CVaR}_\alpha(X)$. Note, however, then the risk measure in this quadrangle is not $\mathrm{CVaR}_\alpha(X)$.

3.2.2 Dual Representation and Conjugate Functionals

This subsection can be viewed as an extension of the Envelope Theorem from Rockafellar and Uryasev [2013] to the subregular quadrangle.

Up to now, we have been working with functionals on $\mathcal{L}^p(\Omega)$ for any $p \in [1, \infty]$, but to bring in duality, we restrict henceforth to $p \in [1, \infty)$ to ensure that the Banach space dual to $\mathcal{L}^p(\Omega)$ is $\mathcal{L}^q(\Omega)$ for $q \in (1, \infty]$, where $\frac{1}{p} + \frac{1}{q} = 1$ when $p > 1$ and $q = \infty$ when $p = 1$. Moreover, in the case of $\mathcal{L}^q(\Omega)$ for $q = \infty$ we replace the norm topology by the weak-$*$ topology induced by the pairing with $\mathcal{L}^1(\Omega)$ so as to have a topological vector space for which the dual is $\mathcal{L}^1(\Omega)$. This should be kept in mind in general statements below that refer to closedness of sets and functionals on $\mathcal{L}^q(\Omega)$ when specializing to $q = \infty$.

In this framework of paired spaces, the conjugate to a closed proper convex functional $\mathcal{F} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is the functional $\mathcal{F}^* : \mathcal{L}^q(\Omega) \to (-\infty, \infty]$ given by

$$\mathcal{F}^*(Q) = \sup_{X \in \mathcal{L}^p(\Omega)} (\mathbb{E}[XQ] - \mathcal{F}(X)), \qquad \forall Q \in \mathcal{L}^q(\Omega). \tag{23}$$

It is well-known that $\mathcal{F}^*$ is closed, proper and convex, and

$$\mathcal{F}(X) = \sup_{Q\in\mathcal{L}^q(\Omega)} (\mathbb{E}[XQ] - \mathcal{F}^*(Q)), \qquad \forall X \in \mathcal{L}^p(\Omega). \tag{24}$$

Let

$$\mathcal{Q} = \{Q \in \mathcal{L}^q(\Omega) \,|\, \mathcal{F}^*(Q) < \infty\} = \operatorname{dom} \mathcal{F}^*$$

be the effective domain of $\mathcal{F}^*$.

Proposition 3.3. Let $\mathcal{F} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ be a closed proper convex functional, and let $\mathcal{F}^*$ be its conjugate. Then

(i) $\mathcal{F}$ is a subregular error measure if and only if $\mathcal{F}^*$ satisfies:

- (E1*) $\mathcal{F}^*(Q) \geq \mathcal{F}^*(0) = 0$ for every $Q \in \mathcal{L}^q(\Omega)$;
- (E2*) for any non-zero $X \in \mathcal{L}^p(\Omega)$ there exists $Q \in \mathcal{Q}$ such that $\mathbb{E}[XQ] > 0$.

(ii) $\mathcal{F}$ is a subregular regret measure if and only if $\mathcal{F}^*$ satisfies:

- (V1*) $\mathcal{F}^*(Q) \geq \mathcal{F}^*(1) = 0$ for every $Q \in \mathcal{L}^q(\Omega)$;
- (V2*) for any non-zero $X \in \mathcal{L}^p(\Omega)$ there exists $Q \in \mathcal{Q}$ such that $\mathbb{E}[XQ] > \mathbb{E}[X]$.

(iii) $\mathcal{F}$ is a subregular deviation measure if and only if $\mathcal{F}^*$ satisfies:

- (D1*) $\mathcal{F}^*(Q) \geq \mathcal{F}^*(0) = 0$ for every $Q \in \mathcal{L}^q(\Omega)$;
- (D2*) $\mathbb{E}[Q] = 0$ for every $Q \in \mathcal{Q}$;
- (D3*) for any non-constant $X \in \mathcal{L}^p(\Omega)$ there exists $Q \in \mathcal{Q}$ such that $\mathbb{E}[XQ] > 0$.

(iv) $\mathcal{F}$ is a subregular risk measure if and only if $\mathcal{F}^*$ satisfies:

- (R1*) $\mathcal{F}^*(Q) \geq \mathcal{F}^*(1) = 0$ for every $Q \in \mathcal{L}^q(\Omega)$;
- (R2*) $\mathbb{E}[Q] = 1$ for every $Q \in \mathcal{Q}$;
- (R3*) for any non-constant $X \in \mathcal{L}^p(\Omega)$ there exists $Q \in \mathcal{Q}$ such that $\mathbb{E}[XQ] > \mathbb{E}[X]$.

Proof. Let us prove (i). If $\mathcal{F}$ is a subregular error measure, then, by (23),

$$\mathcal{F}^*(0) = \sup_{X\in\mathcal{L}^p(\Omega)} (0 - \mathcal{F}(X)) = -\inf_{X\in\mathcal{L}^p(\Omega)} \mathcal{F}(X) = 0,$$

where the last equality follows from (E1). Further, for any $Q \in \mathcal{L}^q(\Omega)$,

$$\mathcal{F}^*(Q) = \sup_{X\in\mathcal{L}^p(\Omega)} (\mathcal{E}[XQ] - \mathcal{F}(X)) \geq \mathcal{E}[0 \cdot Q] - \mathcal{F}(0) = 0.$$

This proves (E1*). Now, let $X \in \mathcal{L}^p(\Omega)$ be any non-zero r.v. Then by (E2) there exists $\lambda > 0$ such that $\mathcal{F}(\lambda X) > 0$. Then (24) implies that

$$0 < \mathcal{F}(\lambda X) = \sup_{Q\in\mathcal{L}^q(\Omega)} (\lambda\mathbb{E}[XQ] - \mathcal{F}^*(Q))$$

Hence, there exists $\bar{Q} \in \mathcal{L}^q(\Omega)$ such that $\lambda\mathbb{E}[X\bar{Q}] - \mathcal{F}^*(\bar{Q}) > 0$. Because $\mathcal{F}^*(\bar{Q}) \geq 0$, this implies that $\mathcal{F}^*(\bar{Q}) < \infty$ and $\mathbb{E}[X\bar{Q}] > 0$, and (E2*) follows.

Conversely, assume that $\mathcal{F}^*$ satisfies (E1*) and (E2*). Then $\mathcal{F}$ satisfies (E1) by the argument exactly as above. To prove (E2), fix any non-zero $X \in \mathcal{L}^p(\Omega)$. Then (E2*) implies that $\mathbb{E}[X\bar{Q}] > 0$ for some $\bar{Q} \in \mathcal{Q}$. Then (24) implies that for any $\lambda > \frac{\mathcal{F}^*(\bar{Q})}{\mathbb{E}[X\bar{Q}]}$,

$$\mathcal{F}(\lambda X) = \sup_{Q\in\mathcal{L}^q(\Omega)} (\lambda\mathbb{E}[XQ] - \mathcal{F}^*(Q)) \geq \lambda\mathbb{E}[X\bar{Q}] - \mathcal{F}^*(\bar{Q}) > 0.$$

This proves (E2).

Let us now prove (ii). $\mathcal{F}$ is a subregular regret measure if and only if $\mathcal{E}(X) = \mathcal{F}(X) - \mathbb{E}[X]$ is a subregular error measure. Then its conjugate

$$\mathcal{E}^*(Q) = \sup_{X\in\mathcal{L}^p(\Omega)} (\mathbb{E}[XQ] - \mathcal{E}(X)) = \sup_{X\in\mathcal{L}^p(\Omega)} (\mathbb{E}[X(Q + 1)] - \mathcal{F}(X)) = \mathcal{F}^*(Q + 1).$$

By (i), $\mathcal{E}(X) = \mathcal{F}(X) - \mathbb{E}[X]$ is a subregular error measure if and only if $\mathcal{E}^*(Q) = \mathcal{F}^*(Q+1)$ satisfies (E1*) and (E2*). This happens if and only if $\mathcal{F}^*(Q)$ satisfies (V1*) and (V2*).

We next prove (iii). The equivalence of $\mathcal{F}(X) \geq \mathcal{F}(0) = 0$ and (D1*) is already established when proving (i). We next prove that $\mathcal{F}(C) \leq 0$ for constants C if and only if (D2*) holds. Indeed,

$$\mathcal{F}(C) = \sup_{Q \in \mathcal{L}^q(\Omega)} (C\mathbb{E}[Q] - \mathcal{F}^*(Q)) \leq 0 \quad \forall C \in \mathbb{R}$$

if and only if

$$C\mathbb{E}[Q] \leq \mathcal{F}^*(Q) \quad \forall Q \in \mathcal{L}^q(\Omega), \quad \forall C \in \mathbb{R}.$$

The last inequality holds if, for every $Q \in \mathcal{L}^q(\Omega)$, we have either $\mathcal{F}^*(Q) = \infty$ or $\mathbb{E}[Q] = 0$. But this property is exactly (D2*). The equivalence of (D2) for $\mathcal{F}$ and (D3*) for $\mathcal{F}^*$ can be proved exactly as the equivalence of (E2) and (E2*) in part (i), with the only difference that we start with non-constant r.v. $X \in \mathcal{L}^p(\Omega)$ instead of non-zero one.

Finally, part (iv) follows from part (iii) in exactly the same way as part (ii) follows from (i). □

A functional $\mathcal{F} : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ is called positively homogeneous if

(F1) $\mathcal{F}(\lambda X) = \lambda \mathcal{F}(X)$ for all $X \in \mathcal{L}^p(\Omega)$ and every $\lambda \geq 0$.

Every closed proper convex functional that is potitively homogeneous can be represented in the form

$$\mathcal{F}(X) = \sup_{Q \in \mathcal{Q}} \mathbb{E}[QX] \tag{25}$$

for some convex closed set $\mathcal{Q} \subseteq \mathcal{L}^q(\Omega)$ called the risk envelope of $\mathcal{F}$. The risk envelope can be recovered from $\mathcal{F}$ by the formula

$$\mathcal{Q} = \{Q \in \mathcal{L}^q(\Omega) \,\big|\, \mathbb{E}[QX] \leq \mathcal{F}(X) \quad \text{for all} \quad X \in \mathcal{L}^p(\Omega)\}. \tag{26}$$

Then Proposition 3.3 implies that:

- (E') $\mathcal{F}$ is a subregular error measure if and only if $\mathcal{Q}$ contains 0, and, for any non-zero X, an r.v. Q such that $\mathbb{E}[XQ] > 0$;
- (V') $\mathcal{F}$ is a subregular regret measure if and only if $\mathcal{Q}$ contains 1, and, for any non-zero X, an r.v. Q such that $\mathbb{E}[XQ] > \mathbb{E}[X]$;
- (D') $\mathcal{F}$ is a subregular deviation measure if and only if $\mathcal{Q}$ contains 0, we have $\mathbb{E}[Q] = 0$ for every $Q \in \mathcal{Q}$, and, for any non-constant X, $\mathcal{Q}$ contains an r.v. Q such that $\mathbb{E}[XQ] > 0$;
- (R') $\mathcal{F}$ is a subregular risk measure if and only if $\mathcal{Q}$ contains 1, we have $\mathbb{E}[Q] = 1$ for every $Q \in \mathcal{Q}$, and, for any non-constant X, $\mathcal{Q}$ contains an r.v. Q such that $\mathbb{E}[XQ] > \mathbb{E}[X]$.

The conditions above can be equivalently reformulated in more geometric way as follows.

- (E') $\mathcal{F}$ is a subregular error measure if and only if $\mathcal{Q}$ is a closed, convex subset of $\mathcal{L}^q(\Omega)$ that contains the constant 0 in its quasi-interior; in other words, $0 \in \mathcal{Q}$ and every closed hyperplane H containing 0 has elements of $\mathcal{Q}$ in both of its associated open half-spaces;
- (V') $\mathcal{F}$ is a subregular regret measure if and only if $\mathcal{Q}$ is a closed, convex subset of $\mathcal{L}^q(\Omega)$ that contains the constant 1 in its quasi-interior; in other words, $1 \in \mathcal{Q}$ and every closed hyperplane H containing 1 has elements of $\mathcal{Q}$ in both of its associated open half-spaces;
- (D') $\mathcal{F}$ is a subregular deviation measure if and only if $\mathcal{Q}$ is a closed, convex subset of the closed hyperplane $H_0 = \{Q : \mathbb{E}[Q] = 0\}$ in $\mathcal{L}^q(\Omega)$ that contains the constant 0 in its quasi-interior relative to H_0; in other words, $0 \in \mathcal{Q}$ and every closed hyperplane $H \neq H_0$ containing 0 has elements of $\mathcal{Q}$ in both of its associated open half-spaces.
- (R') $\mathcal{F}$ is a subregular risk measure if and only if $\mathcal{Q}$ is a closed, convex subset of the closed hyperplane $H_1 = \{Q : \mathbb{E}[Q] = 1\}$ in $\mathcal{L}^q(\Omega)$ that contains the constant 1 in its quasi-interior relative to H_1; in other words, $1 \in \mathcal{Q}$ and every closed hyperplane $H \neq H_1$ containing 1 has elements of $\mathcal{Q}$ in both of its associated open half-spaces.

3.3 Parent Functionals and Corresponding Positive Homogeneous Families

Consider an arbitrary non-constant quasi-convex closed functional $\mathcal{J} : \mathcal{L}^q(\Omega) \to [-\infty, \infty]$. Let

$$A = \inf_{Q \in \mathcal{L}^q(\Omega)} \mathcal{J}(Q), \quad \text{and} \quad B = \sup_{Q \in \mathcal{L}^q(\Omega)} \mathcal{J}(Q). \tag{27}$$

Because $\mathcal{J}$ is non-constant, $A < B$. For any $\tau \in (A, B)$, let

$$\mathcal{Q}_\tau = \{Q \in \mathcal{L}^q(\Omega) \,|\, \mathcal{J}(Q) \leq \tau\}. \tag{28}$$

Then $\mathcal{Q}_\tau$, $A < \tau < B$ is a family of non-empty convex closed proper subsets of $\mathcal{L}^q(\Omega)$, that is nested in the sense that

(Q1) $Q_\tau \subseteq Q_t$ whenever $\tau \le t$.

Let g be any continuous, strictly increasing function mapping the interval $[A, B]$ onto $[0, +\infty]$. Then the functional $\mathcal{J}' : \mathcal{L}^q(\Omega) \to [0, \infty]$ defined by $\mathcal{J}'(X) := g(\mathcal{J}(X))$ produces via (28) exactly the same family of nested convex closed sets, just with parameter ranging over $(0, \infty)$. Hence, if we are interested in studying the family $\mathcal{Q}_\tau$, then without loss of generality, we may assume that $A = 0$ and $B = \infty$, that is, $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ and

$$\inf_{Q \in \mathcal{L}^q(\Omega)} \mathcal{J}(Q) = 0 \quad \text{and} \quad \sup_{Q \in \mathcal{L}^q(\Omega)} \mathcal{J}(Q) = +\infty. \tag{29}$$

Because $\mathcal{Q}_\tau$ are non-empty convex closed sets, the formula

$$\mathcal{F}_\tau(X) = \sup_{Q \in \mathcal{Q}_\tau} \mathbb{E}[QX] = \sup_{Q : \mathcal{J}(Q) \le \tau} \mathbb{E}[QX], \qquad \tau > 0, \tag{30}$$

defines a one-parameter family of positive homogeneous convex closed functionals $\mathcal{F}_\tau : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$. Because $\mathcal{Q}_\tau$ are proper subsets of $\mathcal{L}^q(\Omega)$, the functionals $\mathcal{F}_\tau$ are proper, that is, not identically $+\infty$. The property (Q1) translates into the fact that

(T1) for every fixed $X \in \mathcal{L}^q(\Omega)$, $\mathcal{F}_\tau(X)$ is a non-decreasing function of τ on $(0, +\infty)$.

Conversely, for any one-parameter family of positive homogeneous convex closed proper functionals $\mathcal{F}_\tau : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$, $\tau > 0$ satisfying (T1), the corresponding risk envelopes $\mathcal{Q}_\tau$, $\tau > 0$, can be recovered by

$$\mathcal{Q}_\tau = \{Q \in \mathcal{L}^q(\Omega) \,|\, \mathbb{E}[QX] \le \mathcal{F}_\tau(X) \quad \text{for all} \quad X \in \mathcal{L}^p(\Omega)\}, \quad \tau > 0, \tag{31}$$

see (26), and are non-empty closed convex proper subsets of $\mathcal{L}^q(\Omega)$ satisfying (Q1). This family form level sets (28) of the unique non-constant quasi-convex closed functional $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ defined by

$$\mathcal{J}(Q) = \inf\{\tau > 0 \,|\, Q \in \mathcal{Q}_\tau\} = \inf\{\tau > 0 \,|\, \mathbb{E}[QX] \le \mathcal{F}_\tau(X) \quad \text{for all} \quad X \in \mathcal{L}^p(\Omega)\} \tag{32}$$

Hence, we obtained the following result.

Proposition 3.4. Relations (30)–(32) define a one-to-one correspondence between quasi-convex closed functionals $\mathcal{J} : \mathcal{L}^q(\Omega) \to [-\infty, \infty]$ satisfying (29) and one-parameter families $\mathcal{F}_\tau$, $\tau > 0$, of positive homogeneous convex closed proper functionals $\mathcal{F}_\tau : \mathcal{L}^p(\Omega) \to (-\infty, \infty]$ satisfying (T1).

We next investigate how various properties of $\mathcal{J}$ translate into $\mathcal{F}_\tau$ and vice versa. We first record the obvious equivalence of the following statements.

- The infimum in (29) is attained and can be replaced by the minimum.
- There exists $\bar{Q} \in \mathcal{L}^q(\Omega)$ such that $\mathcal{J}(\bar{Q}) = 0$.
- The intersection of all sets $\mathcal{Q}_\tau$, $\tau > 0$ is non-empty.
- There exists $\bar{Q} \in \mathcal{L}^q(\Omega)$ such that $\mathcal{F}_\tau(X) \ge \mathbb{E}[\bar{Q}X]$ for all $X \in \mathcal{L}^p(\Omega)$ and all $\tau > 0$.

We next record the consequences of the uniqueness of the minimizer. The following statements are equivalent.

- $\operatorname{argmin} \mathcal{J}$ is a singleton.
- The intersection of all sets $\mathcal{Q}_\tau$, $\tau > 0$ is a singleton.
- There exists a unique $\bar{Q} \in \mathcal{L}^q(\Omega)$ such that $\lim_{\tau \to 0+} \mathcal{F}_\tau(X) = \mathbb{E}[\bar{Q}X]$ for all $X \in \mathcal{L}^p(\Omega)$.

If $\bar{Q}$ is the unique minimizer in (29), then $\mathcal{J}(\bar{Q}) = 0$, and we can interpret $\mathcal{J}$ as a "measure of distance" from Q to $\bar{Q}$. Then sets $\mathcal{Q}_\tau$ can be interpreted as sets of r.v.s Q at "distance" at most τ from $\bar{Q}$.

In particular, substituting $\bar{Q} = 1$, we obtain the following equivalence

$$\operatorname{argmin} \mathcal{J} = \{1\} \quad \Leftrightarrow \quad \lim_{\tau \to 0+} \mathcal{F}_\tau(X) = \mathbb{E}[X] \quad \text{for all} \quad X \in \mathcal{L}^p(\Omega),$$

while substituting $\bar{Q} = 0$, we obtain the following

$$\operatorname{argmin} \mathcal{J} = \{0\} \quad \Leftrightarrow \quad \lim_{\tau \to 0+} \mathcal{F}_\tau(X) = 0 \quad \text{for all} \quad X \in \mathcal{L}^p(\Omega).$$

What about the opposite limit, as $\tau \to \infty$? It is easy to see that

$$\lim_{\tau \to \infty} \mathcal{F}_\tau(X) = \sup_{Q \in \mathcal{C}} \mathbb{E}[QX], \qquad \text{where} \qquad \mathcal{C} := \operatorname{cl} \operatorname{dom} \mathcal{J}.$$

In particular, if

$$\operatorname{cl}\operatorname{dom}\mathcal{J} = \mathcal{P} := \{Q \in \mathcal{L}^q(\Omega) \,|\, Q \geq 0 \quad\text{and}\quad \mathbb{E}[Q] = 1\} \tag{33}$$

then

$$\lim_{\tau\to\infty} \mathcal{F}_\tau(X) = \sup X.$$

Also, property $Q \geq 0$ in (33) implies that each $\mathcal{F}_\tau$ is monotonic in sense that $\mathcal{F}_\tau(X) \geq 0$ whenever $X \geq 0$, while property $\mathbb{E}[Q] = 1$ in (33) implies that $\mathcal{F}_\tau(C) = C$ for constants C.

As another example, if

$$\operatorname{cl}\operatorname{dom}\mathcal{J} = \{Q \in \mathcal{L}^q(\Omega) \,|\, Q \geq 0\} \tag{34}$$

then

$$\lim_{\tau\to\infty} \mathcal{F}_\tau(X) = \begin{cases} 0, & \text{if } \sup X \leq 0 \\ +\infty, & \text{if } \sup X > 0\ . \end{cases} \tag{35}$$

We continue with the following observation.

Proposition 3.5. Let $\mathcal{J} : \mathcal{L}^q(\Omega) \to [-\infty, \infty]$ and $\mathcal{F}_\tau$, $\tau > 0$ be related by (30)–(32) as in Proposition 3.4. Then $\mathcal{J}$ is convex if and only if the family $\mathcal{F}_\tau$ satisfies (T1) and

(T2) for every fixed $X \in \mathcal{L}^p(\Omega)$, $\mathcal{F}_\tau(X)$ is a (non-decreasing and) *concave* function of τ on $(0, +\infty)$.

These properties are also equivalent to the property

(Q2) $(1-\lambda)\mathcal{Q}_{\tau_1} + \lambda\mathcal{Q}_{\tau_2} \subseteq \mathcal{Q}_\tau$ for $\tau = (1-\lambda)\tau_1 + \lambda\tau_2$ for all $\tau_1 > 0$, $\tau_2 > 0$ and $\lambda \in [0,1]$

for the corresponding family $\mathcal{Q}_\tau$, $\tau > 0$.

Proof. By definition, concavity (T2) means that $\mathcal{F}_\tau(X) \geq (1-\lambda)\mathcal{F}_{\tau_1} + \lambda\mathcal{F}_{\tau_2}$ for all $\tau_1 > 0$, $\tau_2 > 0$ and $\lambda \in [0,1]$, where $\tau = (1-\lambda)\tau_1 + \lambda\tau_2$. By (26), this translates into (Q2). Further, with (32), (Q2) is equivalent to the statement that if $Q = (1-\lambda)Q_1 + \lambda Q_2$ with $\mathcal{J}(Q_1) \leq \tau_1$ and $\mathcal{J}(Q_2) \leq \tau_2$ then $\mathcal{J}(Q) \leq \tau$, where $\tau = (1-\lambda)\tau_1 + \lambda\tau_2$. But this is exactly the convexity of $\mathcal{J}$. □

We remark that for non-constant convex functionals, the supremum condition in (29) is automatically satisfied, hence (29) reduces to the infimum condition.

Definition 3.12 (Divergence root)**.** A convex closed functional $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ satisfying (29) such that $\operatorname{argmin}\mathcal{J} = \{1\}$ and $\operatorname{cl}\operatorname{dom}\mathcal{J} = \{Q \in \mathcal{L}^q(\Omega) \,|\, Q \geq 0\}$ is called a *divergence root*.

The discussion above implies the following result.

Proposition 3.6. If a functional $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ is a divergence root then the corresponding family of positive homogeneous convex closed functionals $\mathcal{F}_\tau$ in (30) satisfies (T1), (T2), $\lim_{\tau\to 0+} \mathcal{F}_\tau(X) = \mathbb{E}[X]$, (35), and each $\mathcal{F}_\tau$ is monotonic. In particular, $\mathcal{F}_\tau$ is a subregular regret measure for every $\tau > 0$.

Proof. Only the last statement is new and requires proof. The inequality $\mathcal{F}_\tau(X) \geq \mathbb{E}[X]$ is obvious from (T1) and $\lim_{\tau\to 0+} \mathcal{F}_\tau(X) = \mathbb{E}[X]$. Hence, (V1) in Definition 3.9 holds. If (V2) fails, then there exists a non-zero X such that $\mathcal{F}_\tau(X) = \mathbb{E}[X]$. But then (T1) and (T2) imply that $\mathcal{F}_\tau(X) = \mathbb{E}[X]$ for all $\tau > 0$, which contradicts (35). □

Definition 3.13 (Stochastic divergence)**.** A convex closed functional $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ satisfying (29) such that $\operatorname{argmin}\mathcal{J} = \{1\}$ and $\operatorname{cl}\operatorname{dom}\mathcal{J} = \mathcal{P}$ is called a *stochastic divergence*.

The discussion above implies the following result.

Proposition 3.7. If functional $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ is a stochastic divergence, then the corresponding family of positive homogeneous convex closed functionals $\mathcal{F}_\tau$ in (30) satisfies (T1), (T2), $\lim_{\tau\to 0+} \mathcal{F}_\tau(X) = \mathbb{E}[X]$, $\lim_{\tau\to\infty} \mathcal{F}_\tau(X) = \sup X$, each $\mathcal{F}_\tau$ is monotonic and satisfies $\mathcal{F}_\tau(C) = C$ for constants C. In particular, $\mathcal{F}_\tau$ is a subregular (in fact, a coherent) risk measure for every $\tau > 0$.

Proof. Only the last statement is new and requires proof. The inequality $\mathcal{F}_\tau(X) \geq \mathbb{E}[X]$ is obvious from (T1) and $\lim_{\tau\to 0+} \mathcal{F}_\tau(X) = \mathbb{E}[X]$. Hence, (R1) in Definition 3.7 holds. If (R2) fails, then there exists a non-constant X such that $\mathcal{F}_\tau(X) = \mathbb{E}[X]$. But then (T1) and (T2) imply that $\mathcal{F}_\tau(X) = \mathbb{E}[X]$ for all $\tau > 0$, which implies that $\mathbb{E}[X] = \lim_{\tau\to\infty} \mathcal{F}_\tau(X) = \sup X$. But this is possible only if X is a constant. □

Stochastic divergence as a measure of distance. A stochastic divergence $\mathcal{J}$ can also be interpreted as a distance between *probability measures*. Recall that the underlying probability space $(\Omega, \mathcal{M}, \mathbb{P})$ comes with some "reference" probability measure $\mathbb{P}$, and when we write $\mathbb{E}[Q]$ for an r.v. Q, we actually mean $\mathbb{E}_{\mathbb{P}}[Q]$. Any r.v. $Q \in \mathcal{P}$ satisfies $Q \geq 0$ and $\mathbb{E}[Q] = 1$. Hence, we can define the probability measure $\mathbb{P}_Q$ as $\mathbb{P}_Q(A) = \mathbb{E}_{\mathbb{P}}[QI_A]$ for any event A, where I_A is the indicator function. Then $\mathbb{E}_{\mathbb{P}_Q}[X] = \mathbb{E}[QX]$ for any r.v. X. Hence, for any $\mathcal{Q} \subseteq \mathcal{P}$,

$$\sup_{Q\in\mathcal{Q}} \mathbb{E}[QX] = \sup_{\mathbb{P}_Q \in \mathcal{P}(\mathcal{Q})} \mathbb{E}_{\mathbb{P}_Q}[X],$$

where $\mathcal{P}(\mathcal{Q})$ is the set of probability measures corresponsing to r.v.s $Q \in \mathcal{Q}$. In particular, (30) reduces to

$$\mathcal{F}_\tau(X) = \sup_{\mathbb{P}_Q \in \mathcal{P}(\mathcal{Q}_\tau)} \mathbb{E}_{\mathbb{P}_Q}[X], \quad \text{where} \quad \mathcal{P}(\mathcal{Q}_\tau) = \{\mathbb{P}_Q : \mathcal{J}(Q) \leq \tau\} = \{\mathbb{P}_Q : \mathcal{G}(\mathbb{P}_Q) \leq \tau\}, \tag{36}$$

where $\mathcal{G}(\mathbb{P}_Q) := \mathcal{J}(Q)$. The functional $\mathcal{G}$ maps probability measures absolutely continuous with respect to $\mathbb{P}$ to $[0, \infty]$. It is convex, closed with $\mathcal{G}(\mathbb{P}) = 0$, and, for $\mathbb{P}_Q \neq \mathbb{P}$, $\mathcal{G}(\mathbb{P}_Q)$ can be interpreted as the "distance" from $\mathbb{P}_Q$ to $\mathbb{P}$. By Proposition 3.7, $\mathcal{F}_\tau$ is a coherent risk measure in the basic sense for every such $\mathcal{G}$ and every $\tau > 0$.

A popular example of such "distance" is the Wasserstein divergence

$$W(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\gamma \in \Pi(\mathbb{P}_1, \mathbb{P}_2)} \int_{\Omega\times\Omega} d(x, y)\, d\gamma(x, y)$$

where:

- $\Pi(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all joint probability measures γ on $\Omega \times \Omega$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$;
- $d(x, y)$ is a metric on Ω.

A family of risk measures (36) with $\mathcal{G}(\mathbb{P}_Q) = W(\mathbb{P}_Q, \mathbb{P})$ is discussed in Rockafellar [2024].

We next consider the case when

$$\mathcal{J}(Q) = \mathbb{E}(\varphi(Q)), \quad Q \in \mathcal{L}^q(\Omega) \tag{37}$$

and investigate under what conditions on $\varphi : \mathbb{R} \to (-\infty, \infty]$ functional (37) has the properties listed above. We have the following implications.

- If φ is closed proper convex, then so is $\mathcal{J}$;
- If φ is non-constant, then so is $\mathcal{J}$;
- If $\inf_{y\in\mathbb{R}} \varphi(y) = 0$ and $\sup_{y\in\mathbb{R}} \varphi(y) = \infty$, then (29) holds;
- In particular, if φ is convex, non-constant, and $\inf_{y\in\mathbb{R}} \varphi(y) = 0$, then (29) holds;
- If the infimum $\inf_{y\in\mathbb{R}} \varphi(y)$ is attained and can be replaced by a minimum, then the same is true for the infimum in (29);
- If $\operatorname{argmin} \varphi = \{C\}$ is a singleton, then $\operatorname{argmin} \mathcal{J} = \{C\}$;
- In particular, $\operatorname{argmin} \varphi = \{0\} \Rightarrow \operatorname{argmin} \mathcal{J} = \{0\}$ and $\operatorname{argmin} \varphi = \{1\} \Rightarrow \operatorname{argmin} \mathcal{J} = \{1\}$;
- If φ is convex, then $\operatorname{cl} \operatorname{dom} \varphi = [a, b]$ is a closed convex interval, where $-\infty \leq a < b \leq \infty$. Then $\operatorname{cl} \operatorname{dom} \mathcal{J} = \{Q \in \mathcal{L}^q(\Omega) \,|\, a \leq Q \leq b\}$;
- In particular, if $\operatorname{cl} \operatorname{dom} \varphi = [0, \infty]$, then (34) holds;
- If φ is convex, closed, $\min_{y\in\mathbb{R}} \varphi(y) = 0$, $\operatorname{argmin} \varphi = \{1\}$, and $\operatorname{cl} \operatorname{dom} \varphi = [0, \infty]$, then $\mathcal{J}$ is a divergence root.

Assume that $\mathcal{J}$ is convex, and let $\mathcal{F}$ be the conjugate functional to $\mathcal{J}$ defined in (24). Then Rockafellar [2024] points out that it is a basic formula in convex analysis that

$$\mathcal{F}_\tau(X) = \inf_{\lambda>0} \lambda[\mathcal{F}(\lambda^{-1}X) + \tau]. \tag{38}$$

Proposition 3.8. If $\mathcal{F}$ is a subregular error, regret, deviation, or risk measure, then $\mathcal{F}_\tau$ is a positive homogeneous subregular error, regret, deviation, or risk measure, respectively.

Proof. If $\mathcal{F}$ is, for example, an error measure, then $\mathcal{F}^*$ satisfy the conditions (E1*) and (E2*) in Proposition 3.3. Then (E1*) implies (29), while (E2*) implies that for any non-constant r.v. X there exists $\bar{Q}$ such that $\mathcal{F}^*(\bar{Q}) < \infty$ and $0 < \mathbb{E}[X\bar{Q}]$. Let $\lambda = \min\{\frac{\tau}{\mathcal{F}^*(\bar{Q})}, 1\}$ if $\mathcal{F}^*(\bar{Q}) > 0$ and $\lambda = 1$ if $\mathcal{F}^*(\bar{Q}) = 0$. Then the convexity of $\mathcal{F}^*$ implies that

$$\mathcal{F}^*(\lambda\bar{Q}) = \mathcal{F}^*(\lambda\bar{Q} + (1-\lambda)\cdot 0) \leq \lambda\mathcal{F}^*(\bar{Q}) + (1-\lambda)\cdot 0 \leq \tau.$$

Then for $Q' = \lambda\bar{Q}$ we have $Q' \in \mathcal{Q}_\tau$ and $0 < \mathbb{E}[XQ']$. Also, $\mathcal{Q}_\tau$ contains 0. Hence, it satisfies the condition (E') in Section 3.2.2, which implies that it is a dual set of an error measure. The proofs for the regret, deviation, and risk measures are similar. □

If (38) holds, we will call $\mathcal{F}$ the *parent* functional for a family $\mathcal{F}_\tau$, $\tau > 0$. In particular, if $\mathcal{J} : \mathcal{L}^q(\Omega) \to [0, \infty]$ is a divergence root, then its conjugate $\mathcal{F}$ is a subregular regret measure, which is the parent of the family $\mathcal{F}_\tau$ of subregular regret measures corresponding to $\mathcal{J}$ via Proposition 3.6.

Proposition 3.9. Let $(\mathcal{E}, \mathcal{V}, \mathcal{D}, \mathcal{R})$ be a quadrangle generated starting from a subregular error measure $\mathcal{E}$ (or from subregular regret measure $\mathcal{V}$) via relations (Q2), (Q3). Let $\tau > 0$ and let $\mathcal{E}_\tau$, $\mathcal{V}_\tau$, $\mathcal{D}_\tau$, and $\mathcal{R}_\tau$ be positive homogeneous subregular error, regret, deviation, and risk measures defined in (38). Then $\mathcal{E}_\tau$, $\mathcal{V}_\tau$, $\mathcal{D}_\tau$ and $\mathcal{R}_\tau$ are also related by (Q2), (Q3).

Proof. We first prove (6). Indeed,

$$\mathcal{V}_\tau(X) = \inf_{\lambda>0} \lambda[\mathcal{V}(\lambda^{-1}X) + \tau] = \inf_{\lambda>0} \lambda[\mathcal{E}(\lambda^{-1}X) + \mathbb{E}[\lambda^{-1}X] + \tau]$$

$$= \mathbb{E}[X] + \inf_{\lambda>0} \lambda[\mathcal{E}(\lambda^{-1}X) + \tau] = \mathbb{E}[X] + \mathcal{E}_\tau(X).$$

The proof of (5) is similar. Next,

$$\mathcal{D}_\tau(X) = \inf_{\lambda>0}[\lambda\mathcal{D}(\lambda^{-1}X) + \lambda\tau] = \inf_{\lambda>0}[\lambda \inf_{C_1} \mathcal{E}(\lambda^{-1}X - C_1) + \lambda\tau] = \inf_{\lambda>0} \inf_{C} [\lambda\mathcal{E}(\lambda^{-1}(X - C)) + \lambda\tau],$$

where $C = \lambda C_1$. We can then exchange the order of infimums and obtain

$$\mathcal{D}_\tau(X) = \inf_{C} \inf_{\lambda>0} [\lambda\mathcal{E}(\lambda^{-1}(X - C)) + \lambda\tau] = \inf_{C} \mathcal{E}_\tau(X - C).$$

This proves (Q1), and (Q2) can be proved similarly. □

As a simple example, if we start with an error measure

$$\mathcal{E}(X) = \mathbb{E}[X^2],$$

then, for any $\tau > 0$,

$$\mathcal{E}_\tau(X) = \inf_{\lambda>0} \lambda[\mathbb{E}[(\lambda^{-1}X)^2] + \tau] = \inf_{\lambda>0}[\lambda^{-1}\mathbb{E}[X^2] + \lambda\tau] = 2\sqrt{\tau}\sqrt{\mathbb{E}[X^2]} = 2\sqrt{\tau}||X||_2.$$

The projected deviation measure is then

$$\mathcal{D}_\tau(X) = 2\sqrt{\tau}\sigma(X),$$

see Example 1 in Rockafellar and Uryasev [2013]. Equivalently, we may first note that the projected deviation measure for $\mathcal{E}$ is

$$\mathcal{D}(X) = \sigma^2(X),$$

see Example 2 in Rockafellar and Uryasev [2013], and then by (38)

$$\mathcal{D}_\tau(X) = \inf_{\lambda>0} \lambda[\sigma^2(\lambda^{-1}X) + \tau] = 2\sqrt{\tau}\sigma(X).$$

As another example, Rockafellar [2024] proved that if $\mathcal{V}$ is the indicator of the set

$$\mathcal{X} = \{X | \exists Y \geq 1, \text{ such that } \mathbb{E}[X + Y] = 1, X + Y \geq 0\}$$

then the corresponding risk measure $\mathcal{R}_\tau$ is the conditional value-at-risk with $\tau = \frac{\alpha}{1-\alpha}$. We remark that the set $\mathcal{X}$ can be written in a simpler form. Indeed, with $Z = Y - 1$ we obtain

$$\mathcal{X} = \{X | \exists Z \geq 0, \text{ such that } \mathbb{E}[X + Z] = 0, X + Z \geq -1\}.$$

We claim that $X \in \mathcal{X}$ if and only if $\mathbb{E}[\max(-1, X)] \leq 0$. Indeed, $Z \geq 0$ implies that $X + Z \geq X$. Hence, $X + Z \geq \max(-1, X)$, and $\mathbb{E}[\max(-1, X)] \leq \mathbb{E}[X + Z] = 0$. On the other hand, if $\mathbb{E}[\max(-1, X)] \leq 0$, then there exist a constant $c > 0$ such that $\mathbb{E}[\max(-1 + c, X + c)] = 0$, and we may take $Z = \max(-1 + c, X + c) - X$. In conclusion,

$$\mathcal{X} = \{X | \mathbb{E}[\max(-1, X)] \leq 0\} = \{X | \mathbb{E}[X + 1]_+ \leq 1\},$$

and

$$\mathcal{V}(X) = \begin{cases} 0, & \text{if } \mathbb{E}[X + 1]_+ \leq 1 \\ +\infty, & \text{otherwise.} \end{cases}$$

This functional is a subregular regret measure. Indeed, inequality $\mathcal{V}(X) \geq \mathbb{E}[X]$ is trivial if $\mathcal{V}(X) = \infty$. If, conversely, $\mathcal{V}(X) = 0$, then $1 \geq \mathbb{E}[X + 1]_+ \geq \mathbb{E}[X + 1]$ implies that $0 \geq \mathbb{E}[X]$, and (V1) follows. From this argument, it is clear that equality $\mathcal{V}(X) = \mathbb{E}[X]$ holds if and only if X belongs to the set $A = \{X : \mathbb{E}[X] = 0, \mathbb{P}(X \geq -1) = 1\}$. Obviously, for any non-zero $X \in A$ there exists $\lambda > 0$ such that $\lambda X \notin A$. This implies (V2). The corresponding risk measure is

$$\mathcal{R}(X) = \inf\{C \,|\, \mathbb{E}[X + 1 - C]_+ \leq 1\}.$$

Also, for any $\tau > 0$,

$$\mathcal{V}_\tau(X) = \inf_{\lambda>0} \lambda[\mathcal{V}(\lambda^{-1}X) + \tau] = \tau \inf\{\lambda > 0 : \mathbb{E}[\lambda^{-1}X + 1]_+ \leq 1\}.$$

4 Generalized Regression and Statistical Estimation

4.1 Functional Regression

Regression is one of the central concepts in statistical estimation theory. Given a random variable $Y \in \mathcal{L}^p(\Omega)$ (*the regressant* or independent variable) and a collection of random variables $X_i \in \mathcal{L}^p(\Omega)$, $i = 1, \ldots, n$, (*the regressors* or independent variables), the task of *functional regression* (see Kendall [1951, 1952]) is to find a function $f : \mathbb{R}^n \to \mathbb{R}$, belonging to a class of measurable functions $\mathcal{C}$, that minimizes the *regression residual* $Z_f := Y - f(\mathbf{X})$, $\mathbf{X} = (X_1, \ldots, X_n)$, with respect to a particular error $\mathcal{E}$ (e.g., mean squared error, mean absolute error). Specifically, the goal is to solve the following stochastic optimization problem:

$$\min_{f \in \mathcal{C}} \quad \mathcal{E}(Z_f). \tag{39}$$

In general, different choices of error result in different optimal solutions of (39).

From the statistical estimation perspective, given the regressant Y and the vector of regressors $\mathbf{X}$, the aim is to estimate (track) a desired *conditional statistic* $\mathcal{S}(Y|\mathbf{X})$ (e.g., conditional mean $\mathbb{E}[Y|\mathbf{X}]$ or conditional quantile $\mathrm{VaR}_\alpha(Y|\mathbf{X})$) via regression. The classical approach to this problem is to find an appropriate *loss function* $\ell : \mathbb{R} \to \mathbb{R}$ and solve (39), where $\mathcal{E}(Z_f) = \mathbb{E}[\ell(Z_f)]$.[4] Then (see Bach [2024])

$$f^*(\mathbf{x}) \in \underset{C \in \mathbb{R}}{\operatorname{argmin}} \quad \mathbb{E}[Y - C|\mathbf{X} = \mathbf{x}], \tag{40}$$

where the equality $\mathbf{X} = \mathbf{x}$ is understood pointwise.

Of course, the above approach works provided a loss function exists for a given statistic. Statistics for which such a loss function exists are called *elicitable* (see Lambert et al. [2008]). For non-elicitable statistics, however, the expected loss approach is infeasible, and thus other approaches should be considered. The RQ provides a unified framework for both elicitable and non-elicitable statistics by considering axiomatically defined errors beyond expected losses.

Theorem 4.1 (Regression Theorem)**.** *Consider problem*

$$\text{minimize } \mathcal{E}(Z_f) \text{ over all } f \in \mathcal{C}, \text{ where } Z_f = Y - f(\mathbf{X}) \tag{41}$$

for random variables $\mathbf{X}$ *and* Y *in the case of* $\mathcal{E}$ *being a subregular measure of error and* $\mathcal{C}$ *being a class of functions* $f : \mathbb{R}^n \to \mathbb{R}$ *such that*

$$f \in \mathcal{C} \implies f + C \in \mathcal{C} \text{ for all } C \in \mathbb{R}. \tag{42}$$

Let $\mathcal{D}$ *and* $\mathcal{S}$ *correspond to* $\mathcal{E}$ *as in the Quadrangle Theorem. Problem* (41) *is equivalent then to:*

$$\text{minimize } \mathcal{D}(Z_f) \text{ over all } f \in \mathcal{C} \text{ such that } 0 \in \mathcal{S}(Z_f). \tag{43}$$

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

$$\begin{aligned} f(\mathbf{x}) \in \mathcal{S}(Y|\mathbf{x}) \text{ almost surely for } \mathbf{x} \in D, \\ \text{where } Y|\mathbf{x} = Y|\mathbf{X} = \mathbf{x} \text{ (conditional distribution)}, \end{aligned} \tag{44}$$

with D *being the support of the distribution in* $\mathbb{R}^n$ *induced by* $\mathbf{X}$*,*[5] *then that* f *solves the regression problem and tracks this conditional statistic*[6] *in the sense that*

$$f(\mathbf{X}) \in \mathcal{S}(Y|\mathbf{X}) \text{ almost surely.} \tag{45}$$

The Regression Theorem 4.1 of Rockafellar and Uryasev [2013] remains valid for the subregular functionals with the same proof.

In general, the inclusion (45) holds only for the errors of expectation type. However, more can be said in the case of linear regression.

4.2 Linear Regression

Consider the linear regression problem

$$\min_{(c_0, c_1, \ldots, c_n) \in \mathbb{R}^{n+1}} \mathcal{E}\left(Y - c_0 - \sum_{i=1}^{n} c_i X_i\right), \tag{46}$$

where $\mathcal{E}$ is a subregular error measure. Theorem 3.1 in Rockafellar et al. [2008] proves that the solution set in (46) is non-empty under some additional assumptions on $\mathcal{E}$ such as positive homogeneity. Here we prove the same result for any subregular error measure with no additional assumptions.

[4] Such errors are referred to as the errors of expectation type according to Rockafellar and Uryasev [2013].

[5] Almost surely, in (44), refers to this distribution.

[6] It is assumed, for this part, that the distribution of $Y|\mathbf{x}$ for $\mathbf{x} \in D$ belongs to $\mathcal{L}^p(\Omega)$, and the same then for the random variable $Y|\mathbf{X}$ obtained from it.

Proposition 4.1. Let $\mathcal{E}$ be a subregular error measure. Then the set of minimizers in (46) is a non-empty, closed, convex subset of $\mathbb{R}^{n+1}$.

Proof. The convexity and closedness of the set of minimizers follow from the convexity and lower-semicontinuity of the objective function, so we only need to prove its non-emptiness. Let $\mathcal{X}$ be the set of all r.v.s X representable as $X = c_0 + \sum_{i=1}^n c_i X_i$ for some $(c_0, c_1, \dots, c_n) \in \mathbb{R}^{n+1}$. Optimization problem (46) can be rewritten as

$$\min_{X \in \mathcal{X}} \mathcal{E}\left(Y - X\right),$$

and, in this formulation, it is clear that we may assume that the random variables $X_1, \dots, X_n$ satisfy the linear independence condition that $\sum_{i=1}^n c_i X_i$ in not constant unless $c_1 = \dots = c_n = 0$, because otherwise we can remove some of the X_i without changing set $\mathcal{X}$.

Function

$$f(d, c_0, c_1, \dots, c_n) = \mathcal{E}\left(dY - c_0 - \sum_{i=1}^n c_i X_i\right)$$

is a convex lower-semicontinuous function on $\mathbb{R}^{n+2}$, satisfying $f(x) \geq 0$ for all $x \in \mathbb{R}^{n+2}$ and $f(0) = 0$. We need to minimize f subject to the constraint that $d = 1$. If $f(1, c_0, c_1, \dots, c_n)$ is identical $+\infty$ then the statement of the Proposition trivially holds. Otherwise select some $c_0, c_1, \dots, c_n$ such that $f(1, c_0, c_1, \dots, c_n) = C < \infty$. Let $D_C := \{x \in \mathbb{R}^{n+2} : f(x) \leq C\}$ and D'_C be the set of vectors in D_C with first coordinate 1. If D_C is a bounded subset of $\mathbb{R}^{n+2}$, then so is D'_C. Because D'_C is also non-empty and closed, and f is lower-semicontinuous, this implies that the set of minimizers is non-empty.

It is left to consider the case when the set D_C is unbounded. Then there is a sequence $\{x_k\}_{k=0}^\infty$ such that $\lim\limits_{k\to\infty} ||x_k|| = \infty$ and $f(x_k) \leq C$ for all k, where $||\cdot||$ is the usual Euclidean norm in $\mathbb{R}^{n+2}$. Then $y_k = \frac{x_k}{||x_k||}$, $k = 1, 2, \dots$ are unit vector belonging to the compact set $\{x \in \mathbb{R}^{n+2} : ||x|| = 1\}$, hence, by passing to a subsequence if necessary, we may assume that $\lim\limits_{k\to\infty} y_k = y^*$ for some unit vector $y^* \in \mathbb{R}^{n+2}$. Now, for any $\lambda > 0$, let K_λ be an integer such that $||x_k|| \geq \lambda$ for all $k \geq K_\lambda$. Then the convexity of f implies that

$$f(\lambda y_k) = f\left(\left(1 - \frac{\lambda}{||x_k||}\right) 0 + \frac{\lambda}{||x_k||} x_k\right) \leq \left(1 - \frac{\lambda}{||x_k||}\right) f(0) + \frac{\lambda}{||x_k||} f(x_k) \leq \frac{\lambda}{||x_k||} C,$$

for all $k \geq K_\lambda$. Hence,

$$0 \leq \lim_{k\to\infty} f(\lambda y_k) \leq \lim_{k\to\infty} \frac{\lambda}{||x_k||} C = 0,$$

from which we conclude that $\lim\limits_{k\to\infty} f(\lambda y_k) = 0$. Now lower semicontinuity of f implies that

$$0 \leq f(\lambda y^*) = f\left(\lim_{k\to\infty} (\lambda y_k)\right) \leq \lim_{k\to\infty} f(\lambda y_k) = 0,$$

hence $f(\lambda y^*) = 0$ for all $\lambda > 0$. Let us write y^* in the coordinate form as $y^* = (d^*, c_0^*, \dots, c_n^*)$. If $d^* = 0$, then

$$0 = f(\lambda y^*) = \mathcal{E}\left(\lambda X^*\right) \quad \text{for all} \quad \lambda > 0, \quad \text{where} \quad X^* = -c_0^* - \sum_{i=1}^n c_i^* X_i) \in \mathcal{X},$$

which is a contradiction with (E2) unless $X^* = 0$. By the linear independence condition, $X^* = 0$ is possible only if $c_0^* = \dots = c_n^* = 0$, but this is a contradiction with $||y^*|| = 1$.

Hence, $d^* \neq 0$. But then for $\lambda = 1/d^*$ we have

$$0 = f(\lambda y^*) = \mathcal{E}\left(Y - (\lambda c_0^*) - \sum_{i=1}^n (\lambda^* c_i) X_i\right),$$

which implies that the minimum in (46) is 0 and the set of minimizers in non-empty. □

It is easy to check that Proposition 4.1 does not hold without condition (E2). Indeed, fix any non-constant $X \in \mathcal{X}$ and consider functional

$$\mathcal{E}(Z) = \begin{cases} 0, & \text{if} \quad Z = 0; \\ \frac{a^2}{b}, & \text{if} \quad Z = aY + bX \quad \text{for some constants} \quad a \geq 0 \quad \text{and} \quad b > 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

The convexity of $\mathcal{E}(Z)$ follows from the convexity of function $f(a, b) = \frac{a^2}{b}$ in the region $\{a \geq 0, \, b > 0\}$. Also, $\mathcal{E}$ is closed and satisfies (E1). On the other hand, $\mathcal{E}(Z) = 0$ if and only $Z = bX$ for some $b > 0$. If $Y \notin \mathcal{X}$, then $Y - c_0 - \sum_{i=1}^n c_i X_i$ is never of this form, hence the objective function in (46) is never 0. However, for any

$b > 0$, $\mathcal{E}(Y + bX) = \frac{1}{b}$, hence, if $b \to \infty$, then the objective function in (46) can be arbitrary close to 0. Hence, the set of minimizers in this example is the empty set.

The regression decomposition theorem (Theorem 3.2 in Rockafellar et al. [2008]) remains valid for any subregular error measure, with essentially the same proof.

Now we turn to conditional statistic tracking. For this, we need the following definition.

Definition 4.1 (Representation of risk identifiers, Rockafellar and Royset [2015])**.** A risk identifier Q^Y at $Y \in \mathcal{L}^p(\Omega)$ for a regular measure of risk will be called *representable* if there exists a Borel-measurable function $h^Y : \mathbb{R} \to \mathbb{R}$, possibly depending on Y, such that

$$Q^Y(\omega) = h^Y(Y(\omega)) \text{ for a.e. } \omega \in \Omega.$$

The following is a reformulation of Theorem 5.1 of Rockafellar and Royset [2015].

Theorem 4.2 (Statistic tracking in regression)**.** *For given $c_0^* \in \mathbb{R}$ and $\mathbf{c}^* \in \mathbb{R}^n$, assume that*

$$Y(\omega) = c_0^* + \mathbf{c}^{*\top}\mathbf{X} + \varepsilon(\omega) \quad \textit{for all } \omega \in \Omega \tag{47}$$

with $\varepsilon \in \mathcal{L}^p(\Omega)$ independent of X_i, $i = 1, \dots, n$ and $\mathcal{S}(\varepsilon) = 0$. Let $(\mathcal{S}, \mathcal{R}, \mathcal{D}, \mathcal{E})$ be a subregular quadrangle quartet. If $\mathcal{R}$ has a representable risk identifier at ε and $\varepsilon \in \operatorname{int}(\operatorname{dom} \mathcal{R})$, then

$$c_0^* + \mathbf{c}^{*\top}\mathbf{X} \in \mathcal{S}(Y|\mathbf{X}) \quad a.s. \tag{48}$$

The proof of the theorem remains the same.

5 Stochastic Optimization and Distributional Robustness

The notion of risk has become central in modern stochastic optimization and closely related fields such as machine learning. Whenever the uncertainty is modeled probabilistically, the decision-maker aims to select a decision that minimizes the risk of future losses, i.e., solve

$$\min_{\mathbf{w} \in \mathcal{W}} \quad \mathcal{R}(\ell(\mathbf{w}, \omega)), \tag{49}$$

where $\mathcal{R} : \mathcal{L}^p(\Omega, \mathcal{A}, \mathbb{P}_0) \to (-\infty, \infty]$ is a subregular risk measure and $\ell : \mathcal{W} \times \Omega \to \mathbb{R}$ is a real-valued random loss function assumed to be convex in w over a closed, convex, and nonempty set $\mathcal{W} \subseteq \mathbb{R}^n$.

Equation (24) allows equivalently rewriting (49) as follows

$$\min_{\mathbf{w} \in \mathcal{W}} \sup_{Q \in \mathcal{Q}^{\mathcal{R}}} \quad \mathbb{E}_{\mathbb{P}_0}[Q\ell(\mathbf{w}, \omega) - \mathcal{R}^*(Q)]. \tag{50}$$

Then for positive homogeneous risk functionals $\mathcal{R}$, (50) is

$$\min_{\mathbf{w} \in \mathcal{W}} \sup_{Q \in \mathcal{Q}^{\mathcal{R}}} \quad \mathbb{E}_{\mathbb{P}_0}[Q\ell(\mathbf{w}, \omega)], \tag{51}$$

where $\mathcal{Q}^{\mathcal{R}} = \{Q \in \mathcal{L}^q(\Omega) \,|\, \mathbb{E}_{\mathbb{P}_0}[Q] = 1,\ \mathcal{R}^*(Q) < \infty\}$.

The stochastic optimization problem (51) can be interpreted as a *relaxed robust optimization* problem, where the random variable Q serves as a "normalized" random weighting function. Therefore, instead of hedging against the worst-case scenario $\omega \in \Omega$ (see Ben-Tal et al. [2009]), the decision-maker selects the worst weighting function $\bar{Q} \in \mathcal{Q}^{\mathcal{R}}$ and minimizes $\mathbb{E}_{\mathbb{P}_0}[\bar{Q}\ell(\mathbf{w}, \omega)]$ over $\mathbf{w} \in \mathcal{W}$. Classical robust optimization takes $\bar{Q}$ such that it is 0 for almost all $\omega \in \Omega$ except the one $\bar{\omega} \in \Omega$ that maximizes the loss function $\ell(\mathbf{w}^*, \bar{\omega})$ at optimal $\mathbf{w}^* \in \mathcal{W}$.

In turn, for positive homogeneous and monotonic risk functionals $\mathcal{R}$, problem (51) is as follows

$$\min_{\mathbf{w} \in \mathcal{W}} \sup_{Q \in \mathcal{Q}_0^{\mathcal{R}}} \quad \mathbb{E}_{\mathbb{P}_0}[Q\ell(\mathbf{w}, \omega)], \tag{52}$$

where $\mathcal{Q}_0^{\mathcal{R}} = \{Q \in \mathcal{L}^q(\Omega) \,|\, Q \geq 0,\ \mathbb{E}_{\mathbb{P}_0}[Q] = 1,\ \mathcal{R}^*(Q) < \infty\}$. Here the Radon-Nikodym theorem implies that for any probability measure $\mathbb{P}$ absolutely continuous w.r.t. $\mathbb{P}_0$, the random variable $Q \in \mathcal{Q}_0^{\mathcal{R}}$ is the Radon-Nikodym density, i.e., $Q = \dfrac{d\mathbb{P}}{d\mathbb{P}_0}$. Thus the set $\mathcal{Q}_0^{\mathcal{R}}$ has a one-to-one correspondence with the set of probability measures

$$\mathcal{P}^{\mathcal{R}} = \left\{ \mathbb{P} \ll \mathbb{P}_0 \,|\, \mathcal{R}^* \left(\frac{d\mathbb{P}}{d\mathbb{P}_0} \right) < \infty \right\}. \tag{53}$$

and problem (52) can be equivalently rewritten as follows

$$\min_{\mathbf{w} \in \mathcal{W}} \sup_{\mathbb{P} \in \mathcal{P}^{\mathcal{R}}} \quad \mathbb{E}_{\mathbb{P}}[\ell(\mathbf{w}, \omega)]. \tag{54}$$

As Subsection 3.3 mentions, $\mathcal{R}^*$ can be interpreted as a "measure of distance" between two elements from $\mathcal{L}^q(\Omega)$. Hence when the risk is positively homogeneous and monotone, $\mathcal{R}^*$ may serve as a "statistical divergence" between two probability measures $\mathbb{P}$ and $\mathbb{P}_0$. Indeed, by item (R1*) of Proposition 3.3, $\mathcal{R}^*(Q) \geq \mathcal{R}^*(1) = 0$ for every $Q \in \mathcal{L}^q(\Omega)$, where 0 is the minimum value of $\mathcal{R}^*$ hence, whenever $\mathbb{P} = \mathbb{P}_0$, $\mathcal{R}^*\left(\frac{d\mathbb{P}}{d\mathbb{P}_0}\right) = 0$. This observation allows establishing a link between the theory of risk and modern distributionally robust optimization.

Note that the set $\mathcal{P}^{\mathcal{R}}$ considers all probability measures $\mathbb{P}$ having a finite "distance" to $\mathbb{P}_0$ induced by $\mathcal{R}^*$. One may argue that this approach is too conservative, as one may be interested in probability measures $\mathbb{P}$ in the proximity of $\mathbb{P}_0$. This suggests considering sets

$$\mathcal{P}^{\mathcal{R}}_{\tau} = \left\{\mathbb{P} \ll \mathbb{P}_0 \,|\, \mathcal{R}^*\left(\frac{d\mathbb{P}}{d\mathbb{P}_0}\right) \leq \tau\right\}, \quad \tau > 0. \tag{55}$$

Replacing $\mathcal{P}^{\mathcal{R}}$ with $\mathcal{P}^{\mathcal{R}}_{\tau}$ in (54) and relying on (38), problem (54) is as follows

$$\min_{\mathbf{w}\in\mathcal{W}} \quad \mathcal{R}_{\tau}(\ell(\mathbf{w},\omega)), \tag{56}$$

where $\mathcal{R}_{\tau}(\ell(\mathbf{w},\omega)) = \inf_{\lambda>0} \lambda[\mathcal{R}(\lambda^{-1}\ell(\mathbf{w},\omega)) + \tau]$. Moreover, functional $\mathcal{R}_{\tau}$ is itself positively homogeneous and monotone subregular risk (see Proposition 3.8).

In modern distributionally robust optimization (cf. Shapiro [2017]), the decision-maker first selects $\mathcal{R}^*$ and then solves (56). This is equivalent to choosing the risk $\mathcal{R}$ and solving (56). Moreover, the option of choosing a known risk first may be beneficial, as there might exist an efficient way of optimizing it through regret. Proposition 3.9 implies that (56) can be rewritten as

$$\min_{\mathbf{w}\in\mathcal{W},\ C\in\mathbb{R}} \quad C + \mathcal{V}_{\tau}(\ell(\mathbf{w},\omega) - C). \tag{57}$$

Indeed, formulation (57) is usually more computationally efficient than (56).

Coming back to the formulation (51), instead of sets $\mathcal{Q}^{\mathcal{R}}$ one may consider sets

$$\mathcal{Q}^{\mathcal{R}}_{\tau} = \{Q \in \mathcal{L}^q(\Omega) \,|\, \mathbb{E}_{\mathbb{P}_0}[Q] = 1,\ \mathcal{R}^*(Q) \leq \tau\}$$

and solve (56), where the subregular risk is positively homogeneous but no longer a monotone functional. This perspective was used in Peng et al. [2024] to construct the extended φ–Divergence-based Quadrangle.

Example A. The first is an interpretation of Markowitz portfolio optimization [Markowitz, 1952], Large Margin Distribution Machine [Zhang and Zhou, 2014], and least squares regression as relaxed robust loss minimization. For the first example, we define the uncertainty set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}$ of random variables Q as a Euclidean ball of radius $\sqrt{\tau}$ centered at 1 with expected value of 1 :

$$\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau} = \{Q \in \mathcal{L}^p : \mathbb{E}[Q] = 1, \mathbb{E}[\varphi(Q)] \leq \tau\}, \quad \varphi(x) = (x-1)^2. \tag{58}$$

Set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}$ appears in portfolio optimization, classification, and regression problems. Consider a random portfolio loss $\ell(\mathbf{w}) = \mathbf{w}^T\mathbf{r}$, where $\mathbf{w} \in \mathbb{R}^d$ is a vector of portfolio weights and $\mathbf{r}$ is a random vector of negative asset returns. Let $\mathbf{1} = (1,\ldots,1)^\top \in \mathbb{R}^d$. Then the following two problems have the same optimal objective function value and the same set of solution vectors:

Markowitz portfolio optimization

$$\min_{\mathbf{1}^\top\mathbf{w}=1} \mathbb{E}[\ell(\mathbf{w})] + \sqrt{\tau}\sigma(\ell(\mathbf{w})), \tag{59}$$

Robust expected loss minimization

$$\min_{\mathbf{1}^\top\mathbf{w}=1} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[QX(\mathbf{w})]\,. \tag{60}$$

Problem (60) is the robust version of the expected loss minimization problem $\min_{\mathbf{1}^\top\mathbf{w}=1} \mathbb{E}[X(\mathbf{w})]$.

The following is an interpretation of the Large Margin Distribution Machine classification algorithm as a robust optimization. Consider an attribute (random vector of features) $\mathbf{X}$, label Y, and decision vector $\mathbf{w}$. The margin is defined by $L(\mathbf{w},b) = Y(\mathbf{w}^\top\mathbf{X} - b)$. Denote by $\gamma(\mathbf{w})$ a regularization function. The following two problems have the same optimal objective function value and the same set of solution vectors:

Large Margin Distribution Machine

$$\min_{\mathbf{w},b} \mathbb{E}[-L(\mathbf{w},b)] + \sqrt{\tau}\sigma(-L(\mathbf{w},b)) + \gamma(\mathbf{w}), \tag{61}$$

Robust expected margin maximization

$$\min_{\mathbf{w},b} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[-QL(\mathbf{w},b)] + \gamma(\mathbf{w})\,. \tag{62}$$

Problem (62) is the regularized robust version of the expected margin maximization problem $\max_{\mathbf{w}} \mathbb{E}[L(\mathbf{w},b)]$.

The following is an interpretation of least squares regression as robust optimization. Consider a regressant Y, a vector of regressors $\mathbf{X} = (X_1,\ldots,X_n)$, a class of functions $\mathcal{C}$, and an intercept $c \in \mathbb{R}$. The regression residual is defined by $Z_f = Y - f(\mathbf{X}) - c$, and the residual without the intercept C is defined by $\overline{Z}_f = Y - f(\mathbf{X})$. The following two problems have the same optimal solution (f, c):

Least squares regression

$$\min_{f\in\mathcal{C},c\in\mathbb{R}} ||Z_f||_2\ , \tag{63}$$

Deviation minimization

$$\min_{f\in\mathcal{F}} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[Q\overline{Z}_f] - \mathbb{E}[\overline{Z}_f] \tag{64}$$

$$\text{calculate } c = \mathbb{E}[\overline{Z}_f]\ . \tag{65}$$

The interpretation of (60) and (64) as a robust optimization is obtained from the dual representation of risk and deviation in the Extended Pearson Divergence Quadrangle

$$\begin{aligned}
\mathcal{R}_{\varphi,\tau}(X) &= \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[QX],\\
\mathcal{V}_{\varphi,\tau}(X) &= \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau}} \mathbb{E}[QX],\\
\mathcal{D}_{\varphi,\tau}(X) &= \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[QX] - \mathbb{E}[X],\\
\mathcal{E}_{\varphi,\tau}(X) &= \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau}} \mathbb{E}[QX] - \mathbb{E}[X],\\
\mathcal{S}_{\varphi,\tau}(X) &= \mathbb{E}[X],
\end{aligned}$$

where $\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}$ is defined in (58) and the uncertainty set $\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau}$ of random variables Q removes the condition $\mathbb{E}[Q] = 1$ in (58)

$$\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau} = \{Q \in \mathcal{L}^p : \mathbb{E}[\varphi(Q)] \le \tau\}, \quad \varphi(x) = (x-1)^2. \tag{66}$$

Example B. The next example shows the relation between CVaR optimization [Rockafellar and Uryasev, 2000], ν-support vector machine [Schölkopf et al., 2000], quantile regression [Koenker and Bassett Jr, 1978] and robust optimization. Let $\nu = 1-\alpha$. The equivalence of ν-SVM and CVaR optimization was studied by Gotoh and Takeda [2004]; Takeda and Sugiyama [2008]. Define the uncertainty set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}$

$$\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau} = \{Q \in \mathcal{L}^p \mid \mathbb{E}[Q] = 1, \mathbb{E}[\varphi(Q)] \le \tau\}, \quad \varphi(x) = \begin{cases} 0, & x \in [0, \frac{1}{1-\alpha}] \\ +\infty, & \text{otherwise} \end{cases}. \tag{67}$$

$\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}$ will appear in portfolio optimization, classification, and regression problems. Similarly to Example A, in each of the following three pairs of problems, the optimizations on the left and right have the same optimal objective function value and the same set of solution vectors:

CVaR portfolio optimization

$$\min_{\mathbf{1}^T\mathbf{w}=1} \mathrm{CVaR}_\alpha(\ell(\mathbf{w}))\ , \tag{68}$$

Robust loss minimization

$$\min_{\mathbf{1}^T\mathbf{w}=1} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[Q\ell(\mathbf{w})]\ , \tag{69}$$

ν-SVM

$$\min_{\mathbf{w},b} \mathrm{CVaR}_\alpha(-L(\mathbf{w},b)) + \gamma(\mathbf{w}), \tag{70}$$

Robust expected margin maximization

$$\min_{\mathbf{w}} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[-QL(\mathbf{w},b)] + \gamma(\mathbf{w})\ . \tag{71}$$

Quantile regression

$$\min_{f\in\mathcal{C},c\in\mathbb{R}} \mathcal{E}_\alpha(Z_f)\ , \tag{72}$$

Deviation minimization

$$\min_{f\in\mathcal{C}} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[Q\overline{Z}_f] - \mathbb{E}[\overline{Z}_f] \tag{73}$$

$$\text{calculate } c \in \mathrm{VaR}_\alpha[\overline{Z}_f]\ . \tag{74}$$

CVaR portfolio optimization, ν-SVM, and quantile regression are connected by the quantile quadrangle (Example 2, Rockafellar and Uryasev [2013]). The interpretation as robust optimization is obtained from the dual representation, which is presented below together with the primal representation.

Quantile-based Quadrangle

$$\begin{aligned}
\mathcal{R}_\alpha(X) &= \mathrm{CVaR}_\beta(X) = \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[QX],\\
\mathcal{V}_\alpha(X) &= \frac{1}{1-\beta}\mathbb{E}[X_+] = \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau}} \mathbb{E}[QX],\\
\mathcal{D}_\alpha(X) &= \mathrm{CVaR}_\alpha(X) - \mathbb{E}[X] = \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\tau}} \mathbb{E}[QX] - \mathbb{E}[X],\\
\mathcal{E}_\alpha(X) &= \mathbb{E}\Big[\frac{\alpha}{1-\alpha}X_+ + X_-\Big] = \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau}} \mathbb{E}[QX],\\
\mathcal{S}_\alpha(X) &= \mathrm{VaR}_\alpha(X)\ ,
\end{aligned}$$

where the uncertainty set $\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau}$ is defined by

$$\mathcal{Q}^{\mathcal{V}}_{\varphi,\tau} = \{Q \in \mathcal{L}^p : \mathbb{E}[\varphi(Q)] \leq \tau\}, \quad \varphi(x) = \begin{cases} 0, & x \in [0, \frac{1}{1-\alpha}] \\ +\infty, & \text{otherwise} \end{cases}. \tag{75}$$

The robust representations (69), (71), (73) are implied by the dual representations of risk and deviation in the quantile quadrangle.

References

Amir Ahmadi-Javid. Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155:1105–1123, 2012.

Harbir Antil, Drew P. Kouri, Martin-D. Lacasse, and Denis Ridzal, editors. *Frontiers in PDE-Constrained Optimization*, volume 163 of *The IMA Volumes in Mathematics and its Applications*. Springer New York, NY, 2018. ISBN 978-1-4939-8635-4. doi: 10.1007/978-1-4939-8636-1. URL https://doi.org/10.1007/978-1-4939-8636-1.

F. Bach. *Learning Theory from First Principles*. The MIT Press, Cambridge, MA, December 2024. ISBN 9780262049443.

Aharon Ben-Tal and Marc Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007. doi: 10.1111/j.1467-9965.2007.00311.x. URL https://doi.org/10.1111/j.1467-9965.2007.00311.x.

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, Princeton, NJ, 2009.

D. Bertsimas and I. Paskov. Stable regression: On the power of optimization over randomization. *Journal of Machine Learning Research*, 21(230):1–25, 2020. URL http://jmlr.org/papers/v21/19-408.html.

Dimitris Bertsimas, Dessislava Pachamanova, and Melvyn Sim. Robust linear optimization under general norms. *Operations Research Letters*, 39(4):281–286, 2011. doi: 10.1016/j.orl.2011.05.009.

So Yeon Chun, Alexander Shapiro, and Stan Uryasev. Conditional value-at-risk and average value-at-risk: Estimation and asymptotics. *Operations Research*, 60(4):739–756, 2012. doi: 10.1287/opre.1120.1072. URL https://doi.org/10.1287/opre.1120.1072.

Paul Dommel and Alois Pichler. Convex Risk Measures based on Divergence. Papers 2003.07648, arXiv.org, March 2020. URL https://ideas.repec.org/p/arx/papers/2003.07648.html.

H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support Vector Regression Machines. *Advances in Neural Information Processing Systems*, 8:155–161, 1996.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106 (494):746–762, 2011. doi: 10.1198/jasa.2011.r10138. URL https://doi.org/10.1198/jasa.2011.r10138.

Alex Golodnikov, Viktor Kuzmenko, and Stan Uryasev. Cvar regression based on the relation between cvar and mixed-quantile quadrangles. *Journal of Risk and Financial Management*, 12(3), 2019. ISSN 1911-8074. URL https://www.mdpi.com/1911-8074/12/3/107.

Jun-ya Gotoh and Akiko Takeda. A linear classification model based on conditional geometric score. *Institute of Policy and Planning Sciences, Discussion Paper Series*, 2004.

M. G. Kendall. Regression, structure and functional relationship. part i. *Biometrika*, 38(1/2):11–25, 1951. ISSN 00063444. URL http://www.jstor.org/stable/2332313.

M. G. Kendall. Regression, structure and functional relationship. part ii. *Biometrika*, 39(1/2):96–108, 1952. ISSN 00063444, 14643510. URL http://www.jstor.org/stable/2332467.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

P. A. Krokhmal. Higher moment coherent risk measures. *Quantitative Finance*, 7(4):373–387, 2007. doi: 10.1080/14697680701458307. URL https://doi.org/10.1080/14697680701458307.

V. Kuzmenko. A new family of expectiles and its properties. *Cybernetics and Computer Technologies*, 3:43–58, 2020. URL https://doi.org/10.34229/2707-451X.20.3.5.

Nicholas Lambert, David M. Pennock, and Yoav Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138, Chicago, IL, 2008.

A. Mafusalov and S. Uryasev. CVaR (Superquantile) Norm: Stochastic Case. *European Journal of Operational Research*, 249:200—208, 2016.

A. Malandii and S. Uryasev. Support vector regression: Risk quadrangle framework. *arXiv:2212.09178 [stat.ML]*, 2022. URL https://doi.org/10.48550/arXiv.2212.09178.

A. Malandii and S. Uryasev. Biased mean quadrangles and applications. *Working Paper*, 2024.

A. Malandii, V. Kuzmenko, and S. Uryasev. Expectile Risk Quadrangles and Applications. *Journal of Risk*, 27 (1):1–35, October 2024. doi: 10.21314/JOR.2024.010. First published: 11 December 2024.

Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 00221082, 15406261. URL http://www.jstor.org/stable/2975974.

W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, 1987. URL https://doi.org/10.2307/1911031.

K. Pavlikov and S. Uryasev. CVaR Norm and Applications in Optimization. *Optimization Letters*, 8(7):1999––2020, 2014.

Cheng Peng, Anton Malandii, and Stan Uryasev. Risk quadrangle and robust optimization based on φ-divergence. *Working Paper*, 2024.

R. T. Rockafellar and S. Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–42, 2000.

R. T. Rockafellar and S. Uryasev. Conditional Value-at-Risk for General Loss Distributions. *Journal of Banking an Finance*, 26:1443–1471, 2002.

R. T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization, and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1):33–53, 2013. URL https://doi.org/10.1016/j.sorms.2013.03.001.

R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Risk Tuning with Generalized Linear Regression. *Mathematics of Operations Research*, 33(3):712–729, 2008.

R. T. Rockafellar, J. O. Royset, and S. I. Miranda. Superquantile regression with applications to buffered reliability, uncertainty quantification and conditional value-at-risk. *European J. Operations Research*, 234(1): 140–154, 2014. URL https://doi.org/10.1016/j.ejor.2013.10.046.

R. Tyrrell Rockafellar. Distributional robustness, stochastic divergences, and the quadrangle of risk. *Computational Management Science*, 21:34, 2024. doi: 10.1007/s10287-024-00516-z.

R. Tyrrell Rockafellar and Johannes O. Royset. Random variables, monotone relations, and convex analysis. *Mathematical Programming*, 148:297–331, 2014. doi: 10.1007/s10107-014-0801-1. URL https://doi.org/10.1007/s10107-014-0801-1.

R. Tyrrell Rockafellar and Johannes O. Royset. Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity. *SIAM Journal on Optimization*, 25(2):1179–1208, 2015.

Johannes O. Royset. Risk-adaptive approaches to stochastic optimization: A survey. *arXiv preprint arXiv:2212.00856*, 2022. URL https://arxiv.org/abs/2212.00856.

Bernhard Schölkopf, Alex J Smola, Robert C Williamson, and Peter L Bartlett. New support vector algorithms. *Neural computation*, 12(5):1207–1245, 2000.

B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett. New Support Vector Algorithms. *Neural Computation*, 12(5):1207–1245, 2000.

A. Shapiro. On kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, 38 (1), 2012. URL https://doi.org/10.1287/moor.1120.0563.

Alexander Shapiro. Distributionally robust stochastic programming. *SIAM Journal on Optimization*, 27(4): 2258–2275, 2017.

Akiko Takeda and Masashi Sugiyama. ν-support vector machine as conditional value-at-risk minimization. In *Proceedings of the 25th international conference on Machine learning*, pages 1056–1063, 2008.

V. Vapnik, S. E. Golowich, and A. Smola. Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. *Advances in Neural Information Processing Systems*, 9:281–287, 1996.

Robert Williamson and Aditya Menon. Fairness risk measures. In *International conference on machine learning*, pages 6786–6797. PMLR, 2019.

Teng Zhang and Zhi-Hua Zhou. Large margin distribution machine. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 313–322, 2014.