European Journal of Operational Research 249 (2016) 200–208

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

CVaR (superquantile) norm: Stochastic case[☆]

Alexander Mafusalov, Stan Uryasev[*]

Risk Management and Financial Engineering Lab Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, Gainesville, FL 32611, US

A R T I C L E   I N F O

Article history:
Received 17 January 2014
Accepted 26 September 2015
Available online 5 October 2015

Keywords:
CVaR norm
L-p norm
Superquantile
Risk quadrangle
Linear regression

A B S T R A C T

The concept of Conditional Value-at-Risk (CVaR) is used in various applications in uncertain environment. This paper introduces CVaR (superquantile) norm for a random variable, which is by definition CVaR of absolute value of this random variable. It is proved that CVaR norm is indeed a norm in the space of random variables. CVaR norm is defined in two variations: scaled and non-scaled. L-1 and L-infinity norms are limiting cases of the CVaR norm. In continuous case, scaled CVaR norm is a conditional expectation of the random variable. A similar representation of CVaR norm is valid for discrete random variables. Several properties for scaled and non-scaled CVaR norm, as a function of confidence level, were proved. Dual norm for CVaR norm is proved to be the maximum of L-1 and scaled L-infinity norms. CVaR norm, as a Measure of Error, is related to a Regular Risk Quadrangle. Trimmed L1-norm, which is a non-convex extension for CVaR norm, is introduced analogously to function L-p for $p < 1$. Linear regression problems were solved by minimizing CVaR norm of regression residuals.

1. Introduction

The concept of Conditional Value-at-Risk (CVaR) is widely used in risk management and various applications in uncertain environment. This paper introduces a concept of *CVaR norm* in the space of random variables. CVaR norm in $\mathbb{R}^n$ was introduced in (Pavlikov & Uryasev, 2014) and developed in (Gotoh & Uryasev, 2013), and is a particular case of general error measures introduced and developed by Rockafellar, Uryasev, and Zabarankin (2008). The term "superquantile", free from dependence on financial terminology, can be used as a neutral alternative name for "CVaR", like it was done in Rockafellar and Royset (2010); Rockafellar and Uryasev (2013). For the similar reason the alternative name "superquantile norm" is proposed to replace "CVaR norm" when desired. For the sake of consistency within the paper and with the earlier study (Pavlikov & Uryasev, 2014), this paper will use mostly the "CVaR norm" term.

This section provides a short introduction in the CVaR norm in $\mathbb{R}^n$ and shows the relation with the CVaR norm in the space of random variables. This paper is motivated by applications of norms in optimization. We consider norms in $\mathbb{R}^n$ and in the space of random variables. We use symbols $\mathbf{x}$ and x_i for a vector and an ith vector component in $\mathbb{R}^n$, i.e. $\mathbf{x} = (x_1, \ldots, x_n)$. We use symbol X for a random variable.

l_p norms are broadly used in $\mathbb{R}^n$, and L_p norms are considered in the space of random variables. For $p \in [1, \infty]$, the norms l_p and L_p are defined as follows[1]:

$$l_p(\mathbf{x}) = \left(\frac{1}{n}\sum_{i=1}^{n}|x_i|^p\right)^{1/p}, \quad L_p(X) = (E|X|^p)^{1/p},$$

where E is the expectation sign. The most popular cases are $p = 1, 2, \infty$, i.e.,

- $l_1(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}|x_i|, \quad L_1(X) = E|X|;$
- $l_\infty(\mathbf{x}) = \max_{i=1,\ldots,n}|x_i|, \quad L_\infty(X) = \sup|X|;$
- $l_2(\mathbf{x}) = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)^{1/2}, \quad L_2(X) = (EX^2)^{1/2}.$

It is known that $l_1(\mathbf{x}) \le l_2(\mathbf{x}) \le l_\infty(\mathbf{x})$ and $L_1(X) \le L_2(X) \le L_\infty(X)$, which follow from $l_p(\mathbf{x}) \le l_q(\mathbf{x})$ and $L_p(X) \le L_q(X)$ for $p < q$, see (e.g. Brezis, 2010, page 118).

The other family of norms, CVaR norm for $\mathbb{R}^n$ was defined in Pavlikov and Uryasev (2014) and studied in Gotoh and Uryasev (2013). According to Pavlikov and Uryasev (2014, Definition 3), the *non-scaled*

[1] Note that the classic definition for l_p norm is $l_p(\mathbf{x}) = \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p}$ does not satisfy inequality $l_p(\mathbf{x}) \le l_q(\mathbf{x})$ for $p < q$. This paper uses an equivalent scaled version of this norm $l_p(\mathbf{x}) = \left(\frac{1}{n}\sum_{i=1}^{n}|x_i|^p\right)^{1/p}$, which satisfies that inequality. L_p norm is commonly defined as $L_p(f) = \|f\|_p \equiv \left(\int_S |f|^p \, d\mu\right)^{1/p}$, where S is a considered space. It is known (e.g. Brezis, 2010, page 118) that for $\|\cdot\|_p$ and $\|\cdot\|_q$ norms inequality $\|f\|_p \le \mu(S)^{\frac{1}{p}-\frac{1}{q}}\|f\|_q$ holds for $1 \le p \le q \le \infty$, where S is a considered space and $\mu(S)$ is the measure of the space S. When S is a probability space, $\mu(S) = 1$ and inequality $L_p(X) \le L_q(X)$ holds for $1 \le p \le q \le \infty$, where $L_p(X) = (E|X|^p)^{1/p}$.

[☆] This paper was processed by Guest-Editor prof. Ruediger Schultz for an uncompleted EJOR Feature Cluster on Stochastic Optimization.

[*] Corresponding author. Tel.: +1 352 213 3457.
E-mail addresses: mafusalov@ufl.edu (A. Mafusalov), uryasev@ufl.edu (S. Uryasev).

http://dx.doi.org/10.1016/j.ejor.2015.09.058
0377-2217/Published by Elsevier B.V.

CVaR norm with parameter α, or $\langle\langle \mathbf{x} \rangle\rangle_\alpha$, is defined on $\mathbf{x} \in \mathbb{R}^n$ as a sum of absolute values of biggest $n(1-\alpha)$ components. If $n(1-\alpha)$ is not an integer, then $\langle\langle \mathbf{x} \rangle\rangle_\alpha$ is defined as a weighted average of two norms $\langle\langle \mathbf{x} \rangle\rangle_{\alpha_1}$ and $\langle\langle \mathbf{x} \rangle\rangle_{\alpha_2}$ for closest values α_1, α_2 such that $n(1-\alpha_1)$ and $n(1-\alpha_2)$ are integers.

A similar norm, called D-norm, was introduced in Bertsimas, Pachamanova, and Sim (2004, Section 3) in a different way. D-norm is defined as a maximum of a sum of weighted absolute values of vector components. The maximization is performed over all sets of indexes for components in the sum, with a constraint on cardinality. For $\alpha \in \left[0, \frac{n-1}{n}\right]$, the CVaR norm $\langle\langle \mathbf{x} \rangle\rangle_\alpha$ coincides with the D-norm $|||\mathbf{x}|||_p$ with parameter p defined by $p = n(1-\alpha)$, see Pavlikov and Uryasev (2014, Proposition 3.4). (Bertsimas et al., 2004, Proposition 2) find a dual norm to D-norm; this result was generalized with Item 2 of Proposition 2.1 of this paper for the stochastic case.

Both CVaR norm and D-norm can be viewed as important special cases of Ordered Weighting Averaging (OWA) operators, see Merigó and Yager (2013); Torra and Narukawa (2007); Yager (2010). A subfamily of OWA operators with monotonically non-increasing weights, when implied to the absolute values of the vector, were formalized as norms in Yager (2010). The worst-case averages, corresponding to CVaR, were also studied, e.g., in Ogryczak and Zawadzki (2002); Romeijn, Ahuja, Dempsey, and Kumar (2005).

The paper (Pavlikov & Uryasev, 2014, Definition 1) has also defined *scaled* CVaR norm $\langle\langle \mathbf{x} \rangle\rangle^S_\alpha$. Scaled version calculates average value of components instead of sum: $n(1-\alpha)\langle\langle \mathbf{x} \rangle\rangle^S_\alpha = \langle\langle \mathbf{x} \rangle\rangle_\alpha$. This paper defines (scaled) CVaR norm of a random variable X as an expectation of $|X|$ in its right $(1-\alpha)$-tail. It can be shown that proposed norm is a generalization of $\langle\langle \mathbf{x} \rangle\rangle^S_\alpha$ in a following way. Consider mapping $X(\mathbf{x}) : \mathbb{R}^n \to \mathcal{L}^1(\Omega)$ from Eucledian space of dimension n to the space of L_1-finite random variables. Denote $\mathbf{x} = (x_1, \ldots, x_n)$. Let $X(\mathbf{x})$ be discretely distributed over atoms $x_1, \ldots, x_n$ with equal probabilities $\frac{1}{n}$. Then it is easy to see that $\langle\langle \mathbf{x} \rangle\rangle^S_\alpha = \mathrm{CVaR}_\alpha(|X(\mathbf{x})|) \equiv \langle\langle X(\mathbf{x}) \rangle\rangle^S_\alpha$, see Pavlikov and Uryasev (2014, Definition 2).

This paper also defines *non-scaled* CVaR norm $\langle\langle X \rangle\rangle_\alpha = (1-\alpha)\langle\langle X \rangle\rangle^S_\alpha$, which corresponds to $\langle\langle \mathbf{x} \rangle\rangle_\alpha$ from Pavlikov and Uryasev (2014). Non-scaled version has attractive properties with respect to parameter α, see Items 6, 7 from Section 2.

Risk Quadrangle considers risk $\mathcal{R}(X)$, deviation $\mathcal{D}(X)$, regret $\mathcal{V}(X)$, error $\mathcal{E}(X)$ and statistic $\mathcal{S}(X)$, related with a set of equations, called The General Relationships, see Rockafellar and Uryasev (2013, Diagram 3). If a functional satisfies a corresponding set of axioms, it is called regular, see Rockafellar and Uryasev (2013, Section 3). It can be proved that if $\mathcal{R}(X)$ is a coherent and regular Measure of Risk, then $\mathcal{R}(|X|)$ is both a norm and a regular Measure of Error, however, this proof is beyond the scope of this paper. This paper proves that $\langle\langle X \rangle\rangle_\alpha$ is a regular Measure of Error and finds the corresponding functions $\mathcal{R}(X)$, $\mathcal{D}(X)$, $\mathcal{V}(X)$ and $\mathcal{S}(X)$ in risk quadrangle related to the Measure of Error $\mathcal{E}(X) = \langle\langle X \rangle\rangle_\alpha$, see Section 2.

Item 3 from Proposition 2.1 can be viewed as a stochastic generalization of Hall and Tymoczko (2012, Lemma 1). Paper (Hall & Tymoczko, 2012) considers functions $\Sigma_j(\mathbf{x})$ on nonnegative orthant $\mathbb{R}^n_+$, corresponding to $\mathbb{R}^n_+$ reduction of special cases of CVaR norm. Paper (Hall & Tymoczko, 2012) relies on majorization theory, see (Marshall, Olkin, & Arnold, 2011), which is generalized for the stochastic case with a concept of stochastic dominance, see Ogryczak and Ruszczynski (2002); Dentcheva and Ruszczynski (2003, e.g.).

This paper considers also non-convex functions closely related to CVaR norm. In deterministic case, by definition, CVaR norm is the average of biggest $(1-\alpha)n$ absolute values of components of a vector. The *trimmed L1-norm* is also known as a trimmed sum of absolute deviations estimator for LTA regression, see (Hawkins & Olive, 1999; Bassett Jr, 1991; Hössjer, 1994, e.g.). L1-based LTA regression is introduced similarly to the more common L2-based Least-Trimmed-Squares (LTS) regression, see Zabarankin and Uryasev (2014b, e.g.). Trimmed L1-norm is denoted here by t_α, it is the average of smallest αn absolute values of components of a vector. Calculation formulas and mathematical properties for $t_\alpha(\mathbf{x})$ in Eucledian space and $T_\alpha(X)$ in the space of random variables are considered in Section 3. Trimmed L1-norm is also related to the sparse optimization, similar to functions l_p for $0 < p < 1$, see Ge, Jiang, and Ye (2011). Note that the constraint on trimmed L1-norm directly specifies sparsity of the solution vector (see Item 6 of Section 3), compared to "indirect" sparsity specification with l_p function. This paper also provides an illustration for t_α in Euclidean space.

Paper (Krzemienowski, 2009, Definition 2) defines conditional average CAVG function. Both average quantile and CVaR are subfamilies of CAVG family, therefore, both $\langle\langle X \rangle\rangle^S_\alpha$ and $T_\alpha(X)$ are subfamilies of $\mathrm{CAVG}_{\beta,\gamma}(|X|)$ function family. Unfortunately, these functions are not convex or concave in general, and are out of the scope of this study, although robust regression applications based on these functions is a promising research direction.

The paper is organized as follows. Section 2 gives a formal definition of CVaR norm in stochastic case and enlists various mathematical of CVaR norm, including that it is indeed a norm and a regular measure of error. CVaR norm is a parametric family of norms with respect to the confidence parameter α, properties of CVaR norm as a function of α are proved. Dual representation of the CVaR norm is derived, and a dual norm to the CVaR norm is defined. A short introduction to the concept of Risk Quadrangle is given. We derive the quadrangle related to the CVaR norm as a measure of error and we prove that this quadrangle is regular. Section 3 defines the *trimmed L1-norm*, both in $\mathbb{R}^n$ and in the space of random variables, and enlists several basic properties. The trimmed L1-norm is an extension of CVaR norm, but it is not actually a norm. Section 4 illustrates properties of CVaR norm with a case study. Section 5 provides concluding remarks and acknowledgements.

2. CVaR (superquantile) norm properties and connection to risk quadrangle

This section gives a formal definition of CVaR norm in stochastic case and proves various properties of the norm. Let us denote $[x]^+ = \max\{0, x\}$, $[x]^- = \max\{0, -x\}$. Consider cumulative distribution function $F_X(x) = P(X \le x)$. If, for a probability level $\alpha \in (0, 1)$, there is a unique x such that $F_X(x) = \alpha$, then this x is called the α-quantile $q_\alpha(X)$. In general, however, the value x is not unique, or may not even exist. There are two boundary values:

$$q^+_\alpha(X) = \inf\{x | F_X(x) > \alpha\}, \quad q^-_\alpha(X) = \sup\{x | F_X(x) < \alpha\}.$$

We will call by the *quantile* the entire interval between the two boundary values,

$$q_\alpha(X) = [q^-_\alpha(X), q^+_\alpha(X)]. \tag{1}$$

We will use notation $\int q_p(X)dp \equiv \int q^-_p(X)dp$, which is reasonable since $\int q^+_p(X)dp = \int q^-_p(X)dp$.

The CVaR norm is defined as follows.

Definition 1. Let X be a random variable with $E|X| < \infty$. Then CVaR (superquantile) norm of X with parameter $\alpha \in [0, 1]$ is defined by

$$\langle\langle X \rangle\rangle^S_\alpha = \mathrm{CVaR}_\alpha(|X|) = \bar{q}_\alpha(|X|).$$

Following the logic of (Pavlikov & Uryasev, 2014), $\langle\langle X \rangle\rangle^S_\alpha$ is called *scaled* CVaR norm, while Definition 2 introduces $\langle\langle X \rangle\rangle_\alpha$, corresponding to *non-scaled* CVaR norm for $\mathbb{R}^n$ in (Pavlikov & Uryasev, 2014, Definition 3). By default we call by CVaR norm the function $\langle\langle X \rangle\rangle^S_\alpha$. The second version of the norm, *non-scaled* CVaR norm, is defined as follows.

Definition 2. Let X be a random variable with $E|X| < \infty$. Then non-scaled CVaR (superquantile) norm of X with parameter $\alpha \in [0, 1)$ is defined as follows:

$$\langle\langle X \rangle\rangle_\alpha = (1-\alpha)\langle\langle X \rangle\rangle^S_\alpha.$$

Note that by continuity $\langle\langle X\rangle\rangle_1 \equiv \lim_{\alpha\to 1}\langle\langle X\rangle\rangle_\alpha = 0$. That is, for $\alpha = 1$, the function $\langle\langle X\rangle\rangle_\alpha$ is not a norm.

Recall some properties of CVaR (e.g. Rockafellar & Uryasev, 2013, page 20), see also (Rockafellar & Uryasev, 2000; 2002):

- Positive homogeneity:

$$\mathrm{CVaR}_\alpha(\lambda X) = \lambda \mathrm{CVaR}_\alpha(X), \quad \text{for } \lambda > 0. \tag{2}$$

- Subadditivity:

$$\mathrm{CVaR}_\alpha(X+Y) \le \mathrm{CVaR}_\alpha(X) + \mathrm{CVaR}_\alpha(Y). \tag{3}$$

- Monotonicity:

$$\mathrm{CVaR}_\alpha(X) \le \mathrm{CVaR}_\alpha(Y), \quad \text{for } X \le Y. \tag{4}$$

Properties and representations of CVaR (superquantile) norm

1. $\langle\langle X\rangle\rangle_0^S = L_1(X)$, $\langle\langle X\rangle\rangle_1^S = L_\infty(X)$.
(Follows from Definition 1).
2. $\langle\langle X\rangle\rangle_\alpha^S$ is a norm on $\mathcal{L}^1(\Omega)$ space of random variables.
(Norm positive homogeneity and convexity follow from the positive homogeneity and convexity of the CVaR and the absolute value $\|\cdot\|$, whereas $\mathrm{CVaR}_\alpha(|X|) = 0 \Leftrightarrow X \equiv 0$ is obvious.)
3. The representations

$$\langle\langle X\rangle\rangle_\alpha^S = \min_c \left\{ c + \frac{1}{1-\alpha} E[|X| - c]^+ \right\}. \tag{5}$$

$$\langle\langle X\rangle\rangle_\alpha^S = \frac{1}{1-\alpha}\int_\alpha^1 q_p(|X|)dp. \tag{6}$$

$$\langle\langle X\rangle\rangle_\alpha^S = E\big(|X| \,\big|\, |X| > q_\alpha(|X|)\big), \quad \text{if } X \text{ is a continuous random variable}, \tag{7}$$

follow directly from CVaR calculation formulas, see (e.g. Rockafellar & Uryasev, 2000, formulas (3),(5)), (e.g. Rockafellar & Uryasev, 2002, Definitions 3,4, Theorem 10). Also,

$$\langle\langle X\rangle\rangle_\alpha^S = \mathrm{CVaR}_{(1+\alpha)/2}(Y), \quad \text{where } Y = \begin{cases} X, & \text{with probability } \frac{1}{2}; \\ -X, & \text{with probability } \frac{1}{2}; \end{cases} \tag{8}$$

see the proof in Appendix A.
4. Let X be a discrete random variable, i.e., it takes values $\{x_i\}_{i=1}^N$ with positive probabilities $\{p_i\}_{i=1}^N$, where $\sum_{i=1}^N p_i = 1$ and $N \in \mathbb{N}\cup\infty$ (N also can be ∞). Let us denote by $\{|x|_{(i)}\}_{i=1}^N$ an ordered sequence $\{|x_i|\}_{i=1}^N$, i.e., $|x|_{(i)} \le |x|_{(i+1)}$.[2] We also denote by $\{|p|_{(i)}\}_{i=1}^N$ a corresponding to the $\{|x|_{(i)}\}_{i=1}^N$ sequence of probabilities from the $\{p_i\}_{i=1}^N$. Then
(a) for $\alpha = 1$,

$$\langle\langle X\rangle\rangle_1^S = \begin{cases} |x|_{(N)}, & \text{for } N < \infty; \\ \lim_{i\to\infty} |x|_{(i)}, & \text{for } N = \infty; \end{cases}$$

(b) for $\alpha_j = \sum_{i=1}^j |p|_{(i)}$ and $j < N$, $j \in \mathbb{Z}_+ \equiv \mathbb{N}\cup\{0\}$,

$$\langle\langle X\rangle\rangle_{\alpha_j}^S = \begin{cases} \dfrac{1}{1-\alpha_j}\displaystyle\sum_{i=j+1}^N |x|_{(i)}|p|_{(i)}, & \text{for } N < \infty; \\ \dfrac{1}{1-\alpha_j}\displaystyle\sum_{i=j+1}^\infty |x|_{(i)}|p|_{(i)}, & \text{for } N = \infty; \end{cases}$$

(c) for $\alpha_j < \alpha < \alpha_{j+1}$,

$$\langle\langle X\rangle\rangle_\alpha^S = (1-\lambda)\frac{1-\alpha_j}{1-\alpha}\langle\langle X\rangle\rangle_{\alpha_j}^S + \lambda\frac{1-\alpha_{j+1}}{1-\alpha}\langle\langle X\rangle\rangle_{\alpha_{j+1}}^S,$$

where $\lambda = \dfrac{\alpha - \alpha_j}{\alpha_{j+1} - \alpha_j}$.

See the proof in Appendix A.
5. $\langle\langle X\rangle\rangle_\alpha^S$ is a continuous increasing function of α.
(Follows from the fact that CVaR is a continuous increasing function of α, see Rockafellar and Uryasev (2002, Proposition 13).)
6. $\langle\langle X\rangle\rangle_\alpha$ is a concave and decreasing function of α.
(The integral $\int_\alpha^1 q_p(|X|)dp$ is a convex function of α as shown in Ogryczak and Ruszczynski (2002, page 6), and $\langle\langle X\rangle\rangle_\alpha = E|X| - \int_\alpha^1 q_p(|X|)dp$.)
7. $\langle\langle X\rangle\rangle_\alpha$ and $(1-\alpha)\mathrm{CVaR}_\alpha(X)$ are piecewise-linear functions of α for a discretely distributed random variable X.
(Since quantile function for discretely distributed random variable is a step function, then its integral is piecewise-linear, and $\langle\langle X\rangle\rangle_\alpha = E|X| - \int_\alpha^1 q_p(|X|)dp$.)
8. The norm $\langle\langle X\rangle\rangle_\alpha^S$ generates a Banach space for $\alpha \in [0,1]$.
(From the definition of Banach space[3] follows that since $L_1(X) = E|X|$ generates[4] a Banach space and $E|X| \le \langle\langle X\rangle\rangle_\alpha^S \le \frac{1}{1-\alpha}E|X|$, then for $\alpha \in [0,1)$ CVaR norm also generates a Banach space, and $\langle\langle X\rangle\rangle_\alpha^S = L_\infty(X)$ for $\alpha = 1$.)
9. $\mathcal{E}(X) = \langle\langle X\rangle\rangle_\alpha^S$ is a regular measure of error.
See the proof in Appendix A.

The asterisk $*$ denotes the dual norm[5] to a norm. Therefore, $\langle\langle Y\rangle\rangle_\alpha^{S*}$ denotes the norm dual to the CVaR norm $\langle\langle X\rangle\rangle_\alpha^S$.

Proposition 2.1. *For $X \in \mathcal{L}^1(\Omega)$ and the norm $\langle\langle\cdot\rangle\rangle_\alpha^S$, the following statements hold:*

1. $\langle\langle X\rangle\rangle_\alpha^S = \sup_{Y\in\mathcal{Y}} EXY$, *where* $\mathcal{Y} = \left\{Y \,\middle|\, |Y| \le \frac{1}{1-\alpha}, E|Y| \le 1\right\} = \left\{Y \,\middle|\, EXY \le \langle\langle X\rangle\rangle_\alpha^S\right\}$ *is a closed convex set.*
2. *The norm* $\langle\langle Y\rangle\rangle_\alpha^{S*} = \max\{E|Y|, (1-\alpha)\sup|Y|\}$ *is dual to the norm* $\langle\langle X\rangle\rangle_\alpha^S$ *for* $\alpha \in (0,1)$.
3. *Problem* $\min_{d\in\mathbb{R}}\langle\langle X-d\rangle\rangle_\alpha^S$ *has the following solution:*

$$\arg\min_d \langle\langle X-d\rangle\rangle_\alpha^S = \frac{1}{2}\big(q_{(1-\alpha)/2}(X) + q_{(1+\alpha)/2}(X)\big),$$

$$\min_d \langle\langle X-d\rangle\rangle_\alpha^S = \frac{1}{1-\alpha}\left(\frac{1+\alpha}{2}\mathrm{CVaR}_{(1-\alpha)/2}(X) + \frac{1-\alpha}{2}\mathrm{CVaR}_{(1+\alpha)/2}(X) - EX\right).$$

Proof.

1. Papers (Rockafellar, Uryasev, & Zabarankin, 2006), (Rockafellar & Uryasev, 2013, (6.9)) proved that $\mathrm{CVaR}_\alpha(X) = \sup_{Q\in\mathcal{Q}} EXQ$, where

$$\mathcal{Q} = \left\{Q \,\middle|\, 0 \le Q \le \frac{1}{1-\alpha}, EQ = 1\right\}.$$

---

[2] Note that ordered sequence $\{|x|_{(i)}\}$ may not exist for some sets $\{x_i\}$. As a counterexample, consider $x_i = 2^{-i}$, for $i = 1,\dots,\infty$.

[3] A Banach space is a vector space $\mathbb{X}$ over $\mathbb{R}$, which is equipped with a norm $\|\cdot\|$ and which is complete with respect to that norm. By definition, completeness means that for every Cauchy sequence $\{x_n\}_{n=1}^\infty$ in $\mathbb{X}$ (i.e., for every $\varepsilon > 0$ exists N such that $\|x_m - x_n\| < \varepsilon$ for all $m, n > N$), there exists an element x in $\mathbb{X}$ such that

$$\lim_{n\to\infty} x_n = x, \quad \text{i.e.,} \quad \lim_{n\to\infty}\|x_n - x\| = 0.$$

[4] We say that norm L generates space $(\mathcal{X}_L, L)$, where $\mathcal{X}_L = \{X | L(X) < \infty\}$.

[5] Let $\mathbb{X}$ be a normed space over $\mathbb{R}$ with norm $\|\cdot\|$ (i.e., $\|X\| \in \mathbb{R}$ for $X \in \mathbb{X}$). Then, the dual (or conjugate) normed space $\mathbb{X}^*$ is defined as the set of all continuous linear functionals from $\mathbb{X}$ into $\mathbb{R}$. For $f \in \mathbb{X}^*$, the dual norm $\|\cdot\|^*$ of f is defined by

$$\|f\|^* = \sup\{|f(x)| : x \in \mathbb{X}, \|x\| \le 1\} = \sup\left\{\frac{|f(x)|}{\|x\|} : x \in \mathbb{X}, x \ne 0\right\}.$$

Denote $\mathcal{Y} = \{Y \mid |Y| \le \frac{1}{1-\alpha}, E|Y| \le 1\}$ and $\mathcal{Y}^* = \{Y \mid EXY \le \langle\langle X\rangle\rangle_\alpha^S\}$. Then

$$\langle\langle X\rangle\rangle_\alpha^S = \sup_{Q\in\mathcal{Q}} E|X|Q \le \sup_{Y\in\mathcal{Y}} EXY \le \sup_{Y\in\mathcal{Y}} E|X||Y| \le \sup_{Q\in\mathcal{Q}} E|X|Q.$$

The first inequality holds since for any $Q \in \mathcal{Q}$ exists $Y = (I(X>0) - I(X<0))Q \in \mathcal{Y}$ such that $|X|Q = XY$. Hence, $\langle\langle X\rangle\rangle_\alpha^S = \sup_{Y\in\mathcal{Y}} EXY$. Note that $\langle\langle X\rangle\rangle_\alpha^S = \sup_{Y\in\mathcal{Y}^*} EXY$. Then, $\mathcal{Y}^*$ is a closed convex hull of $\mathcal{Y}$. $\mathcal{Y}$ is closed convex as an intersection of two closed convex sets $\{Y \mid |Y| \le \frac{1}{1-\alpha}\}$ and $\{Y \mid E|Y| \le 1.\}$. That is, $\mathcal{Y}^* = \mathcal{Y}$.

2. Item 1 implies that $\mathcal{Y} = \{Y \mid |Y| \le \frac{1}{1-\alpha}, E|Y| \le 1\}$ is a unit ball for the dual norm $\langle\langle Y\rangle\rangle_\alpha^{S*} = \sup_{X\neq 0} \frac{EXY}{\langle\langle X\rangle\rangle_\alpha^S}$. Then, the unit sphere $\langle\langle Y\rangle\rangle_\alpha^{S*} = 1$ for the dual norm is the set

$$\left\{Y \,\middle|\, \sup|Y| = \frac{1}{1-\alpha}, E|Y| \le 1\right\} \cup \left\{Y \,\middle|\, \sup|Y| \le \frac{1}{1-\alpha}, E|Y| = 1\right\}.$$

Therefore, the dual norm equals $\langle\langle Y\rangle\rangle_\alpha^{S*} = \max\{E|Y|, (1-\alpha)\sup|Y|\}$.

3. Note that for $c \ge 0$ and arbitrary $x, d \in \mathbb{R}$, we have

$$[|x-d|-c]^+ = \begin{cases} d-c-x, & \text{for } x \le d-c; \\ 0, & \text{for } x \in [d-c, d+c]; \\ & \equiv [d-c-x]^+ + [x-(d+c)]^+ = \\ x-(d+c), & \text{for } x \ge d+c. \end{cases}$$
$$= [x-(d-c)]^+ + d - c - x + [x-(d+c)]^+.$$

Let $c_1 = d - c$ and $c_2 = d + c$, then this relationship yields

$$(1-\alpha)c + [|x-d|-c]^+ = (1-\alpha)\frac{c_2-c_1}{2} + [x-c_1]^+ + c_1 - x + [x-c_2]^+ = \frac{1+\alpha}{2}c_1 + [x-c_1]^+ + \frac{1-\alpha}{2}c_2 + [x-c_2]^+ - x. \quad (9)$$

An optimal value c in $\langle\langle X\rangle\rangle_\alpha^S = \min_{c\in\mathbb{R}}\left\{c + (1-\alpha)^{-1}E[|X|-c]^+\right\}$ is the quantile $q_\alpha(|X|)$ and hence nonnegative. Thus, c can be restricted in this optimization problem to be nonnegative, and, applying (9),

$$\min_{d\in\mathbb{R}} \langle\langle X-d\rangle\rangle_\alpha^S = \min_{c\ge 0, d\in\mathbb{R}} \left\{c + (1-\alpha)^{-1}E[|X-d|-c]^+\right\}$$
$$= \frac{1}{1-\alpha}\min_{c_1,c_2\in\mathbb{R}}\left\{\frac{1+\alpha}{2}c_1 + E[X-c_1]^+ + \frac{1-\alpha}{2}c_2 + E[X-c_2]^+ - EX\right\}$$
$$= \frac{1}{1-\alpha}\left(\frac{1+\alpha}{2}\mathrm{CVaR}_{(1-\alpha)/2}(X) + \frac{1-\alpha}{2}\mathrm{CVaR}_{(1+\alpha)/2}(X) - EX\right),$$

where optimal c_1 and c_2 are determined by $q_{(1-\alpha)/2}(X)$ and $q_{(1+\alpha)/2}(X)$, respectively, and yield optimal $d = \frac{1}{2}(q_{(1-\alpha)/2}(X) + q_{(1+\alpha)/2}(X))$. □

Risk Quadrangle (Rockafellar & Uryasev, 2013) relates risk $\mathcal{R}(X)$, deviation $\mathcal{D}(X)$, regret $\mathcal{V}(X)$, error $\mathcal{E}(X)$ and statistic $\mathcal{S}(X)$ with the following equations (Rockafellar & Uryasev, 2013, Diagram 3):

$$\mathcal{V}(X) = EX + \mathcal{E}(X), \mathcal{R}(X) = EX + \mathcal{D}(X), \quad (10)$$

$$\mathcal{D}(X) = \min_C\{\mathcal{E}(X-C)\}, \mathcal{R}(X) = \min_C\{C + \mathcal{V}(X-C)\}, \quad (11)$$

$$\mathcal{S}(X) = \arg\min_C\{\mathcal{E}(X-C)\} = \arg\min_C\{C + \mathcal{V}(X-C)\}. \quad (12)$$

Following the paper (Rockafellar & Uryasev, 2013, Section 3), we consider the $\mathcal{L}^2(\Omega)$ space of random variables with finite second moment, $EX^2 < \infty$, which implies finite first moment, $E|X| < \infty$. The natural ("strong") convergence in $\mathcal{L}^2(\Omega)$ of a sequence of random variables X^k to a random variable X is characterized as follows:

$$\mathcal{L}^2\text{-}\lim_{k\to\infty} X^k = X \Leftrightarrow \lim_{k\to\infty} L_2(X^k - X) = 0.$$

The functional $\mathcal{F}$ is closed if for any $C \in \mathbb{R}$ the set $\{X \mid \mathcal{F}(X) \le C\}$ is closed with respect to $\mathcal{L}^2$-convergence. The functional $\mathcal{F}$ is convex if $\mathcal{F}(\lambda X + (1-\lambda)Y) \le \lambda\mathcal{F}(X) + (1-\lambda)\mathcal{F}(Y)$ for all X, Y and $\lambda \in (0,1)$.

Measure of error $\mathcal{E}(X)$ is regular if: 1) $\mathcal{E}(X) \in [0,\infty]$; 2) $\mathcal{E}(X)$ is closed convex; 3) $\mathcal{E}(0) = 0$; 4) $\mathcal{E}(X) > 0$ for any $X \neq 0$; 5) $\mathcal{E}(X) \ge \psi(EX)$ with a convex function ψ on $(-\infty,\infty)$ having $\psi(0) = 0$ but $\psi(t) > 0$ for $t \neq 0$. Definitions for regular measures of risk, deviation and regret are available in Rockafellar and Uryasev (2013, Section 3). The *quadrangle* $(\mathcal{R}, \mathcal{D}, \mathcal{E}, \mathcal{V}, \mathcal{S})$ *is regular* if Eqs. (10)–(12) hold and if also $\mathcal{R}(X)$ is a regular measure of risk, $\mathcal{D}(X)$ is a regular measure of deviation, $\mathcal{V}(X)$ is a regular measure of regret, and $\mathcal{E}(X)$ is a regular measure of error.

(Rockafellar & Uryasev, 2013, Quadrangle Theorem) implies that if Eqs. (10)–(12) hold for functions $\mathcal{R}$, $\mathcal{D}$, $\mathcal{E}$, $\mathcal{V}$, $\mathcal{S}$, and if also $\mathcal{E}(X)$ is a regular measure of error, then $(\mathcal{R}, \mathcal{D}, \mathcal{E}, \mathcal{V}, \mathcal{S})$ is a regular quadrangle. Since $\langle\langle X\rangle\rangle_\alpha^S$ is a regular measure of error, then $\langle\langle X\rangle\rangle_\alpha$ is a regular measure of error, and the quadrangle, related to CVaR norm as a measure of error, is regular. If $\mathcal{E}(X) = \langle\langle X\rangle\rangle_\alpha$ and Eqs. (10)–(12) hold, then the corresponding measure of risk and statistic are calculated from Item 3 of Proposition 2.1, and the whole corresponding quadrangle is presented below.

Proposition 2.2 (CVaR (superquantile) Norm Quadrangle)**.** *For $\alpha \in [0, 1)$ the error measure $\mathcal{E}(X) = \langle\langle X\rangle\rangle_\alpha$ is related to the following regular quadrangle:*

$$\mathcal{S}(X) = \frac{1}{2}\left(q_{(1-\alpha)/2}(X) + q_{(1+\alpha)/2}(X)\right),$$
$$\mathcal{R}(X) = \frac{1-\alpha}{2}\mathrm{CVaR}_{(1+\alpha)/2}(X) + \frac{1+\alpha}{2}\mathrm{CVaR}_{(1-\alpha)/2}(X),$$
$$\mathcal{D}(X) = \frac{1-\alpha}{2}\mathrm{CVaR}_{(1+\alpha)/2}(X-EX) + \frac{1+\alpha}{2}\mathrm{CVaR}_{(1-\alpha)/2}(X-EX),$$
$$\mathcal{V}(X) = \langle\langle X\rangle\rangle_\alpha + EX,$$
$$\mathcal{E}(X) = \langle\langle X\rangle\rangle_\alpha.$$

CVaR norm quadrangle is similar to the Mixed-Quantile-Based quadrangle, see Rockafellar and Uryasev (2013), for

$$\alpha_1 = (1+\alpha)/2, \alpha_2 = (1-\alpha)/2, \lambda_1 = (1-\alpha)/2, \lambda_2 = (1+\alpha)/2. \quad (13)$$

For $k = 1, 2$ define

$$\mathcal{E}_{\alpha_k}(X) = E\left[\frac{\alpha_k}{1-\alpha_k}X^+ + X^-\right], \mathcal{V}_{\alpha_k}(X) = \frac{1}{1-\alpha_k}EX^+.$$

With parameters from (13) we obtain the Mixed-Quantile-Based quadrangle with the same risk and deviation as in CVaR norm quadrangle, but with different statistic, error and regret:

$$\mathcal{S}(X) = \frac{1-\alpha}{2}q_{(1+\alpha)/2}(X) + \frac{1+\alpha}{2}q_{(1-\alpha)/2}(X),$$
$$\mathcal{V}(X) = \min_{B_1,B_2}\left\{\lambda_1\mathcal{V}_{\alpha_1}(X-B_1) + \lambda_2\mathcal{V}_{\alpha_2}(X-B_2) \mid \lambda_1B_1 + \lambda_2B_2 = 0\right\},$$
$$\mathcal{E}(X) = \min_{B_1,B_2}\left\{\lambda_1\mathcal{E}_{\alpha_1}(X-B_1) + \lambda_2\mathcal{E}_{\alpha_2}(X-B_2) \mid \lambda_1B_1 + \lambda_2B_2 = 0\right\}.$$

Suppose one is optimizing measure of error over some parametric family $X(\theta)$:

$$\min_\theta \quad \mathcal{E}_i(X(\theta)), \quad (14)$$

where $i = 1$ for error from CVaR norm quadrangle, and $i = 2$ for error from Mixed-Quantile-Based quadrangle. Assume that $X(\theta) =$

$\theta_0 + Y(\tilde{\theta})$, where $\theta = (\theta_0, \tilde{\theta})$, and θ_0 is a free parameter. Define $\theta_i^* = \arg\min_\theta \quad \mathcal{E}_i(X(\theta))$. Then $\tilde{\theta}_1^* = \tilde{\theta}_2^* = \arg\min_{\tilde{\theta}} \mathcal{D}(Y(\tilde{\theta})) = \tilde{\theta}^*$. Therefore, $Y(\tilde{\theta}_1^*) = Y(\tilde{\theta}_2^*)$ and two optimal points $X(\theta_1^*)$ and $X(\theta_2^*)$ for problems (14) can be obtained from each other by adding a constant shift

$$X(\theta_1^*) = (\theta_0^*)_1 + Y(\tilde{\theta}^*), X(\theta_2^*) = (\theta_0^*)_2 + Y(\tilde{\theta}^*), X(\theta_1^*) - X(\theta_2^*) = (\theta_0^*)_1 - (\theta_0^*)_2.$$

3. Trimmed L1-norm

Paper (Ge et al., 2011) considers a class of functions defined similar to L_p norms, but for $p \in [0, 1)$. These functions are not norms and they are concave for some regions of the space they are defined[6]. Such functions are used in optimization problems to achieve a sparse solution vector. We will define a similar functions in terms of CVaR concept.

Further we define *trimmed L1-norm*. Contrary to CVaR norm, this function takes average over *smallest* α-fraction of absolute values $|X|$ of a random variable X. The word "norm" here is potentially deceptive, since it corresponds to L1 and not the resulting function itself: trimmed L1-norm is not actually a norm. The term "trimmed" is widely used in robust regression, when the average of a few smallest regression residuals is minimized. Before averaging, residuals are usually transformed with some function $\phi : \mathbb{R} \to [0, \infty)$. The case of $\phi(x) = x^2$ is the most commonly used and corresponds to the least trimmed squares regression (LTS), see Rousseeuw (1984); Rousseeuw and Van Driessen (1999); Zabarankin and Uryasev (2014b). The case of $\phi(x) = |x|$ corresponds to the least trimmed sum of absolute deviations (LTA) regression, see Bassett Jr (1991); Hawkins and Olive (1999); Hössjer (1994), it also corresponds to the trimmed L1-norm function here, see below. The general case for arbitrary function ϕ was described in Neykov, Čížek, Filzmoser, and Neytchev (2012, formula (3)) and formalized for random variables via average quantile function in Zabarankin and Uryasev (2014b, problem (6.5.9)).

One possible way to define trimmed L1-norm, or T_α, is as follows.

Definition 3. Let X be a random variable with $E|X| < \infty$. Trimmed L1-norm $T_\alpha(X)$ for $\alpha \in [0, 1]$ is defined by

$$T_\alpha(X) = -\text{CVaR}_{1-\alpha}(-|X|).$$

Below we provide some mathematical properties for trimmed L1-norm. These properties mostly follow from the ones presented in Section 2.

Properties and representations of trimmed L1-norm

1. For $\alpha = 0$, trimmed L1-norm $T_\alpha(X) = \inf|X|$. For $\alpha \in (0, 1]$ trimmed L1-norm can be calculated using one of the formulas below:

$$T_\alpha(X) = \frac{1}{\alpha}(E|X| - (1-\alpha)\text{CVaR}_\alpha(|X|)), \tag{15}$$

$$T_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha q_p(|X|)dp, \tag{16}$$

$$T_\alpha(X) = \max_c \left\{c - \frac{1}{\alpha}E[|X| - c]^-\right\}. \tag{17}$$

2. $0 \le T_\alpha(X) \le L_1(X) = E|X|$,
3. $T_\alpha(\lambda X) = |\lambda| T_\alpha(X)$,
4. if $XY \ge 0$, then $T_\alpha(\lambda X + (1-\lambda)Y) \ge \lambda T_\alpha(X) + (1-\lambda)T_\alpha(Y)$,
5. $T_\alpha(0) = 0$, however, there exists $X \ne 0$ such that $T_\alpha(X) = 0$,
6. $T_\alpha(X) \le 0 \Leftrightarrow T_\alpha(X) = 0$ implies that $P(X = 0) \ge \alpha$.
7. $T_\alpha(X)$ is a continuous non-decreasing function w.r.t. α.
8. $\alpha T_\alpha(X)$ is a convex non-decreasing function w.r.t. α.

Formula (15) follows from Pflug (2000, Proposition 2 (iii)), when $Y = |X|$ is taken. $T_\alpha(X)$ can be interpreted as an expectation of $|X|$ in left α-tail. Note that $T_\alpha(X)$ is then the average quantile of the random variable $|X|$, hence, formula (16) holds, see Zabarankin and Uryasev (2014a, formula (1.4.1)). Formula (17) follows from Rockafellar et al. (2006, formula (5)).

For $p \in (0, 1)$ the following inequality holds $L_p(X) \le L_1(X)$, where $L_p(X) = (E|X|^p)^{1/p}$. Since x^p is a concave function for $0 < p < 1$, using Jensen's inequality we have $E|X|^p \le (E|X|)^p$, therefore, $(E|X|^p)^{1/p} \le E|X|$, and Item 2 shows that for trimmed L1-norm. Items 2–4 follow from the fact that $-\frac{1}{\alpha}\int_0^\alpha q_p(X)dp$ is a coherent, expectation bounded risk measure, see Rockafellar et al. (2006, Example 3). Item 5 follows from example of $X = 0$ with probability 0.5 and $X = 1$ with probability 0.5. Then, for $\alpha \in [0, 0.5]$ function $T_\alpha(X) = 0$. Item 6 follows from formula (16). Note that Item 6 can be used in optimization problem settings for a constraint $T_\alpha(X) \le 0$ to achieve a given sparsity of a solution variable vector, or as $\max \alpha$ subject to $T_\alpha(X) \le 0$ to maximize the sparsity of a solution variable vector.

Item 4 implies that $T_\alpha(X)$ is a concave function for $X \ge 0$. Notice that this property cannot be strengthened to concavity in the whole space of random variables. Consider a function $g(X)$ such that $g(X) \ge 0$, $g(0) = 0$ and $g(X) \not\equiv 0$. Assume that $g(X)$ is concave in the space of random variables. Since $g(X) \not\equiv 0$, then there exists X such that $g(X) > 0$. Then

$$g(X) + g(-X) > 0 = g(0) = g(X - X),$$

which implies that $g(X)$ is not a concave function.

Consider $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. Let $X(\mathbf{x})$ be a discretely distributed random variable taking values $x_1, \dots, x_n$ with equal probabilities $\frac{1}{n}$. Then the trimmed L1-norm on $\mathbb{R}^n$ is defined as $t_\alpha(\mathbf{X}) = T_\alpha(X)$. Properties of $t_\alpha(\mathbf{x})$ are similar to properties of the $T_\alpha(X)$ and follow directly from Definition 3 and enlisted properties. For $\mathbf{x} \in \mathbb{R}^n$, trimmed L1-norm is calculated as follows:

$$t_\alpha(\mathbf{x}) = \frac{1}{\alpha}(l_1(\mathbf{x}) - (1-\alpha)\langle\langle \mathbf{x}\rangle\rangle_\alpha^S),$$

$$t_{\alpha_j}(\mathbf{x}) = (|x|_{(1)} + \dots + |x|_{(j)})/j, \qquad \text{for } \alpha = \alpha_j = j/n,$$

$$t_0(\mathbf{x}) = \min_i |x_i|, \qquad \text{for } \alpha = 0,$$

$$t_1(\mathbf{x}) = l_1(\mathbf{x}), \qquad \text{for } \alpha = 1.$$

Fig. 1 shows level-sets of $\langle\langle \mathbf{x}\rangle\rangle_\alpha^S$ and $t_\alpha(\mathbf{x})$ in $\mathbb{R}^2$ for different values of α. The function $t_\alpha(\cdot)$ is a natural extension of $\langle\langle\cdot\rangle\rangle_\alpha^S$. When α varies from 0 to 1, the function $t_\alpha(\mathbf{x})$ changes from $\min_i |x_i|$ to $l_1(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n |x_i|$, and the function $\langle\langle \mathbf{x}\rangle\rangle_\alpha^S$ changes from $l_1(\mathbf{x})$ to $\max_i |x_i|$.

4. Case study

4.1. Linear regression: financial optimization dataset

We illustrate CVaR norm quadrangle, see Proposition 2.2, with the following case study. The case study results, data, and codes are posted at this link[7].

We considered a linear regression problem with CVaR norm error. Let $\mathbf{X}$ be a $n \times d$ design matrix, where n is a number of observations, d is a number of explanatory variables. Let $\mathbf{y} \in \mathbb{R}^n$ be a vector of observations on the dependent variable. Let $\mathbf{e} \in \mathbb{R}^n$ be a vector of ones. We denote by $\tilde{\mathbf{X}} = [\mathbf{e}, \mathbf{X}]$ an extended design matrix including additional unit column. We considered a linear regression: $\hat{\mathbf{y}} = \tilde{\mathbf{X}}\mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^{d+1}$ is a vector of parameters. To solve this regression problem we minimized CVaR norm of vector of residuals $\mathbf{y} - \hat{\mathbf{y}}$:

$$\min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad \langle\langle \mathbf{y} - \tilde{\mathbf{X}}\mathbf{a}\rangle\rangle_\alpha. \tag{18}$$

[6] For $p \in [0, 1)$ there is $l_p(\mathbf{x})$ in $\mathbb{R}^n$ and $L_p(X)$ in the space of random variables. Concavity holds, for example, for region $\mathbf{x} \ge 0$ in $\mathbb{R}^n$ and for region $X \ge 0$ in the space of random variables.

[7] http://www.ise.ufl.edu/uryasev/research/testproblems/advanced-statistics/cvar-norm-regression/

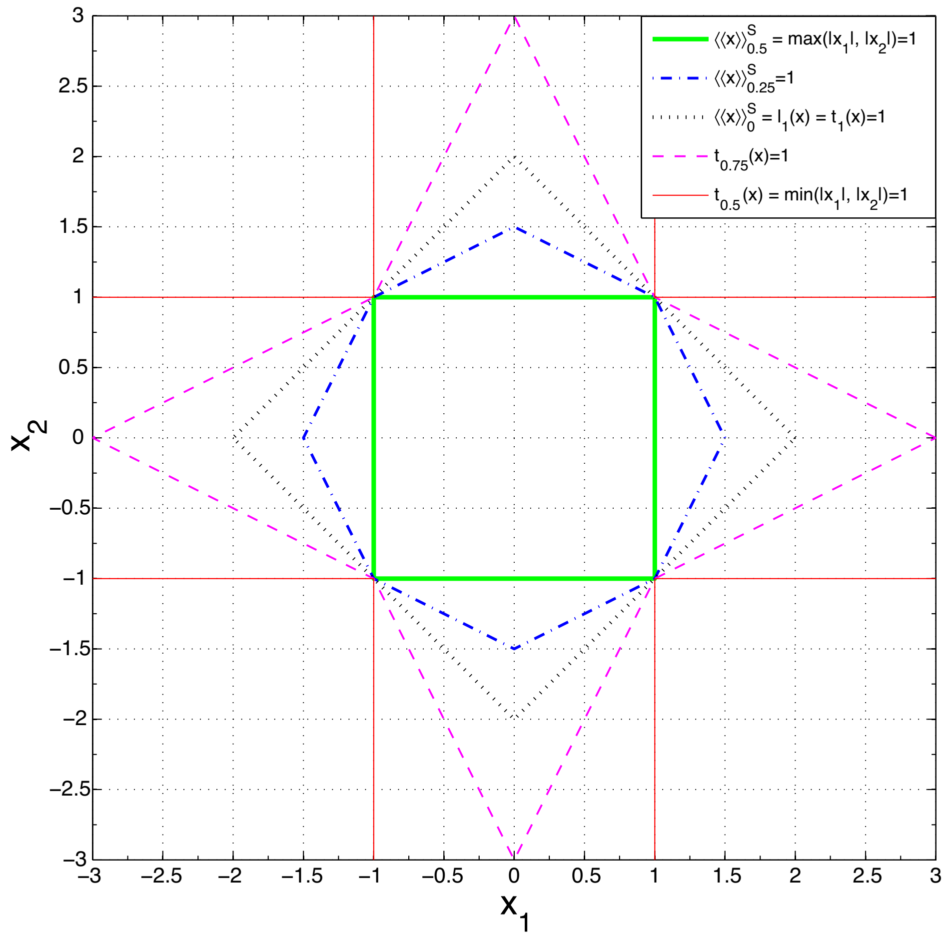

Fig. 1. Level-sets of CVaR norm $\langle\langle \mathbf{x} \rangle\rangle^S_\alpha$ for $\alpha = 0, 0.25, 0.5$ and level-sets of trimmed L1-norm $t_\alpha(\mathbf{x})$ for $\alpha = 0.5, 0.75, 1$ in $\mathbb{R}^2$ space. For $\alpha \in [0.5, 1]$ norm $\langle\langle \mathbf{x} \rangle\rangle^S_\alpha = \max_i |x_i|$. For $\alpha \in [0, 0.5]$ function $t_\alpha(\mathbf{x}) = \min_i |x_i|$. Equality $\langle\langle \mathbf{x} \rangle\rangle^S_0 = l_1(\mathbf{x}) = t_1(\mathbf{x})$ holds.

It is desirable to use CVaR norm in regression when we want to control directly large absolute values of residuals. We are indifferent to the sign of the residual. We just do not want to have large absolute values, but are tolerant to small absolute values. Similar purpose can be achieved by minimizing L_p norm, but in this case we do not control directly some specific percentage of largest outcomes. We can directly specify the percentage of largest absolute residuals with CVaR norm, e.g., 10% of largest outcomes. We also want to mention that the percentile regression (Koenker & Bassett Jr, 1978) with Koenker and Bassett function is quite close to CVaR norm regression. However, percentile regression is concentrated on large outcomes in one tail, while CVaR norm regression pays attention to large outcomes without identifying the sign of the residual.

Similar type of error has been considered earlier in OR literature. For instance, Zabarankin and Uryasev (2014b, formula (6.5.9)) considered so called "average alfa-quantile" error minimization applied to the transformed residual. In this problem, the average is taken over the left tail of the distribution. Such approach corresponds to trimmed error measures and to robust regression, it produces regression which is stable to outliers. By selecting the absolute value as a transformation function in "average alfa-quantile" error we are coming to trimmed L1-norm minimization, or to LTA regression. Here we consider averaging over the right tail of distribution, which leads to CVaR, or superquantile, functions. Similarly, by selecting the absolute value as a transformation function in CVaR we are coming to CVaR norm minimization. CVaR norm regression is not stable to outliers, moreover, it is a "pessimistic" estimator focused is on a fraction of the most "problematic" observations. Aside from potential benefits of the pessimistic approach, problem (18) is a convex problem and can be solved precisely and efficiently.

We consider the dataset from the case study "Estimation of CVaR through Explanatory Factors with Mixed Quantile Regression"[8]. The data contains returns of the Fidelity Magellan Fund as a dependent variable. Russell Value Index (RUJ), Russell 1000 Value Index (RLV), Russell 2000 Growth Index (RUO) and Russell 1000 Growth Index (RLG) are taken as independent variables. Data include 1264 observations. Solving Time on a PC with 2.83GHz is 0.01 s.

The CVaR norm is minimized with Portfolio Safeguard (AORDA, 2009) software package. Confidence level α in CVaR norm equals $\alpha = 0.9$. We minimized CVaR instead of CVaR norm, according to calculation formula (8). Denote $\bar{\mathbf{y}} = [\mathbf{y}; -\mathbf{y}] \in \mathbb{R}^{2n}$ and $\bar{\mathbf{X}} = [\tilde{\mathbf{X}}; -\tilde{\mathbf{X}}] \in \mathbb{R}^{2n\times d}$. Formula (8) implies

$$\langle\langle \mathbf{y} - \tilde{\mathbf{X}}\mathbf{a} \rangle\rangle^S_\alpha = \text{CVaR}_{(1+\alpha)/2}(\bar{\mathbf{y}} - \bar{\mathbf{X}}\mathbf{a}).$$

Then, problem (18) is equivalently stated as follows

$$\min_{\mathbf{a}\in\mathbb{R}^{d+1}} \quad \text{CVaR}_{(1+\alpha)/2}(\bar{\mathbf{y}} - \bar{\mathbf{X}}\mathbf{a}).$$

Optimization results for this problem are in Table 1.

CVaR norm quadrangle is a regular quadrangle, see Proposition 2.2. According to Rockafellar and Uryasev (2013, Regression Theorem), first stated as the Error-Shaping Decomposition

[8] http://www.ise.ufl.edu/uryasev/research/testproblems/financial_engineering/estimation-of-cvar-through-explanatory-factors-with-mixed-quantile-regression/

Table 1
Optimal vector of parameters and objective for linear regression with CVaR norm.

rlv	rlg	ruj	ruo	Intercept	Objective
0.578	0.484	−0.07	−0.008	−0.002	0.015

of Regression Theorem (Rockafellar et al., 2008, Theorem 3.2, page 722), the intercept, obtained in regression, equals to the Statistic of a modified residuals. In CVaR norm quadrangle, statistic equals $\mathcal{S}(X) = (q_{(1+\alpha)/2}(X) + q_{(1-\alpha)/2}(X))/2$. Denote the optimal vector of parameters obtained in regression by $\mathbf{a}^* = [c^*, \mathbf{b}^*]$, where $c^* \in \mathbb{R}$ is an optimal intercept. According to the Regression Theorem, $c^* \in \mathcal{S}(\mathbf{y} - \mathbf{X}\mathbf{b}^*)$ (we write $\in$ because, in general, quantile $q_p(X)$ is an interval, see (1), therefore $\mathcal{S}(X)$ is also an interval). At the optimal point, $c^* = -0.002$, $q^-_{0.05}(\mathbf{y} - \mathbf{X}\mathbf{b}^*) = -0.013$, $q^-_{0.95}(\mathbf{y} - \mathbf{X}\mathbf{b}^*) = 0.009$. Therefore,

$$\mathcal{S}(\mathbf{y} - \mathbf{X}\mathbf{b}^*) \approx (q^-_{0.05}(\mathbf{y} - \mathbf{X}\mathbf{b}^*) + q^-_{0.95}(\mathbf{y} - \mathbf{X}\mathbf{b}^*))/2 = (-0.013 + 0.009)/2 = -0.002 = c^*.$$

That is, numerical experiment confirms theoretical results for CVaR norm quadrangle, namely statistics $\mathcal{S}(\mathbf{y} - \mathbf{X}\mathbf{b}^*) = \frac{1}{2}(q_{(1-\alpha)/2}(\mathbf{y} - \mathbf{X}\mathbf{b}^*) + q_{(1+\alpha)/2}(\mathbf{y} - \mathbf{X}\mathbf{b}^*))$ including the optimal intercept, c^*.

4.2. *Linear regression on simulated data*

In the following case study we consider L1-norm as an out-of-sample criterion. For the in-sample criteria we consider elements of the CVaR norm parametric family. L1-norm is an element of this family with the corresponding parameter value 0. We vary value of the parameter between 0 and 1 for in-sample learning to optimize L1-norm of residuals in out-of-sample. We show that using CVaR norms with parameter value bigger than 0 can lead, for small training samples, to better out-of-sample performance than direct in-sample minimization of L1-norm.

We illustrate CVaR norm regression with a controlled numerical experiment. In this case study we set true law as $y(x) = x$ for $x \in [0, 1]$. We pick sample points $(x_1, \ldots, x_{11}) = (0, 0.1, \ldots, 0.9, 1)$. Then we simulate dependent variable as $y_i = x_i + \varepsilon_i$, where error terms $\varepsilon_i \propto \text{Laplace}(0, 0.5)$ are distributed according to the Laplace distribution (double exponential distribution, probability density function $f(x; \mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$), see Fig. 2.

L1-norm of regression residuals is chosen as the out-of-sample error criterion. Similar to exponential distribution, Laplace distribution has heavy tails, which makes L1-based regression a reasonable choice. Moreover, minimization of L1-norm of residuals is equivalent to likelihood maximization for the case of Laplace-distributed error. Since the true distribution is known, there is no need to divide sample into training and testing subsamples to measure error of the estimated model. Let $\hat{y}_i$ denote model estimation for the data point x_i. First, "expected error" is calculated as

$$\text{EE} = \frac{1}{n}\sum_{i=1}^{n} E_\varepsilon |y(x_i) + \varepsilon - \hat{y}_i|,$$

where $n = 11$ and each expectation is taken for $\varepsilon \propto \text{Laplace}(0, 0.5)$. Also, since the true law is know, then "true error" is calculated as

$$\text{TE} = \frac{1}{n}\sum_{i=1}^{n} |y(x_i) - \hat{y}_i|.$$

As in Section 4.1, the problem (18) is solved. L1-norm minimization is a special case of (18) when $\alpha = 0$. The goal of the case study is to test whether "pessimistic" estimation provided by CVaR norm minimization with $\alpha > 0$ can achieve better out-of-sample quality of estimators, measured by EE and TE, than direct L1-norm minimization with $\alpha = 0$. To smooth the results obtained with the randomly generated sample, number of different samples generated $N = 1000$,

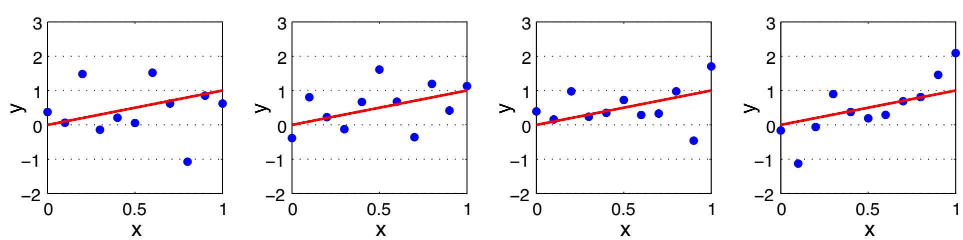

Fig. 2. Samples generated for regression. A true law is the line. Observations are the dots.

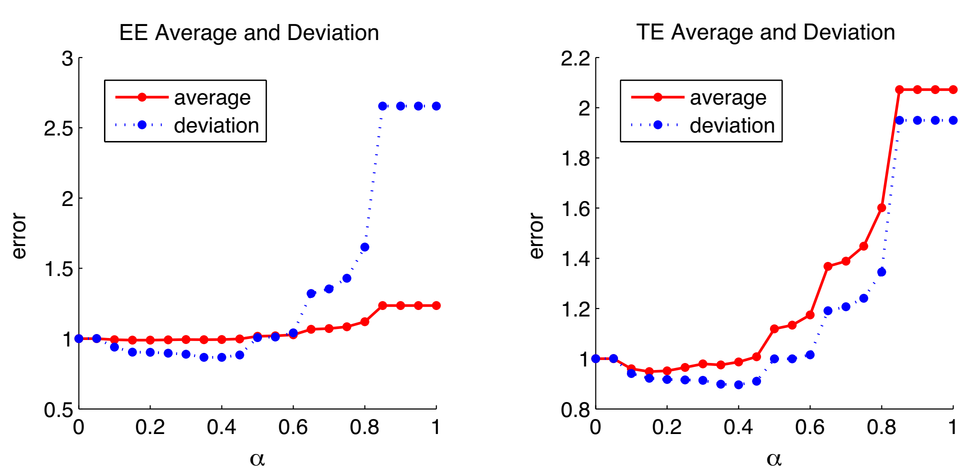

Fig. 3. Average error and standard deviation of error, scaled to the average error for $\alpha = 0$ and the standard deviation of error for $\alpha = 0$ correspondingly, as functions of α. On the left: "Expected Error". On the right: "True Error".

each sample size $n = 11$. For each sample, the problem (18) is solved for values $\alpha_k = (k-1)/(K-1)$ for $k = 1, \ldots, K$, $K = 21$ (that is, values of α are $(0, 0.05, 0.01, \ldots, 1)$). For sample j, where $1 \le j \le N$, and parameter α_k, regression estimator is found, and errors EE_j^k and TE_j^k are calculated. Fig. 3 shows dependence on α for average error and standard deviation of error, scaled to the average error for $\alpha = 0$ and the standard deviation of error for $\alpha = 0$ correspondingly. These values are calculated for the two types of error we consider: "expected error" EE, and "true error" TE. Minimal average EE is achieved for $\alpha = 0.15$ and is 1% lower than average EE for $\alpha = 0$. Despite the modest drop in average error, CVaR norm model is more stable than L1-norm model: standard deviation of EE is 10% lower for $\alpha = 0.15$ than for $\alpha = 0$. Average TE is minimized by the same value $\alpha = 0.15$ and is 5% lower than for $\alpha = 0$; standard deviation of TE is 8% lower for $\alpha = 0.15$ than for $\alpha = 0$.

This case study showed that even if CVaR norm is not an out-of-sample criterion itself, as it was assumed in Section 4.1, the minimization of CVaR norm can be more advantageous than direct optimization of the out-of-sample criterion.

5. Conclusion

This paper extended the definition of CVaR norm from Eucledian space to a $\mathcal{L}^1(\Omega)$ space of random variables and proved that CVaR norm is indeed a norm. CVaR norm is defined in two variations: scaled and non-scaled. Several properties for scaled and non-scaled CVaR norm, as a function of confidence level, were proved. Dual norm for CVaR norm was proved to be the maximum of L_1 and scaled L_∞ norms. CVaR norm was proved to be a regular Measure of Error and components of the corresponding CVaR (superquantile) Norm Risk Quadrangle were found. Trimmed L1-norm, which is a non-convex extension for CVaR norm, was introduced analogously to function L_p for $p < 1$. Linear regression problems were solved by minimizing CVaR norm of regression residuals. CVaR norm has an intuitive interpretation to be chosen as an out-of-sample criterion, and also minimization of CVaR norm can be more advantageous than direct optimization of the out-of-sample criterion.

Acknowledgments

Authors would like to thank Prof. R. Tyrrell Rockafellar, Prof. Michael Zabarankin, Prof. Jun-Ya Gotoh and Dr. Konstantin Pavlikov for their helpful comments and suggestions. Authors would also like to thank Prof. Georg Pflug and two anonymous reviewers for their tremendous help shaping the material of the paper in a clearer way and for the help connecting this study with existing work in related fields.

This work was partially supported by the USA AFOSR grants: "Design and Redesign of Engineering Systems", FA9550-12-1-0427, and "New Developments in Uncertainty: Linking Risk Management, Reliability, Statistics and Stochastic Optimization", FA9550-11-1-0258.

Appendix A. Proofs for properties of CVaR (superquantile) norm

Proof for Item 3 from Section 2

Proof. Formulas (5)–(7) follow directly from CVaR calculation formulas, see (e.g. Rockafellar & Uryasev, 2000, formulas (3),(5)), (e.g. Rockafellar & Uryasev, 2002, Definitions 3,4, Theorem 10). Let us prove formula (8). For any $x \ge 0$,

$$\begin{aligned} F_Y(x) &= P(Y \le x) = \frac{P(X \le x) + P(-X \le x)}{2} \\ &= \frac{P(X \le x) + 1 - P(X < -x)}{2} = \frac{1 + P(-x \le X \le x)}{2} \\ &= \frac{1 + F_{|X|}(x)}{2}, \end{aligned}$$

hence,

$$\begin{aligned} q^+_{(1+\alpha)/2}(Y) &= \inf\left\{x \,\middle|\, \frac{1 + F_{|X|}(x)}{2} > \frac{1+\alpha}{2}\right\} = \inf\left\{x \,\middle|\, F_{|X|}(x) > \alpha\right\} \\ &= q^+_\alpha(|X|), \end{aligned}$$

and finally,

$$\begin{aligned} \langle\langle X\rangle\rangle^S_\alpha &= \frac{1}{1-\alpha}\int_\alpha^1 q^+_p(|X|)dp = \frac{1}{1-\alpha}\int_\alpha^1 q^+_{(1+p)/2}(Y)dp \\ &= \frac{2}{1-\alpha}\int_{(1+\alpha)/2}^1 q^+_{(1+p)/2}(Y)d\left(\frac{1+p}{2}\right) = \mathrm{CVaR}_{(1+\alpha)/2}(Y). \end{aligned}$$

□

Proof for Item 4 from Section 2

Proof. Let us proceed with the proof item by item.

1. The proof follows directly from $\mathrm{CVaR}_1(|X|) = \sup\{|x_i|\}_{i=1}^N$.
2. If X is a discrete random variable, then $q_p(|X|)$ is a step function of p. Step number i is having length $\alpha_i - \alpha_{i-1} = |p|_{(i)}$ and height $q_{\alpha_i}(|X|) = |x|_{(i)}$. Then, using additivity property of the integral (6),
$$\begin{aligned} (1-\alpha_j)\langle\langle X\rangle\rangle^S_{\alpha_j} &= \int_{\alpha_j}^1 q_p(|X|)dp = \sum_{i=j+1}^N \int_{\alpha_{i-1}}^{\alpha_i} q_p(|X|)dp \\ &= \sum_{i=j+1}^N |p|_{(i)}|x|_{(i)}. \end{aligned}$$
3. Since $q_p(|X|)$ is a step function of p and $\alpha_j < \alpha < \alpha_{j+1}$, then, using additivity property of the integral (6),
$$\begin{aligned} \int_\alpha^1 q_p(|X|)dp &= (1-\lambda)\int_{\alpha_j}^1 q_p(|X|)dp + \lambda\int_{\alpha_{j+1}}^1 q_p(|X|)dp, \\ \lambda &= \frac{\alpha - \alpha_j}{\alpha_{j+1} - \alpha_j}, \\ \langle\langle X\rangle\rangle^S_\alpha &= (1-\lambda)\frac{1-\alpha_j}{1-\alpha}\langle\langle X\rangle\rangle^S_{\alpha_j} + \lambda\frac{1-\alpha_{j+1}}{1-\alpha}\langle\langle X\rangle\rangle^S_{\alpha_{j+1}}. \end{aligned}$$

□

Proof for Item 9 from Section 2

Proof. We further prove that axioms of the regular measure of error hold for $\langle\langle X\rangle\rangle^S_\alpha$.

- $\mathcal{E}(X) \in [0, \infty]$, $\mathcal{E}(0) = 0$ and $\mathcal{E}(X) > 0$ for any $X \ne 0$ follows from the fact that $\langle\langle X\rangle\rangle^S_\alpha$ is a norm.
- $\mathcal{E}(X)$ is closed and convex, which follows from Item 1 from Proposition 2.1.
- $\mathcal{E}(X) = \langle\langle X\rangle\rangle^S_\alpha \ge E|X| \ge |EX| = \psi(EX)$ for $\psi(x) = |x|$ on $(-\infty, \infty)$ having $\psi(0) = 0$ and $\psi(t) > 0$ for $t \ne 0$. □

References

AORDA (2009). Portfolio safeguard (PSG) version 2.1. American Optimal Decisions, Inc.Gainesville, FL. http://www.aorda.com/aod/welcome.action.

Bassett Jr, G. W. (1991). Equivariant, monotonic, 50% breakdown estimators. *The American Statistician, 45*(2), 135–137.

Bertsimas, D., Pachamanova, D., & Sim, M. (2004). Robust linear optimization under general norms. *Operations Research Letters, 32*(6), 510–516.

Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media.

Dentcheva, D., & Ruszczynski, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization, 14*(2), 548–566.

Ge, D., Jiang, X., & Ye, Y. (2011). A note on the complexity of lp minimization. *Mathematical Programming, 129*(2), 285–299.

Gotoh, J.-y., & Uryasev, S. (2013). Two pairs of families of polyhedral norms versus lp-norms: proximity and applications in optimization. *University of Florida, Research Report 2013-3*. http://www.ise.ufl.edu/uryasev/files/2013/10/TwoPairsOfPolyhedralNormsVersusLpNorms_20131025_UFISEPD.pdf

Hall, R. W., & Tymoczko, D. (2012). Submajorization and the geometry of unordered collections. *The American Mathematical Monthly, 119*(4), 263–283.

Hawkins, D. M., & Olive, D. (1999). Applications and algorithms for least trimmed sum of absolute deviations regression. *Computational Statistics & Data Analysis, 32*(2), 119–134. http://dx.doi.org/10.1016/S0167-9473(99)00029-8.
Hössjer, O. (1994). Rank-based estimates in the linear model with high breakdown point. *Journal of the American Statistical Association, 89*(425), 149–158.
Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society , 46*(1), 33–50.
Krzemienowski, A. (2009). Risk preference modeling with conditional average: an application to portfolio optimization. *Annals of Operations Research, 165*(1), 67–95.
Marshall, A. W., Olkin, I., & Arnold, B. C. (2011). *Inequalities: theory of majorization and its applications*. New York: Springer.
Merigó, J. M., & Yager, R. R. (2013). Norm aggregations and OWA operators. In *Aggregation functions in theory and in practise* (pp. 141–151). Springer.
Neykov, N., Čížek, P., Filzmoser, P., & Neytchev, P. (2012). The least trimmed quantile regression. *Computational Statistics & Data Analysis, 56*(6), 1757–1770.
Ogryczak, W., & Ruszczynski, A. (2002). Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization, 13*(1), 60–78.
Ogryczak, W., & Zawadzki, M. (2002). Conditional median: a parametric solution concept for location problems. *Annals of Operations Research, 110*(1-4), 167–181.
Pavlikov, K., & Uryasev, S. (2014). CVaR norm and applications in optimization. *Optimization Letters, 8*(7), 1999–2020.
Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization* (pp. 272–281). Springer.
Rockafellar, R. T., & Royset, J. O. (2010). On buffered failure probability in design and optimization of structures. *Reliability Engineering and System Safety, 95*(5), 499–510. http://dx.doi.org/10.1016/j.ress.2010.01.001.
Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk, 2*, 21–41.
Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance, 26*(7), 1443–1471.
Rockafellar, R. T., & Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science, 18*(1), 33–53.
Rockafellar, R. T., Uryasev, S., & Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics, 10*(1), 51–74.
Rockafellar, R. T., Uryasev, S., & Zabarankin, M. (2008). Risk tuning with generalized linear regression. *Mathematics of Operations Research, 33*(3), 712–729.
Romeijn, H. E., Ahuja, R. K., Dempsey, J. F., & Kumar, A. (2005). A column generation approach to radiation therapy treatment planning using aperture modulation. *SIAM Journal on Optimization, 15*(3), 838–862.
Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association, 79*(388), 871–880.
Rousseeuw, P. J., & Van Driessen, K. (1999). Computing LTS regression for large data sets. In *Institute of mathematical statistics bulletin*. Citeseer.
Torra, V., & Narukawa, Y. (2007). *Modeling decisions: information fusion and aggregation operators*. Berlin: Springer.
Yager, R. R. (2010). Norms induced from OWA operators. *IEEE Transactions on Fuzzy Systems, 18*(1), 57–66.
Zabarankin, M., & Uryasev, S. (2014a). Random variables. In *Statistical decision problems* (pp. 3–17). Springer.
Zabarankin, M., & Uryasev, S. (2014b). Regression models. In *Statistical decision problems* (pp. 71–87). Springer.