Evaluations of Risk Measures for Different Probability Measures

Alois Pichler [a,*]

[a] *University of Vienna, Austria. Department of Statistics and Operations Research.*

Abstract

Stochastic optimization problems, which arise in different areas – for example simple asset allocation problems or problems in insurance – often involve coherent risk measures. In these real-world problems the considered risk measure is frequently built on empirical distributions. It is therefore of interest to understand the potential deviation, which occurs when evaluating a risk measure for a perturbed, or slightly changed probability measure.

This paper addresses the potential deviation. It turns out that the Wasserstein distance, a well-known distance for probability measures, provides a valuable notion of distance in the present context: many risk measures allow a precise quantification in terms of the Wasserstein distance, and important risk measures are continuous with respect to this distance.

For specific random variables, which often occur in concrete, real-world problems, it is moreover demonstrated that the derived constants, describing the continuity properties, cannot be improved. The associated worst probability measures can be given explicitly in important situations.

Keywords: Risk Measures, Continuity Properties, Wasserstein and Kantorovich Distance, Optimal Transport
2010 MSC: 90C15, 60B05, 62P05

1. Introduction

In this paper we discuss evaluations of risk measures for different probability measures. A major motivation to study this problem is given by the fact that applications and optimization problems, which arise in finance or energy, use risk measures in their objective. Moreover decisions on optimal investment, which are often driven by (stochastic) optimization and which involve risk measures, frequently build on empirical distributions. But similar observations are possibly as likely as the observations at hand, so it is desirable to ensure that the risk measure employed will result in some similar values as well for similar observations.

The impact of different observations is reflected by different probability measures. In view of *sensitivity analysis* the results in this paper allow quantifying the deviation of considered optimization problems, and various kinds of considerations on *ambiguity* are naturally included as well.

It turns out that the Wasserstein distance (in Russian literature the term Kantorovich distance is preferred, cf. [31]) is a useful notion of distance which ensures the desired properties: typical risk measures allow an estimation in terms of the Wasserstein distance, and for important ones these estimates can be given in a sufficiently precise form.

Above that we describe – for some selected risk measures – probability measures which are worst in the sense that they modify the resulting risk measure utmost.

An important part of the analysis in the present paper is based on the Legendre-Fenchel transform and on Kusuoka's representation.

The paper is organized as follows: a preliminary discussion introduces the concepts and tools required. Section 3 contains a main result, an upper bound for the change in the risk measure,

*Corresponding Author
Email address: `alois.pichler@univie.ac.at` (Alois Pichler)

whenever the probability measure is being perturbed. Section 4 characterizes measures which change the risk functional to the largest possible extent, and particular evaluations are collected in Section 5. Some illustrations in Section 6 complete the paper.

2. Preliminary Discussion, And Continuity With Respect To The Random Variable

The concept of risk measures has been initiated and discussed in the pioneering paper [2], and it was developed further in a sequence of important contributions, among them [5], [7] and [25]. In the following definition introducing the *risk measure* we adopt the general concept, we follow the convex setting elaborated in [28].

Definition 1. A function $\rho\colon \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$ defined on a linear space of $\mathbb{R}-$valued random variables $\mathcal{Y}$ is said to be a *risk measure* (or *risk functional*) if the following axioms are satisfied:

(i) Monotonicity: if $Y_1 \le Y_2$, then $\rho(Y_1) \le \rho(Y_2)$ [1] $(Y_1, Y_2 \in \mathcal{Y})$;
(ii) Convexity: $\rho(tY_1 + (1-t)Y_2) \le t\rho(Y_1) + (1-t)\rho(Y_2)$ whenever $0 \le t \le 1$ $(Y_1, Y_2 \in \mathcal{Y})$;
(iii) Translation Equivariance: if $c \in \mathbb{R}$, then $\rho(Y + c) = \rho(Y) + c$; [2]
(iv) Positive Homogeneity: for $t > 0$, $\rho(tY) = t\,\rho(Y)$ $(Y \in \mathcal{Y})$.

In this definition we allow ρ to take the value ∞. However, throughout this paper we shall assume that ρ is *proper*, that is to say there is at least one Y with $\rho(Y) < \infty$.

Remark 2. It is important to note that similar and equivalent definitions are often used when investigating risk measures. The notion, however, is not used consistently in the literature, the preferred setting is often driven by the application or interpretation in mind. To give an example, $\tilde{\rho}(Y) := \rho(-Y)$ is a coherent measure of risk in [7], whereas [19] use the term *acceptability functional* for the concave functional $\mathcal{A}(Y) := -\rho(-Y)$. The term *acceptability functional* is motivated by the objective of an investor to accept only those asset allocation strategies Y, for which $\mathcal{A}(Y)$ exceeds a certain threshold. The threshold represents the risk the investor is willing to accept, his/ her risk appetite.

In an actuarial context the random variable Y is naturally associated with loss, and $\rho(Y)$ represents a premium principle, that is an insurance pricing strategy.

In another financial context the random variable Y is often associated with the random net worth of an asset allocation strategy. It should be emphasized that Y is even *linear* in this situation: for example suppose that 30 % of the available cash amount is invested in stock i_1, 20 % in bond i_2 etc., such that $Y(x) = 30\,\% \cdot x_1 + 20\,\% \cdot x_2 + \dots$ is the random net worth of the concrete allocation strategy (x_1 is the random stock price of i_1, etc.): then the asset allocation $Y = (30\,\%, 20\,\%, \dots)$ represents a linear functional on $\mathbb{R}^d$.

Examples Of Risk Measures

Average Value-at-Risk

The most prominent risk measure probably is the *Average Value-at-Risk at level* α,[3] which is defined as

$$\mathrm{AV@R}_\alpha(Y) := \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{V@R}_p(Y)\,\mathrm{d}p, \tag{1}$$

where $\mathrm{V@R}_p(Y) = \inf\{q\colon P(Y \le q) \ge p\}$ is the left side quantile. The equivalent expression

$$\mathrm{AV@R}_\alpha(Y) = \inf_{q\in\mathbb{R}}\; q + \frac{1}{1-\alpha}\mathbb{E}\,(Y-q)^+ \tag{2}$$

was elaborated in [24].

[1] For Y_1 and Y_2 random variables we write $Y_1 \le Y_2$ whenever $Y_1 \le Y_2$ almost everywhere.
[2] The random variable $Y + c$ is $Y + c \cdot \mathbb{1}$ where $\mathbb{1}$ is the constant random variable, $\mathbb{1}(x) = 1$.
[3] Average Value-at-Risk is also called Conditional Value-at-Risk, in an actuarial context Conditional Tail Expectation.

Spectral Risk Measures And Kusuoka Representation

The Average Value-at-Risk, by (1), assigns the same weight ($\frac{1}{1-\alpha}$) to any of the quantiles which arise with probability $p \geq \alpha$. One may assign different weights, which gives rise to the following definition of a weighted Average Value-at-Risk, which is known as spectral risk measure as well (cf. [4, 1]).

Definition 3 (Spectral risk measure and Kusuoka representation)**.** Let $\rho\colon \mathcal{Y} \to \mathbb{R} \cup \{\infty\}$ be a risk measure.

(i) ρ is a *spectral risk measure* provided that

$$\rho(Y) = \int_0^1 \mathrm{V@R}_p(Y)\, h(p)\, \mathrm{d}p \tag{3}$$

with h, the spectral density, satisfying

(a) $h \geq 0$ (to ensure monotonicity),

(b) $1 = \int_0^1 h(p)\, \mathrm{d}p$ (to ensure translation equivariance) and

(c) h nondecreasing (to ensure convexity).

(ii) A representation

$$\rho(Y) = \sup_{\mu \in \mathscr{M}} \int_0^1 \mathrm{AV@R}_p(Y)\, \mu(\mathrm{d}p), \tag{4}$$

where $\mathscr{M}$ is a set of probability measures on $[0,1]$, is called *Kusuoka representation*.

In the setting $\mathcal{A}(Y) = -\rho(-Y)$ spectral risk measures are called *distortion acceptability functionals*. Discussion of distortion acceptability functionals are for example contained in [15] and [29], cf also [3] for applications.

Law Invariant Risk Measures

All risk measures introduced above are already determined by the law (the cumulative distribution function) of the considered random variable Y, for which reason they are called *law invariant*, or sometimes *version independent*.

It was elaborated by Kusuoka ([12, 10]) that every lower semi-continuous and law invariant risk functional on L^∞ has a representation as in (4).

Lemma 4. *Let ρ_h be a spectral risk measure. Then there is a probability measure μ_h on $[0,1]$ such that*

$$\rho_h(Y) = \int_0^1 \mathrm{V@R}_p(Y)\, h(p)\, \mathrm{d}p = \int_0^1 \mathrm{AV@R}_\alpha(Y)\, \mu_h(\mathrm{d}\alpha). \tag{5}$$

Conversely, for any probability measure μ on $[0,1]$ there is a function h such that (5) holds for ρ_h, provided that $\int_0^1 \frac{1}{1-\alpha}\mu(\mathrm{d}\alpha) < \infty$.

Moreover, for every law invariant risk functional ρ there is a set $\mathscr{H}$ of bounded *spectral densities such that $\rho(Y) = \sup_{h \in \mathscr{H}} \rho_h(Y)$.*

Proof. One may associate with the density h of a spectral risk measure ρ_h the measure $\mu_h(A) := h(0)\,\delta_0(A) + \int_A 1 - \alpha\, \mathrm{d}h(\alpha)$ [4] on $[0,1]$. This is a positive measure, as h is nondecreasing so that $\int_A 1 - \alpha\, \mathrm{d}h(\alpha) \geq 0$. And it is a probability measure, as

$$\mu_h([0,1]) = h(0) + \int_0^1 1 - \alpha\, \mathrm{d}h(\alpha) = h(0) + 0 \cdot h(1) - 1 \cdot h(0) + \int_0^1 h(\alpha)\, \mathrm{d}\alpha = 1$$

by integration by parts. With this choice of μ_h it is easily observed that (5) holds, any spectral risk measure ρ_h thus is a combination of AV@Rs at different risk levels α.

[4] Without loss of generality one may require $h(0) = 0$, then $\mu_h(A) = \int_A 1 - \alpha\, \mathrm{d}h(\alpha)$.

As for the converse relation, consider $\rho_\mu(Y) := \int_0^1 \text{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha)$ for some probability measure μ on $[0,1]$. The function $h_\mu(p) := \int_0^p \frac{1}{1-\alpha}\mu(\mathrm{d}\alpha)$ is positive, nondecreasing, and

$$\int_0^1 h_\mu(p)\,\mathrm{d}p = \int_0^1 \int_0^p \frac{1}{1-\alpha}\mu(\mathrm{d}\alpha)\,\mathrm{d}p = \int_0^1 \int_\alpha^1 \frac{1}{1-\alpha}\mathrm{d}p\,\mu(\mathrm{d}\alpha) = \int_0^1 \mu(\mathrm{d}\alpha) = 1.$$

Employing (1) again and Fubini's Theorem reveals

$$\int_0^1 \text{V@R}_p(Y)\,h_\mu(p)\,\mathrm{d}p = \int_0^1 \text{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha) = \rho_\mu(Y),$$

so that ρ_μ is a spectral risk measure with spectral density h_μ.

For the latter statement consider a probability measure μ on $[0,1]$ and define the measures $\mu_n(A) := \mu\left(A \cap \left[0, 1-\frac{1}{n}\right]\right) + \mu\left(\left[1-\frac{1}{n}, 1\right[\right) \cdot \delta_{1-\frac{1}{n}}(A)$. As $\alpha \mapsto \text{AV@R}_\alpha(Y)$ is increasing it follows that $\int_0^1 \text{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha) \geq \int_0^1 \text{AV@R}_\alpha(Y)\,\mu_n(\mathrm{d}\alpha)$, and as $\alpha \mapsto \text{AV@R}_\alpha(Y)$ is continuous and bounded we see that $\int_0^1 \text{AV@R}_\alpha(Y)\,\mu_n(\mathrm{d}\alpha) \to \int_0^1 \text{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha)$ whenever n tends to infinity. Note next that $\int_0^1 \frac{1}{1-p}\mu_n(\mathrm{d}p) \leq \int_0^{1-\frac{1}{n}} n\mu(\mathrm{d}p) + n \cdot \mu\left(\left[1-\frac{1}{n}, 1\right[\right) = n < \infty$. It follows that $h_{\mu;n}(\alpha) := \int_0^\alpha \frac{1}{1-p}\mu_n(\mathrm{d}p) \leq n$ is bounded (by n) and $\rho_\mu(Y) = \sup_n \rho_{h_n}(Y)$. Employing Kusuoka's representation (4) and collecting all spectral functions into $\mathscr{H} := \{h_{\mu;n} : \mu \in \mathscr{M},\, n \in \mathbb{N}\}$ reveals the assertion. □

2.1. Continuity Of The Risk Measure

In the sequel we shall restrict the analysis to the Banach lattice $\mathcal{Y} = L^p(X, \mathcal{F}, P)$ where $1 \leq p \leq \infty$. The axioms imposed on risk measures have strong regularizing properties. For future reference in this paper we repeat here some of the key properties of a risk measure with fixed probability measure P.

Monotonicity, together with translation equivariance in the definition of the risk measure, force ρ to be Lipschitz-continuous on the subspace L^∞, as the following lemma reveals.

Lemma 5. *Suppose that $\mathcal{Y} = L^\infty(X, \mathcal{F}, P)$ and there is one almost surely bounded random variable $\tilde{Y} \in L^\infty$ such that $\rho(\tilde{Y}) < \infty$. Then ρ is finite valued ($\rho(Y) < \infty$)* for any *$Y \in L^\infty$ and ρ has Lipschitz constant 1, that is*

$$|\rho(Y_1) - \rho(Y_2)| \leq \|Y_1 - Y_2\|_\infty.$$

Proof. For the proof we refer to Lemma 4.3 in [7]. □

Similar statements hold for the general situation $1 \leq p < \infty$ as well. They are more involved than the latter Lemma, we cite a precise statement from [28, Proposition 6.7], its proof is built on Baire's lemma.

Proposition 6 ([28, Proposition 6.7])**.** *Suppose that $\rho : L^p \to \bar{\mathbb{R}}$ ($1 \leq p < \infty$) is monotone, translation equivariant and convex, and further let $\{\rho < \infty\}$ have non-empty interior. Then ρ is finite valued and continuous.*

2.2. Subdifferential

Typical regularity properties of convex functionals include not only continuity, in many situations they are even subdifferentiable.

Definition. The *subdifferential* $\partial\rho(Y)$ of an $\mathbb{R}-$valued function $\rho : \mathcal{Y} \to \mathbb{R}$ is the collection of all *subgradients*,

$$\partial\rho(Y) = \{Z^* \in \mathcal{Y}^* : \rho(Y') - \rho(Y) \geq Z^*(Y' - Y) \text{ for all } Y' \in \mathcal{Y}\}.$$

ρ is called *sub-differentiable* at Y iff there is at least one subgradient, that is $\partial\rho(Y)$ is non-empty.

For the Banach space $\mathcal{Y} = L^p$ $(1 \leq p < \infty)$ one may choose $\mathcal{Y}^*$ to be the dual space of $\mathcal{Y}$. The inner product is then $Z^*(Y) = \mathbb{E}\, Z \cdot Y$ for some $Z \in L^{p'}$ where p' is the conjugate exponent, $\frac{1}{p} + \frac{1}{p'} = 1$.

As for $\mathcal{Y} = L^\infty$ we follow common practice and choose $\mathcal{Y}^* := L^1$ for the case $p = \infty$, L^∞ thus is not paired with its dual space, but with the pre-dual instead.

Proposition 7 ([28, Proposition 6.5])**.** *Let $\rho : L^p \to \mathbb{R}$ ($p \le \infty$) be a real-valued, monotone and convex risk measure. Then ρ is continuous and subdifferentiable on the entire L^p.*

The preceding lemma and propositions elaborate on the situations where subgradients are available. The subgradients can be used to derive a statement on continuity, as the following lemma reveals:

Lemma 8. *Let ρ be a risk measure on L^p, $1 \le p < \infty$. Then*

$$|\rho(Y) - \rho(Y')| \le \|Y - Y'\|_p \cdot \max\left\{\inf_{Z\in\partial\rho(Y)} \|Z\|_{p'},\ \inf_{Z\in\partial\rho(Y')} \|Z\|_{p'}\right\},$$

in particular

$$|\rho(Y) - \rho(Y')| \le \|Y - Y'\|_p \cdot \sup_{Z\in\partial\rho(Y), Z\in\partial\rho(Y')} \|Z\|_{p'}.$$

Proof. Let $Z \in \partial\rho(Y)$ be chosen, that is

$$\rho(Y') - \rho(Y) \ge \mathbb{E}Z \cdot (Y' - Y).$$

Then by Hölder's inequality

$$\rho(Y) - \rho(Y') \le \mathbb{E}Z \cdot (Y - Y') \le \|Y - Y'\|_p \cdot \|Z\|_{p'}.$$

Interchanging the role of Y and Y' gives the first assertion, from which the other follows. □

2.3. Legendre-Fenchel Transformation

Associated with the subdifferential of a convex function ρ is the Legendre-Fenchel transformation. The *conjugate function* ρ^* *of* ρ is defined as

$$\rho^*(Z^*) := \sup_{Y\in\mathcal{Y}} Z^*(Y) - \rho(Y)$$

on the space $\mathcal{Y}^*$. Then, by the Fenchel-Moreau Theorem (cf. – for example – [26] or [23]), ρ has the representation

$$\rho(Y) = \sup_{Z^*\in\mathcal{Y}^*} Z^*(Y) - \rho^*(Z^*)$$

provided that ρ is lower semi-continuous.

We shall exploit this representation in various investigations below for functionals ρ defined on some $L^p(X, \mathcal{F}, P)$.

2.4. Wasserstein Metric

On a Polish space (X, d) consider the probability measures on its Borel sets. The collection of all probability measures P, which satisfy for some – and thus any – $x_0 \in X$ the moment-like condition

$$\int_\Omega d(x_0, x)^r P(\mathrm{d}x) < \infty$$

is denoted by $\mathcal{P}_r(X; d)$.

On this space of probability measures define the function

$$\mathsf{d}_r(P, P'; d) := \left(\inf_\pi \int_{X\times X} d(x, x')^r \pi(\mathrm{d}x, \mathrm{d}x')\right)^{\frac{1}{r}} \qquad (r \ge 1), \tag{6}$$

where the infimum is taken over all (bivariate) probability measures π on $X \times X$ with marginals P and P', that is

$$\pi(A \times X) = P(A) \text{ and } \pi(X \times B) = P'(B)$$

for all measurable sets $A \subset X$ and $B \subset X$ (in symbols $\pi_1 = P$, $\pi_2 = P'$). We shall call such a feasible measure π a *transport plan.*

d_r is called r^{th}-*Wasserstein distance.* It is well-defined, as for example the product measure $\pi := P \otimes P'$ has the required marginals and hence

$$\mathsf{d}_r\left(P, P'; d\right)^r \leq \int_X \int_X d\left(x, x'\right)^r P\left(\mathrm{d}x\right) P'\left(\mathrm{d}x'\right).$$

A very comprehensive and beautiful discussion and treatment of the distance d_r can be found in Villani's books ([32] and [33]), but we want to mention the books by Rachev and Rüschendorf as well, [22] and [8, 16].

We shall use the properties that the infimum in (6) is actually attained, and $\mathsf{d}_r\left(.,.; d\right)$ turns out to be a metric on the space $\mathcal{P}_r\left(X; d\right)$, so in particular satisfies the triangle inequality

$$\mathsf{d}_r\left(P, P'; d\right) \leq \mathsf{d}_r\left(P, \tilde{P}; d\right) + \mathsf{d}_r\left(\tilde{P}, P'; d\right).$$

We are using the symbol d for the distance in the original space X, and $\mathsf{d}_r\left(.; d\right)$ to account for the distance on probabilities in $\mathcal{P}_r\left(X; d\right)$ induced by d. However, as no confusion may occur we will omit the distance d in the sequel and simply write $\mathsf{d}_r\left(P, P'\right) = \mathsf{d}_r\left(P, P'; d\right)$ for the distance on $\mathcal{P}_r$ specified by d.

Remark 9. In honor of G. Monge[5] (cf. [13]) and L. Kantorovich[6] (cf. [11]) the distance d_r is sometimes called *Monge-Kantorovich distance* of order r, and d_2 is called quadratic Wasserstein distance as well. Moreover, the distance d_1 is also called Kantorovich-Rubinstein distance and sometimes denoted $\mathsf{d}_{KA} := \mathsf{d}_1$.

3. Continuity With Respect To Changing The Measure

Let $Y : X \to \mathbb{R}$ be a function, measurable with respect to $\mathcal{F}$. Then $Y \in \mathcal{Y} = L^p\left(X, \mathcal{F}, P\right)$, provided that $\int |X|^p \mathrm{d}P < \infty$, and the analysis in (the previous) Section 2 is concerned with the mapping

$$Y \mapsto \rho\left(Y\right).$$

All risk functionals described above can be used now to compute $\rho\left(Y\right)$, however, the result will depend on the probability measure P employed (cf. (1), (2), (3) and (4)); we shall make this evident by writing $\rho_P\left(Y\right)$. The natural question, which arises in this context, is the question of *how* ρ_P will vary, whenever the measure P is changed. This is the topic of this section, we shall investigate the continuity properties of

$$P \mapsto \rho_P\left(Y\right).$$

Theorem 10. *Let $Y : X \to \mathbb{R}$ be Hölder continuous with constant L_β, $\left|Y\left(x\right) - Y\left(x'\right)\right| \leq L_\beta \cdot d\left(x, x'\right)^\beta$ for some $\beta \leq 1$ and ρ law invariant. Assume that $Y \in L^p\left(P\right)$ and $Y \in L^p\left(P'\right)$. Then*

$$\rho_P\left(Y\right) - \rho_{P'}\left(Y\right) \leq L_\beta\left(Y\right) \cdot \mathsf{d}_r\left(P, P'\right) \cdot \inf_{\rho_P^*(Z) < \infty} \|Z\|_{P, r_\beta}$$

and a fortiori

$$\left|\rho_P\left(Y\right) - \rho_{P'}\left(Y\right)\right| \leq L_\beta\left(Y\right) \cdot \mathsf{d}_r\left(P, P'\right) \cdot \sup_{\rho_P^*(Z) < \infty} \|Z\|_{P, r_\beta},$$

where $r_\beta \geq \frac{r}{r - \beta}$.

[5] Gaspard Monge (1746 - 1818) investigated how to efficiently construct dugouts.
[6] Leonid Kantorovich was awarded the price in Economic Sciences in Memory of Alfred Nobel in 1975.

Proof. Recall the representation (cf. [26])

$$\rho_P(Y) = \sup_{Z\in\mathcal{Y}^*} \mathbb{E}_P YZ - \rho_P^*(Z) = \sup_{Z\in\partial\rho(0)} \mathbb{E}_P YZ.$$

Note that $\partial\rho(0)$ is convex, closed (in weak* and norm topology) and bounded (because ρ is continuous), hence weakly* compact. Let Z_Y be chosen with $\rho_P^*(Z_Y) < \infty$ such that $\rho_P(Y) = \mathbb{E}_P YZ_Y - \rho_P^*(Z_Y)$. For any Z' with $\rho_{P'}(Z') < \infty$ thus

$$\begin{aligned}\rho_P(Y) - \rho_{P'}(Y) &\le \mathbb{E}_P YZ_Y - \rho_P^*(Z_Y) - \mathbb{E}_{P'} YZ' + \rho_{P'}(Z') \qquad (7)\\ &= \int Y(x)\, Z_Y(x)\, P(\mathrm{d}x) - \int Y(x')\, Z'(x')\, P'(\mathrm{d}x') + \rho_{P'}^*(Z') - \rho_P^*(Z_Y)\\ &= \int Y(x)\, Z_Y(x) - Y(x')\, Z'(x')\, \pi(\mathrm{d}x, \mathrm{d}x') + \rho_{P'}^*(Z') - \rho_P^*(Z_Y)\end{aligned}$$

where π has marginals $\pi_1 = P$ and $\pi_2 = P'$. Taking the infimum over all these bivariate measures π and random variables Z' which have the same distribution as Z_Y under their respective measures (i.e. $P(Z_Y \le z) = P'(Z' \le z)$) one obtains

$$(7) \quad \le \quad \inf_{\pi,\, Z\sim Z_Y} \int Y(x)\, Z_Y(x) - Y(x')\, Z(x')\, \pi(\mathrm{d}x, \mathrm{d}x') + \rho_{P'}^*(Z) - \rho_P^*(Z_Y).$$

To ensure that such a random variable Z exists consider the cumulative distribution functions $G(z) := P(Z_Y \le z)$ and $G'(z) := P'(Z' \le z)$ with respective quantiles $G^{-1}(p) = \inf\{u\colon G(u) \ge p\}$ and $G'^{-1}(p) = \inf\{u\colon G'(u) \ge p\}$. Let $U'\colon X' \to \mathbb{R}$ be independent of Z' and uniformly distributed ($P'(U' \le u) = u$), and define the random variable $F(Z', U') := (1-U')\cdot G'^{-}(Z') + U'\cdot G'(Z')$ where $G'^{-}(z) = P'(Z' < z)$. It is well-known (cf. [6] for the so-called *generalized quantile transform*) that $F(Z', U')$ is uniformly distributed and moreover $Z' = G'^{-1}(F(Z', U'))$ $\;P'-$almost everywhere.

With these preparations define the random variable $Z := G^{-1}(F(Z', U'))$. It holds that

$$\begin{aligned}P(Z_Y \le z) &= G(z) = P'(F(Z', U') \le G(z))\\ &\le P'\left(G^{-1}(F(Z', U')) \le G^{-1}(G(z))\right)\\ &\le P'\left(G^{-1}(F(Z', U')) \le z\right) = P'(Z \le z)\end{aligned}$$

because $F(Z', U')$ is uniform under P' and $G^{-1}(G(z)) \le z$. Moreover

$$\begin{aligned}P'(Z \le z) &= P'\left(G^{-1}(F(Z', U')) \le z\right)\\ &\le P'(F(Z', U') \le G(z)) = G(z) = P(Z_Y \le z),\end{aligned}$$

so that $P(Z_Y \le z) = P'(Z \le z)$, that is Z and Z_Y have the same distribution given their respective measures and Z_Y can be replicated.

ρ is law invariant by assumption, hence so is ρ^* and thus $\rho_{P'}^*(Z) = \rho_P^*(Z_Y)$ (cf. [27]) and

$$(7) \quad \le \quad \inf_{\pi,\, Z\sim Z_Y} \int Y(x)\, Z_Y(x) - Y(x')\, Z(x')\, \pi(\mathrm{d}x, \mathrm{d}x').$$

The infimum over Z in the latter expression is attained for some Z which is coupled in an comonotone way with Y (cf. [14]). By employing Hoeffding's Lemma ([9]) one may allow the random variable Z to be defined on the larger space $X \times X$ without impacting the inf, the latter expression thus may be rewritten as

$$(7) \quad \le \quad \inf_{\pi,\, Z\sim Z_Y} \int Y(x)\, Z_Y(x) - Y(x')\, Z(x, x')\, \pi(\mathrm{d}x, \mathrm{d}x').$$

Hence, for the special choice $Z=Z_Y$,

$$\begin{aligned}(7) &\leq \inf_\pi \int Y(x)\,Z_Y(x)-Y(x')\,Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x')\\ &\leq \inf_\pi \int \left|Y(x)-Y(x')\right| Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x')\\ &\leq \inf_\pi \int L_\beta\cdot d(x,x')^\beta\, Z_Y(x)\,\pi(\mathrm{d}x,\mathrm{d}x').\end{aligned}$$

For the relation $\frac{1}{\frac{r}{\beta}}+\frac{1}{\frac{r}{r-\beta}}=1$ one may apply Hölder's inequality to obtain

$$\begin{aligned}(7) &\leq L_\beta\cdot\inf_\pi\left(\int d^r\mathrm{d}\pi\right)^{\frac{\beta}{r}}\cdot\left(\int Z_Y^{\frac{r}{r-\beta}}\mathrm{d}P\right)^{\frac{r-\beta}{r}}\\ &= L_\beta\cdot \mathsf{d}_r(P,P')^\beta\cdot\left\|Z_Y\right\|_{P,\frac{r}{r-\beta}},\end{aligned}$$

which proves the first statement.

The other expression of the assertion follows by interchanging the roles of P and P'. □

The following corollary quantifies the constants for the Average Value-at-Risk (cf. [20] for an initial statement in this direction) and for spectral risk functionals.

Corollary 11. *Let $Y\colon X\to\mathbb{R}$ be Hölder continuous with constant L_β, $\left|Y(x)-Y(x')\right|\leq L_\beta\cdot d(x,x')^\beta$ for some $\beta\leq 1$. Then*

$$\left|\rho_{h,P}(Y)-\rho_{h,P'}(Y)\right|\leq L_\beta(Y)\cdot\mathsf{d}_r(P,P')\cdot\left\|h\right\|_{r_\beta},$$

where ρ_h is a spectral risk measure and $\left\|h\right\|_r=\left(\int_0^1 h^r\right)^{\frac{1}{r}}$ is the norm for $h\in L^r([0,1],\lambda)$ on the standard space $([0,1],\lambda)$ with Lebesgue measure λ. In particular,

$$\left|\mathrm{AV@R}_{\alpha,P}(Y)-\mathrm{AV@R}_{\alpha,P'}(Y)\right|\leq L_\beta(Y)\cdot\mathsf{d}_r(P,P')\cdot(1-\alpha)^{-\frac{\beta}{r}}.$$

Proof. To prove the assertion we employ the relation

$$\rho_h(Y)=\sup\left\{\mathbb{E}\,Y h(U): U\colon X\to[0,1]\ \text{ uniformly distributed}\right\} \tag{8}$$

from [19, Proposition 2.65]. As

$$\left\|h(U)\right\|_{r_\beta}^{r_\beta}=\int_X h(U)^{r_\beta}\,\mathrm{d}P=\int_0^1 h(u)^{r_\beta}\,\mathrm{d}u=\left\|h\right\|_{r_\beta}^{r_\beta}$$

for *any* uniformly distributed random variable $U\colon X\to[0,1]$ with law $P(U\leq u)=u$ the result follows in view of Theorem 10.

For the Average Value-at-Risk recall the representations (cf. [19, 15])

$$\begin{aligned}\mathrm{AV@R}_\alpha(Y) &= \sup\left\{\mathbb{E}\,YZ\colon 0\leq Z\leq\frac{1}{1-\alpha}\text{ and }\mathbb{E}Z=1\right\}\\ &= \sup\left\{\mathbb{E}\,Yh_\alpha(U): U\text{ uniformly distributed}\right\},\end{aligned}$$

where h_α is the distortion function $h_\alpha(u)=\begin{cases}\frac{1}{1-\alpha} & \text{if } u\geq\alpha\\ 0 & \text{if } u<\alpha\end{cases}$. The assertion is immediate, as

$$\left\|h_\alpha\right\|_{r_\beta}=\left(\int_\alpha^1\frac{1}{(1-\alpha)^{r_\beta}}\mathrm{d}P\right)^{\frac{1}{r_\beta}}=\left((1-\alpha)^{1-r_\beta}\right)^{\frac{1}{r_\beta}}=(1-\alpha)^{\frac{1}{r_\beta}-1}=(1-\alpha)^{\frac{r-\beta}{r}-1}=(1-\alpha)^{-\frac{\beta}{r}}.$$

□

Corollary 12. *Let $Y\colon X\to\mathbb{R}$ be Hölder continuous with constant L_β, $\left|Y(x)-Y(x')\right|\le L_\beta\cdot d(x,x')^\beta$ for some $\beta\le 1$ and ρ have the Kusuoka representation $\rho(Y)=\sup_{\mu\in\mathscr{M}}\int \mathrm{AV@R}_\alpha(Y)\,\mu(\mathrm{d}\alpha)$. Then ρ is continuous with respect to the Wasserstein distance provided that*

$$K:=\sup_{\mu\in\mathscr{M}}\int_0^1\frac{\mu(\mathrm{d}\alpha)}{(1-\alpha)^{\frac{\beta}{r}}}$$

is finite:

$$\left|\rho_P(Y)-\rho_{P'}(Y)\right|\le \mathsf{d}_r(P,P')\cdot L_\beta(Y)\cdot\sup_{\mu\in\mathscr{M}}\int_0^1(1-\alpha)^{-\frac{\beta}{r}}\mu(\mathrm{d}\alpha)\,.$$

Remark 13. As $\int_0^1(1-p)^{-\frac{\beta}{r}}\mu(\mathrm{d}p)\le\int_0^1\frac{1}{1-p}\mu(\mathrm{d}p)$ the statement includes in particular all spectral risk measures due to Lemma 4.

Proof. For the straightforward proof choose $\mu_\varepsilon\in\mathscr{M}$ such that

$$\int_0^1\mathrm{AV@R}_{\alpha,P'}(Y)\,\mu_\varepsilon(\mathrm{d}\alpha)>\rho_{P'}(Y)-\varepsilon.$$

Then

$$\begin{aligned}
\rho_{P'}(Y)-\rho_P(Y)\le&\\
&\le\int_0^1\mathrm{AV@R}_{\alpha,P'}(Y)\,\mu_\varepsilon(\mathrm{d}\alpha)-\int_0^1\mathrm{AV@R}_{\alpha,P}(Y)\,\mu_\varepsilon(\mathrm{d}\alpha)+\varepsilon\\
&=\int_0^1\mathrm{AV@R}_{\alpha,P'}(Y)-\mathrm{AV@R}_{\alpha,P}(Y)\,\mu_\varepsilon(\mathrm{d}\alpha)+\varepsilon\\
&\le L_\beta(Y)\cdot\mathsf{d}_r(P,P')^\beta\cdot\int_0^1\frac{\mu^\varepsilon(\mathrm{d}\alpha)}{(1-\alpha)^{\frac{\beta}{r}}}+\varepsilon\le K\cdot L_\beta(Y)\cdot\mathsf{d}_r(P,P')^\beta+\varepsilon.
\end{aligned}$$

Let $\varepsilon\to 0$ and interchange the role of P and P' to observe the desired assertion. □

4. Worst Probability Measures

The problem investigated in the previous section can be restated as

$$\begin{array}{rl}
\underset{(\text{in } P')}{\text{maximize}} & \rho_{P'}(Y)\\
\text{subject to} & \mathsf{d}_r(P,P')\le K,\\
& P'\in\mathcal{P}_r(X)\,,
\end{array}\tag{9}$$

where the minimum (infimum) is over all probability measures P' whose Wasserstein-distance to P do not exceed K. An upper bound for the objective in (9) was found in Theorem 10, as

$$\rho_{P'}(Y)\le\rho_P(Y)+L(Y)\cdot K\cdot\inf_{\rho_P^*(Z)>-\infty}\|Z\|_{r'}\,.\tag{10}$$

We shall develop our ideas further now and give some general situations for which this bound is sharp. It is possible in some situations to characterize the probability measure, for which problem (9) attains its maximal value. It will turn out that these measures have an interesting description as a transport map. Moreover, situations will occur where the bounds are not attained, and for some of them we will prove that no such bound exists in general.

Most of the results in this section are based on linear functionals Y. This is motivated from finance and was already addressed in Remark 2.

Before turning to the general case (9) we start with the simpler problem

$$\begin{array}{rl} \text{maximize} & \\ \text{(in } P'\text{)} & \mathbb{E}_{P'}Y \\ \text{subject to} & \mathsf{d}_r\left(P,P'\right)\le K, \\ & P'\in\mathcal{P}_r\left(X\right) \end{array} \tag{11}$$

to develop the main ideas.

As above, let $(X,\mathcal{F},P)$ denote a probability triple. We shall assume in addition that X is a linear space – in the simplest case $X=\mathbb{R}^d$ – equipped with an appropriate norm $\|\cdot\|$.

On this space there is the usual notion of a dual X^*, collecting all continuous, linear functionals on X. Recall that any linear functional $Y\colon X\to\mathbb{R}$ is a random variable itself, and $Y\in X^*$. The Lipschitz constant of Y, $L\left(Y\right)=\sup\frac{|Y(x)|}{\|x\|}=\|Y\|$ is the norm in the dual space.

Lemma 14. *Let $(X,\|\cdot\|)$ be a reflexive Banach space and Y a linear functional. Then, for all $1\le r<\infty$, the bound*

$$\mathbb{E}_PY+K\cdot L\left(Y\right)$$

for (11) is sharp: There exists $x_Y\in X$, $\|x_Y\|=1$, such that the maximizing measure in (11) is the push-forward (image measure of P)[7]

$$P^*:=T_*\left(P\right)=P\circ T^{-1}$$

for the transport map (translation map) $T\left(x\right):=x+K\cdot x_Y$.

Proof. By Kantorovich's duality theorem (cf. [17])

$$\begin{aligned} \left|\mathbb{E}_{P'}Y-\mathbb{E}_PY\right| &\le L\left(Y\right)\cdot\mathsf{d}_{KA}\left(P,P'\right) \\ &= L\left(Y\right)\cdot\mathsf{d}_1\left(P,P'\right)\le L\left(Y\right)\cdot\mathsf{d}_r\left(P,P'\right) \end{aligned} \tag{12}$$

for $r\ge1$, establishing that $\mathbb{E}_{P'}Y\le\mathbb{E}_PY+L\left(Y\right)\cdot\mathsf{d}_r\left(P,P'\right)\le\mathbb{E}_PY+L\left(Y\right)\cdot K$.

To observe that this bound is sharp recall that X and Y are linear. By the Hahn-Banach theorem $\|Y\|=x^{**}\left(Y\right)$ for some $x^{**}\in X^{**}$ with $\|x^{**}\|=1$, and as X is reflexive one may identify x^{**} with some $x_Y\in X$, which satisfies

$$\|x_Y\|=1\text{ and }Y\left(x_Y\right)=\|Y\|=L\left(Y\right). \tag{13}$$

Define $T\left(x\right):=x+K\cdot x_Y$ and $P^*:=T_*\left(P\right)=P\circ T^{-1}$ as above, and the push-forward transport plan as

$$\pi:=\left(\mathrm{id}\times T\right)_*\left(P\right),$$

where $\mathrm{i}\times T$ is the mapping $\left(\mathrm{id}\times T\right)\left(x\right):=\left(x,T\left(x\right)\right)$, that is $\pi\left(A\times B\right)=P\left(A\cap T^{-1}\left(B\right)\right)$.

The Wasserstein distance of P and P^* is bounded by K, because

$$\begin{aligned} \mathsf{d}_r\left(P,P^*\right)^r &\le \int d\left(x,x'\right)^r\pi\left(\mathrm{d}x,\mathrm{d}x'\right)=\int d\left(x,T\left(x\right)\right)^rP\left(\mathrm{d}x\right) \\ &= \int\left\|K\cdot x_Y\right\|^rP\left(\mathrm{d}x\right)=K^r\cdot\left\|x_Y\right\|^r=K^r. \end{aligned}$$

Given this measure P^* the objective of the primal function is

$$\begin{aligned} \mathbb{E}_{P^*}Y &= \int Y\left(x\right)P\circ T^{-1}\left(\mathrm{d}x\right)=\int Y\left(T\left(x\right)\right)P\left(\mathrm{d}x\right) \\ &= \int Y\left(x+K\cdot x_Y\right)P\left(\mathrm{d}x\right)=\mathbb{E}_PY+K\cdot Y\left(x_Y\right) \\ &= \mathbb{E}_PY+K\cdot L\left(Y\right), \end{aligned}$$

which is the maximum value we can achieve in view of (12). □

[7] Villani uses the notation $T\#\mathbb{P}:=T_*\left(\mathbb{P}\right)$ for the push-forward measure, the notation $\mathbb{P}^T$ is in frequent use as well.

We shall now turn to the general case.

Theorem 15 (Optimal transport plan)**.** *Let Y be linear on a reflexive Banach space $(X, \|\cdot\|)$ and ρ law invariant.*

(i) *The problem*

$$\begin{array}{rl} \textit{maximize} & \\ \textit{(in } P'\textit{)} & \rho_{P'}(Y) \\ \textit{subject to} & \mathsf{d}_r(P', P) \le K \end{array} \tag{14}$$

has minimal value $\rho_P(Y) + K \cdot L(Y) \cdot \min_{Z \in \partial\rho_P(Y)} \|Z\|_{r'}$.

(ii) *For $1 < r < \infty$ there is a transport map T such that*

$$P^* := T_*(P) = P \circ T^{-1}$$

minimizes (14)*.*

Proof. Choose $x_Y \in X$ as in (13) and select

$$Z_Y \in \operatorname{argmin}\left\{\|Z\|_{\frac{r}{r-1}} : Z \in \partial\rho_P(Y)\right\}. \tag{15}$$

Without loss of generality we assume that the respective argmin −set is nonempty, as otherwise one may continue with $Z_Y \in \partial\rho_P(Y)$ almost (up to some $\varepsilon > 0$) minimal in (15).

Define the transport map

$$\begin{aligned} T(x) &:= x + K \cdot \left|\frac{Z_Y(x)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y \\ &= x + \frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}} \cdot |Z_Y(x)|^{\frac{r}{r-1}-2} \cdot Z_Y(x) \cdot x_Y, \end{aligned}$$

and again consider the transport plan

$$\pi := (\mathrm{id} \times T)_*(\mathbb{P})$$

with marginals P and P^*. Observe that

$$\begin{aligned} \mathsf{d}_r(P, P^*)^r &\le \int \|x - x'\|^r \, \pi(\mathrm{d}x, \mathrm{d}x') = \int \|x - T(x)\|^r \, P(\mathrm{d}x) \\ &= \int \left\| K \cdot \left|\frac{Z_Y(x)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot x_Y \right\|^r P(\mathrm{d}x) \\ &= \frac{K^r}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{r}{r-1}}} \cdot \int |Z_Y|^{\frac{r}{r-1}} \, \mathrm{d}P = K^r, \end{aligned}$$

that is to say P^* has an accepted distance from P.

To observe that the transport map T is injective choose x_1 and x_2 and note that

$$\begin{aligned} &T(x_1) - T(x_2) = \\ &\quad = x_1 - x_2 + \\ &\quad + K \cdot x_Y \left(\left|\frac{Z_Y(x_1)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x_1) - \left|\frac{Z_Y(x_2)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x_2) \right). \end{aligned}$$

One may assume – without loss of generality – that $Z_Y(x_1) \ge Z_Y(x_2)$ (otherwise reverse x_1 and x_2) and distinguish the following two cases:

(i) If $Z_Y(x_1) = Z_Y(x_2)$, then $T(x_1) - T(x_2) = x_1 - x_2$ and thus T is injective on this subset.
(ii) If $Z_Y(x_1) > Z_Y(x_2)$, then $Y(x_1) \geq Y(x_2)$ a.s., because Y and Z_Y are coupled in a comonotone way. In this situation

$$\begin{aligned} & Y(T(x_1) - T(x_2)) = \\ & \quad = Y(x_1 - x_2) + \\ & \quad\quad + K\|Y\| \left(\left| \frac{Z_Y(x_1)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x_1) - \left| \frac{Z_Y(x_2)}{\|Z_Y\|_{\frac{r}{r-1}}} \right|^{\frac{1}{r-1}} \cdot \operatorname{sign} Z_Y(x_2) \right) \\ & \quad > Y(x_1 - x_2) \geq 0 \end{aligned}$$

because the map $x \mapsto \operatorname{sign}(x) \cdot |x|^{\frac{1}{r-1}}$ is increasing.

Hence, $T(x_1) \neq T(x_2)$ unless $x_1 = x_2$.

Define the random variable

$$Z_Y^T := \mathbb{E}[Z_Y | T]$$

by conditional expectation. Due to its definition Z_Y^T obeys the defining property

$$\int_{T^{-1}(B)} Z_Y \mathrm{d}P = \int_{T^{-1}(B)} \mathbb{E}[Z_Y | T] \circ T \mathrm{d}P = \int_B \mathbb{E}[Z_Y | T] \mathrm{d}T_*(P) = \int_B Z_Y^T \mathrm{d}P^* \tag{16}$$

for all measurable sets B (cf. [34]). Notice that

$$\int_{T^{-1}(B)} Z_Y \mathrm{d}P = \int_{T^{-1}(B)} Z_Y^T \circ T \mathrm{d}P = \int_B Z_Y^T \mathrm{d}P^* = \int_B Z_Y^T \mathrm{d}T_*(P)$$

by the change of variable formula again for all measurable sets B, thus

$$Z_Y = Z_Y^T \circ T$$

$P-$almost everywhere, and

$$Z_Y^T = Z_Y \circ T^{-1}$$

P^*-almost everywhere as T is injective.

One further deduces from (16) that

$$\mathbb{E}_P Z_Y \cdot \mathbb{1}_{\{Z_Y \geq q\}} = \int_{T^{-1}T(\{Z_Y \geq q\})} Z_Y \mathrm{d}P = \int_{T\{Z_Y \geq q\}} Z_Y^T \mathrm{d}P^* = \mathbb{E}_{P^*} Z_Y^T \cdot \mathbb{1}_{\left\{Z_Y^T \geq q\right\}}$$

and hence

$$q + \frac{1}{1-\alpha} \mathbb{E}_P (Z_Y - q)^+ = q + \frac{1}{1-\alpha} \mathbb{E}_{P^*} \left(Z_Y^T - q\right)^+,$$

which is a well-known identity – cf. (2). Taking the infimum with respect to q, it will be attained for the same q at the left and at the right:

$$\begin{aligned} G_{Z_Y}^{-1}(\alpha) &= \min \operatorname*{argmin}_q \, q + \frac{1}{1-\alpha} \mathbb{E}_P (Z_Y - q)^+ \\ &= \min \operatorname*{argmin}_q \, q + \frac{1}{1-\alpha} \mathbb{E}_{P^*} \left(Z_Y^T - q\right)^+ = G_{Z_Y^T}^{-1}(\alpha) \end{aligned}$$

(a.e.) and so it follows that Z_Y and Z_Y^T have the same cumulative distribution function under their respective measures, $P(Z_Y \leq z) = P^*\left(Z_Y^T \leq z\right)$, so finally $\rho_{P^*}^*\left(Z_Y^T\right) = \rho_P^*(Z_Y)$.

As Z_Y is optimal by (15), $\rho_P(Y) = \mathbb{E}_P Y \cdot Z_Y - \rho_P^*(Z_Y)$ and thus

$$\begin{aligned} \rho_{P^*}(Y) - \rho_P(Y) &\geq \mathbb{E}_{P^*} Y \cdot Z_Y^T - \rho_{P^*}^*\left(Z_Y^T\right) - \mathbb{E}_P Y \cdot Z_Y + \rho_P^*(Z_Y) \\ &= \mathbb{E}_{P^*} Y \cdot Z_Y^T - \mathbb{E}_P Y \cdot Z_Y \\ &= \mathbb{E}_P (Y \circ T) \cdot Z_Y - \mathbb{E}_P Y \cdot Z_Y = \mathbb{E}_P Y (T - \mathrm{id}) \cdot Z_Y, \end{aligned}$$

by linearity of Y. Using $Y(x_Y) = \|Y\| = L(Y)$ one further finds that

$$\begin{aligned}
\rho_{P^*}(Y) - \rho_P(Y) \geq \\
&\geq \mathbb{E}_P Y\left(K \cdot \left|\frac{Z_Y(x)}{\|Z_Y\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \operatorname{sign} Z_Y(x) \cdot x_Y\right) \cdot Z_Y \\
&= \frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}} \|Y\| \cdot \mathbb{E}_P |Z_Y|^{\frac{1}{r-1}} \cdot |Z_Y| = \frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}} \|Y\| \cdot \mathbb{E}_P |Z_Y|^{\frac{r}{r-1}} \\
&= \frac{K}{\|Z_Y\|_{\frac{r}{r-1}}^{\frac{1}{r-1}}} \cdot L(Y) \cdot \|Z_Y\|_{\frac{r}{r-1}}^{\frac{r}{r-1}} \geq K \cdot L(Y) \cdot \min_{Z \in \partial\rho(Y)} \|Z\|_{\frac{r}{r-1}},
\end{aligned}$$

whence

$$\rho_{P^*}(Y) - \rho_P(Y) \geq K \cdot L(Y) \cdot \min_{Z \in \partial\rho(Y)} \|Z\|_{\frac{r}{r-1}}.$$

In view of (10) this is smallest difference achievable. □

The situation is a bit more involved for the Kantorovich distance, $r = 1$.

Theorem 16. *Problem* (14) *has an optimal solution provided that* $P(Z_Y = \|Z_Y\|) > 0$, *where* Z_Y *is as in* (15). *The corresponding transport map is*

$$T(x) := x + K \cdot \frac{\mathbb{1}_{\{|Z_Y| = \|Z_Y\|_\infty\}}(x)}{\mathbb{P}(|Z_Y| = \|Z_Y\|_\infty)} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y.$$

Proof. For the Kantorovich distance ($r = 1$) the proof needs a slight modification, it may read as follows:

$$\begin{aligned}
\mathsf{d}_{KA}(P, P^*) &\leq \int \|x - T(x)\| P(\mathrm{d}x) \\
&= \int \left\| K \cdot \mathbb{1}_{\{|Z_Y| = \|Z_Y\|_\infty\}}(x) \frac{\operatorname{sign} Z_Y(x)}{\mathbb{P}(|Z_Y| = \|Z_Y\|_\infty)} \cdot x_Y \right\| P(\mathrm{d}x) \\
&= K \cdot \int \frac{\mathbb{1}_{\{|Z_Y| = \|Z_Y\|_\infty\}}(x)}{P(|Z_Y| = \|Z_Y\|_\infty)} P(\mathrm{d}x) = K
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\rho_{P^*}(Y) - \rho_P(Y) &= \mathbb{E}_P Y\left(K \cdot \mathbb{1}_{\{|Z_Y| = \|Z_Y\|_\infty\}} \frac{\operatorname{sign} Z_Y}{P(|Z_Y| = \|Z_Y\|_\infty)} \cdot x_Y\right) \cdot Z_Y \\
&= K \cdot \|Y\| \cdot \int \frac{\mathbb{1}_{\{|Z_Y| = \|Z_Y\|_\infty\}} |Z_Y|}{P(|Z_Y| = \|Z_Y\|_\infty)} \mathrm{d}P \\
&= K \cdot L(Y) \cdot \|Z_Y\|_\infty \geq K \cdot L(Y) \cdot \min_{Z_Y \in \partial\rho(Y)} \|Z_Y\|_\infty,
\end{aligned}$$

which establishes the result in this particular case. □

5. The Spectral Risk Measure ρ_h

As was elaborated in Theorem 15 the optimal transport map may always be given provided that $r > 1$. For $r = 1$ the additional requirement

$$P(|Z_Y| = \|Z_Y\|_\infty) > 0$$

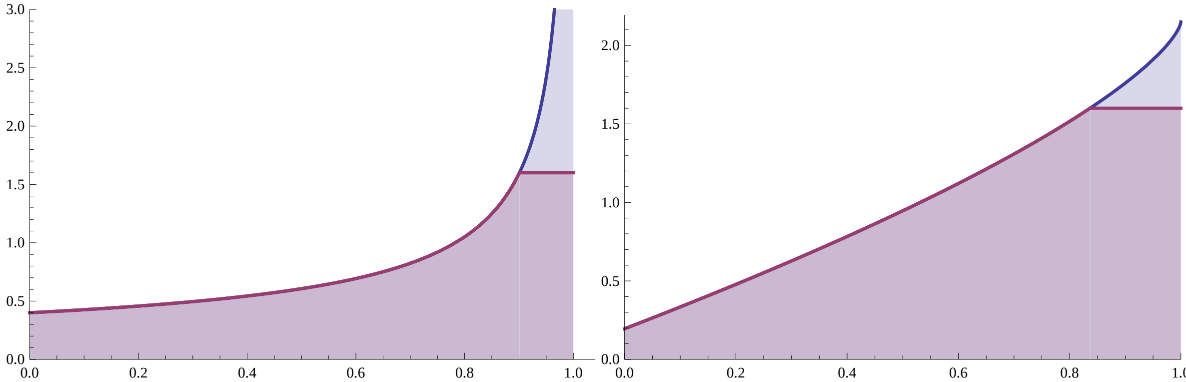

Figure 1: Exemplary shape for a bounded, and an unbounded spectral density h. The area under both charts is one.

is needed to guarantee existence of the furthermost measure. We shall continue the discussion at this point and elaborate the continuity properties for the Kantorovich distance further.

Recall that $Z_Y = h(U)$ (cf. (8)) for some uniform distribution U for the spectral risk measure. The latter condition $P(|Z_Y| = \|Z_Y\|_\infty) > 0$ thus holds iff

$$\lambda(h = \|h\|_\infty) > 0,$$

where λ is the Lebesgue measure on $[0, 1]$: this particularly holds for the AV@R$_\alpha$ distortion function $h_a = \frac{1}{1-\alpha}\mathbb{1}_{[\alpha,1]}$, as

$$\lambda(h_\alpha = \|h_\alpha\|_\infty) = \alpha > 0,$$

and the optimal measure can be given in this situation for any $r \geq 1$ as indicated.

We shall now further discuss the properties in the degenerate case where $\lambda(h = \|h\|_\infty) = 0$, in particular where h is

(i) unbounded (cf. Figure 1, left graph), and
(ii) bounded, but not flat at its top (Figure 1, right graph).

It will turn out that the first problem is pretty easy, whereas the second involves tough mathematical results.

5.1. Unbounded Spectral Densities

Theorem 17. *Suppose that h is not bounded and Y is linear on a linear space. Then the problem*

$$\begin{array}{rl} \text{maximize} & \\ \text{(in } P'\text{)} & \rho_{h;P'}(Y) \\ \text{subject to} & \mathsf{d}_{KA}(P, P') \leq K \end{array}$$

is not bounded either, i.e. the objective is ∞.

Proof. Consider the measures

$$P'_n := (T_n)_*(P),$$

for the transport plans

$$T_n(x) := x + K \cdot \frac{\mathbb{1}_{\{|Z_Y| \geq n\}}(x)}{P(|Z_Y| \geq n)} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y$$

which cut the (possibly sub-optimal) dual variable $Z_Y = h(U)$.

As above $\mathsf{d}_{KA}(P'_n, P) = K$, but

$$\rho_{h,P'_n}(Y) - \rho_{h,P}(Y) \geq K \cdot L(Y) \cdot \int |Z_Y| \cdot \frac{\mathbb{1}_{\{|Z_Y| \geq n\}}(x)}{P(|Z_Y| \geq n)} P(\mathrm{d}x) \geq K \cdot L(Y) \cdot n$$

for any $n \in \mathbb{N}$, the problem thus does not allow a bounded (real) solution. □

5.2. Bounded Spectrum Densities

Theorem 18. *Let Y be a (continuous) linear functional on $\left(\mathbb{R}^d, \|.\|\right)$. Moreover, assume that h is bounded, but $\lambda\left(|h| = \|h\|_\infty\right) = 0$ (cf. Figure 1). Then the problem*

$$\begin{array}{rl} \textit{maximize} & \\ (\textit{in } P') & \rho_{h;P'}(Y) \\ \textit{subject to} & \mathsf{d}_{KA}(P', P) \leq K \end{array} \tag{17}$$

is bounded, but there does not *exist a measure P' with $\mathsf{d}_{KA}(P, P') \leq K$ attaining the maximum in (17), that is to say the respective* argmax*-set is empty.*

Remark 19. Notice, that the latter statement holds true on finite-dimensional $\left(\mathbb{R}^d, \|.\|\right)$, so there is no chance on infinite dimensional spaces either to find a minimizing measure.

Proof. Define the set

$$C := \operatorname{argmax}\left\{\rho_{h;P'}(Y) : \mathsf{d}_{KA}(P, P') \leq K\right\},$$

which is the argmax-set, consisting of all measures maximizing problem (17) in consideration.

In order to prove the statement by contradiction suppose that C were not empty. As the optimal value is known precisely the minimum value of the problem may be written as

$$C = \left\{P' : \rho_{h;P'}(Y) - \rho_{h;P}(Y) = K \cdot L(Y) \cdot \|h\|_\infty, \, \mathsf{d}_{KA}(P, P') \leq K\right\}.$$

Furthermore, one may write

$$C = \bigcap_{n>1} C_n,$$

where the sets C_n originate from the relaxed problem

$$C_n = \left\{P' : \rho_{h;P'}(Y) - \rho_{h;P}(Y) \geq K \cdot L(Y) \cdot \left(\|h\|_\infty - \frac{1}{n}\right), \, \mathsf{d}_{KA}(P, P') \leq K\right\};$$

those sets C_n are certainly non-empty.

Consider the measures

$$P'_n := (T_n)_*(P),$$

defined via the transport maps

$$T_n(x) := x + K \cdot \frac{\mathbb{1}_{\left\{|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right\}}(x)}{P\left(|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right)} \cdot \operatorname{sign} Z_Y(x) \cdot x_Y$$

by appropriately cutting the dual variable Z_Y at its top.

By the same computation as above they satisfy $\mathsf{d}_{KA}(P'_n, P) = K$ by construction, and

$$\begin{aligned} \rho_{h,P'_n}(Y) - \rho_{h,P}(Y) &\geq K \cdot L(Y) \cdot \int |Z_Y| \cdot \frac{\mathbb{1}_{\left\{|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right\}}(x)}{P\left(|Z_Y| > \|Z_Y\|_\infty - \frac{1}{n}\right)} P(\mathrm{d}x) \\ &\geq K \cdot L(Y) \cdot \left(\|Z_Y\|_\infty - \frac{1}{n}\right), \end{aligned}$$

and thus $P'_n \in C_n$.

As $\left(\mathbb{R}^d, \|.\|\right)$ is locally compact, the space of continuous functions vanishing at infinity, $C_0\left(\mathbb{R}^d, \|.\|\right)$, is a Banach space, and Riesz' theorem identifies its dual with the space of regular Borel measures (cf. [35]).

The probability measures P'_n may be considered themselves as elements of this dual via the natural setting

$$\begin{aligned} P'_n : C_0\left(\mathbb{R}^d\right) &\to \mathbb{R} \\ \varphi &\mapsto \int \varphi \, \mathrm{d}P'_n, \end{aligned}$$

but moreover

$$|P'_n(\varphi)| \le \int \|\varphi\|_\infty \, \mathrm{d}P'_n = \|\varphi\|_\infty$$

for any function $\varphi \in C_0\left(\mathbb{R}^d, \|.\|\right)$, and thus $\|P'_n\| \le 1$: That is to say all those measures P'_n are within the unit ball $B_1(0)$ of the dual of $C_0\left(\mathbb{R}^d, \|.\|\right)$.

Alaoglu's theorem states that the closed unit ball $B_1(0)$ in the dual is weakly* compact, thus there is an accumulation point $\tilde{P}' \in B_1(0)$ such that

$$P'_{n_k} \to \tilde{P}'$$

in the weak* topology for some sub-sequence $(n_k)_k$. Again by Riesz' theorem $\tilde{P}'$ has a representation as a measure, although not necessarily as a probability measure.

We shall prove next that C is convex. This holds true, because

(i) the distance d_{KA} is convex for the situation $r = 1$, that is [8]

$$\mathsf{d}_{KA}(P, (1-\lambda) P_0 + \lambda P_1) \le (1-\lambda)\, \mathsf{d}_{KA}(P, P_0) + \lambda \mathsf{d}_{KA}(P, P_1). \tag{18}$$

Indeed, let π_0 (π_1, resp.) have marginals P and P_0 (P and P_1, resp.), then $\pi_\lambda := (1-\lambda)\pi_0 + \lambda \pi_1$ has marginals P and $P_\lambda := (1-\lambda) P_0 + \lambda P_1$, so that (18) is immediate.

(ii) $P \mapsto \rho_{h,P}(Y)$ is convex: to see this consider $P_0 \in C$, $P_1 \in C$ and observe that the distribution functions

$$\begin{aligned} G_{Y,\lambda}(z) &:= P_\lambda(Y \le z) = (1-\lambda) P_0(Y \le z) + \lambda P_1(Y \le z) \\ &= (1-\lambda) G_{Y,0}(z) + \lambda G_{Y,1}(z) \end{aligned}$$

are convex-combinations. Hence

$$\begin{aligned} \rho_{h,P_\lambda}(Y) &= \int G_{Y,\lambda}(z)\, h(z)\, \mathrm{d}z = \int \left((1-\lambda) G_{Y,0}(z) + \lambda G_{Y,1}(z)\right) h(z)\, \mathrm{d}z \\ &= (1-\lambda) \int G_{Y,0}(z)\, h(z)\, \mathrm{d}z + \lambda \int G_{Y,1}(z)\, h(z)\, \mathrm{d}z \\ &= (1-\lambda)\, \rho_{h,P_0}(Y) + \lambda \rho_{h,P_1}(Y) \end{aligned}$$

is a convex combination as well.

So C is convex. By Mazur's theorem the norm-closure and its weak* closure coincide for convex sets,

$$\tilde{P}' \in \overline{C}^{\text{weak}*} = \overline{C}^{\|\cdot\|},$$

we thus deduce in particular that

$$\left\|\tilde{P}'\right\| = 1,$$

and the limiting measure $\tilde{P}'$ thus is a probability measure.

Now define the increasing sets $X_n := \left\{|Z_Y| \le \|Z_Y\|_\infty - \frac{1}{n}\right\} \subset X$. Observe that

$$P'_n\left(\bigcup_n X_n\right) \ge P'_n(X_n) \ge \mathbb{P}\left(|Z_Y| \le \|Z_Y\|_\infty - \frac{1}{n}\right) = \lambda\left(h \le \|h\|_\infty - \frac{1}{n}\right) \longrightarrow 1$$

due to our assumptions, and particularly because $Z_Y = h(U)$. Hence $\tilde{P}'_n\left(\bigcup_n X_n\right) = 1$, and consequently $\tilde{P}'\left(\bigcup_n X_n\right) = 1$, because

$$P'_{n_k} \to \tilde{P}'$$

[8] (18) does not hold whenever $r > 1$.

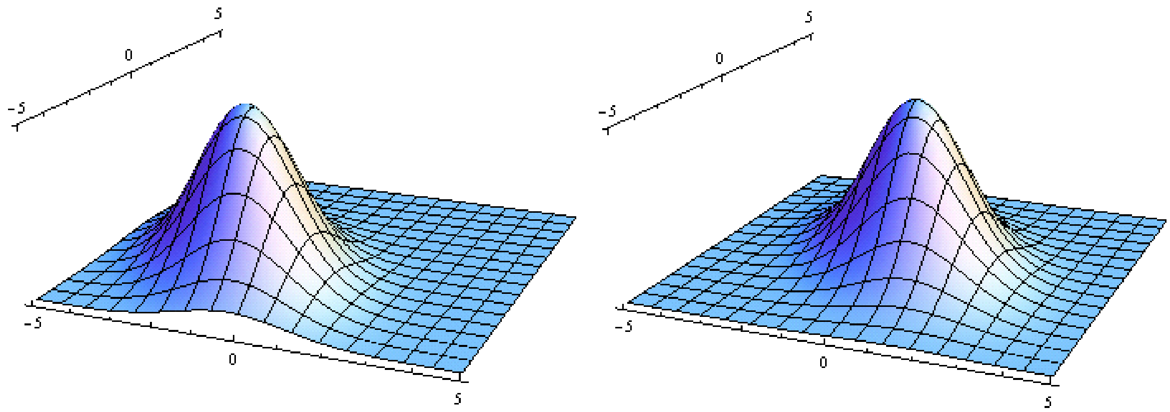

Figure 2: For the expectation, the worst measure is a simple translate. Here, $Y(x_1, x_2) = x_1 + x_2$ and whence the direction of the translation is $x_Y = \frac{1}{\sqrt{2}}(1,1)$ for the Euclidean distance.

by Portmanteau's Lemma (cf. [30]). By construction (recall the definition of the transport map T_n), P'_n and P coincide on any X_n, so $\tilde{P}'$ and P coincide on every set $A \subseteq \bigcup_n X_n$. This, however, means that $\tilde{P}' = P$, because

$$\tilde{P}'\left(\bigcup_n X_n\right) = P\left(\bigcup_n X_n\right) = 1.$$

This is a contradiction, because the measure P certainly is not optimal for the problem (17).

Hence C is the empty set,

$$C = \emptyset,$$

and there is no optimal measure P' for the problem (17). □

6. Illustration Of Optimal Transport Maps

6.1. Expectation

The expectation is the simplest risk measure, $\rho := \mathbb{E} = \mathrm{AV@R}_0$. By Theorem 15 the pushforward measure $P' := P^T$, where T is the translation $T(x) := x + K \cdot x_Y$, satisfies

$$\rho_{P'}(Y) = \rho_P(Y) - K \cdot L(Y)$$

for all $1 \leq r < \infty$. This is illustrated in Figure 2.

6.2. Distortion Measures

For spectral measures we have elaborated that Z_Y is coupled in an comonotone way with Y and moreover $Z_Y = h(U)$: We thus can give the dual variable as $Z_Y = h(G_Y(Y))$[9], and the transport map can be written as

$$T(x) = x + K \cdot \left|\frac{h(G_Y(Y(x)))}{\|h\|_{\frac{r}{r-1}}}\right|^{\frac{1}{r-1}} \cdot x_Y$$

for non-negative spectral densities h. This enables us to illustrate the geometry by plotting some densities, which we want to do here in providing some examples: two distortions of the same bivariate

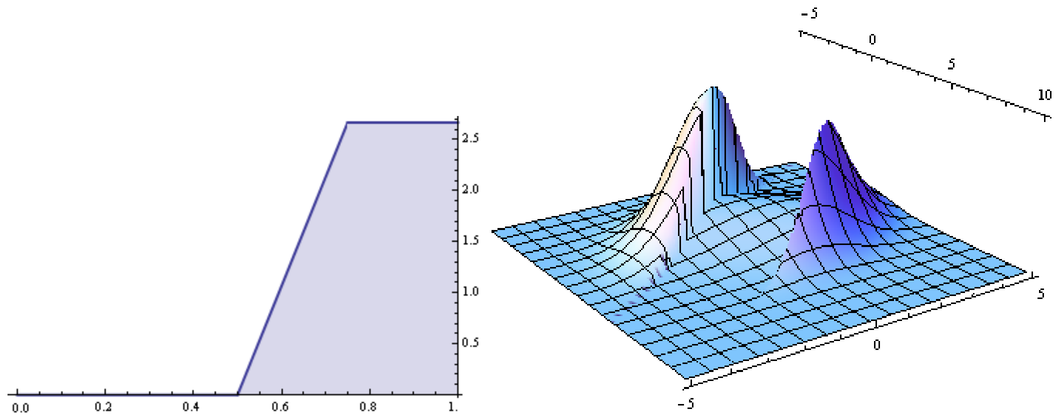

Figure 3: Spectral density h (left chart) and distorted probability measure (right chart) of the same measure as in Figure 2: 50 % of the mass stays at the same place, 25 % is simply being shifted in direction x_Y, and the remaining 25 % are brutally distorted in between (Y as in the previous example).

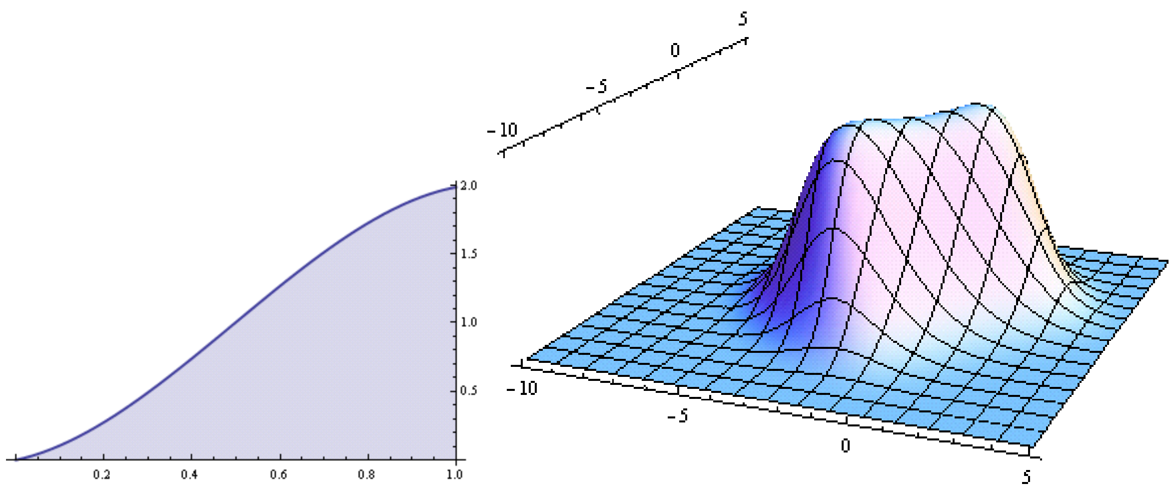

Figure 4: Resulting probability distribution (right chart) by applying the spectral density indicated (left chart) to the initial distribution from Figure 2.

distribution as in Figure 2 are depicted in Figure 3 and Figure 4, each with its corresponding spectral density function h.[10]

[9] Recall that h is *in*creasing.

[10] Let P denote a bivariate normal probability measure with mean μ and covariance matrix Σ ($P \sim \mathcal{N}(\mu, \Sigma)$) and Y a linear functional of the form $Y(x) = Y^\top x = \sum_i Y_i x_i$, then $Y_* P \sim \mathcal{N}\left(Y^\top \mu, Y^\top \Sigma Y\right)$, that is $Y_*(P) \sim \mathcal{N}\left(\sum_i Y_i \mu_i, \sum_{i,j} Y_i \Sigma_{i,j} Y_j\right)$; whence, $G_Y(y) = \frac{1}{\sqrt{2\pi Y^\top \Sigma Y}} \int_{-\infty}^{y} e^{-\frac{1}{2}\frac{\left(x - Y^\top \mu\right)^2}{Y^\top \Sigma Y}} \mathrm{d}x$ and $G_Y(Y(x)) = \frac{1}{\sqrt{2\pi Y^\top \Sigma Y}} \int_{-\infty}^{Y^\top x} e^{-\frac{1}{2}\frac{\left(x' - Y^\top \mu\right)^2}{Y^\top \Sigma Y}} \mathrm{d}x'$ is a $\mathbb{R}-$valued random variable.

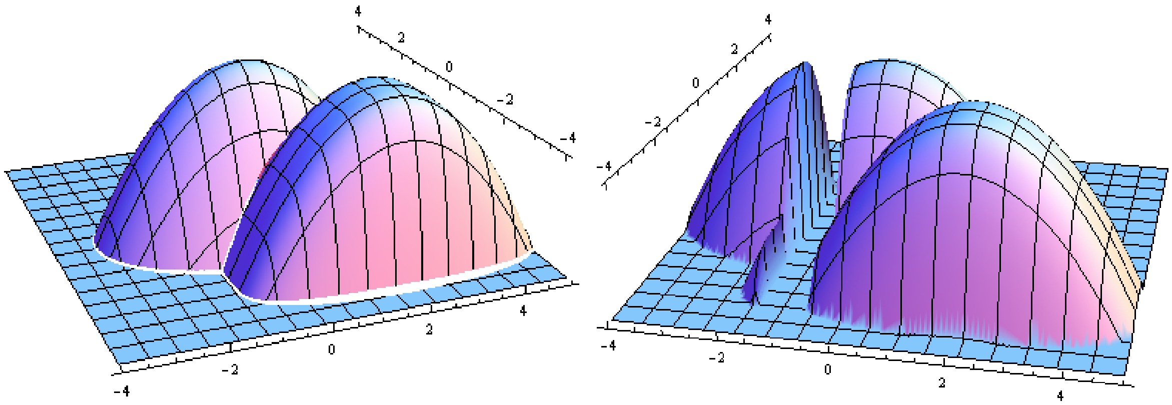

Figure 5: Initial (left) and split (right) probability measure, as it is worst with respect to the Average Value-at-Risk. Displayed from different perspectives.

6.3. The Average Value-at-Risk

As for the Average Value-at-Risk at level α the optimal dual variable is basically $Z_Y = \mathbb{1}_{\left\{Y<G_Y^{-1}(\alpha)\right\}}$. The transport map, for all $1 \leq r < \infty$, is

$$T(x) = x + \frac{K}{\alpha} \cdot x_Y \cdot \mathbb{1}_{\left\{Y>G_Y^{-1}(\alpha)\right\}}(x),$$

which again includes the expectation for $\alpha = 1$.

This transport map splits the sample space X according the α-quantile: Those samples, which *do not* contribute to the computation of $\mathrm{AV@R}_\alpha$ (which have quantile $\left(\left\{Y < G_Y^{-1}(\alpha)\right\}\right)$, are left unchanged in their place, while all other samples, which *do* contribute to the $\mathrm{AV@R}_\alpha$ $\left(\left\{Y > G_Y^{-1}(\alpha)\right\}\right)$, are being simply worsened by shifting them the distance $\frac{K}{\alpha}$; moreover, all of them are being shifted

- in *parallel*
- in the *same direction* x_Y and
- the *same distance* $\frac{K}{\alpha}$,

as illustrated in Figure 5.

7. Summary

In the present paper we investigate the impact of probability measures on the evaluation of risk functionals. It is demonstrated that the Wasserstein distance is a very useful concept of distance in this context. Precise and useful bounds can be given when evaluating the risk functionals.

The measures within a given Wasserstein-ball of radius K, which have the highest impact on the evaluation of a risk functional, can be described in an intuitive way, in particular for the Average Value-at-Risk. For linear random variables it is a distorted measure, which can be constructed by employing a translation. The Average Value-at-Risk is a special case, the respective worst measure divides the sample space into two parts: one part is left unchanged, whereas the second part is being shifted uniformly away from the first one.

The precise understanding of the respective worst measure is an important achievement for many areas which employ risk measures (or acceptability functionals). It is the basis for specific applications in insurance, finance (cf. [18]), energy and other areas.

The results are particularly important for optimization under uncertainty or robust optimization, as they can be used to quantify the error, which is intrinsic in empirical approximations.

Acknowledgment

The author is very grateful to Georg Ch. Pflug for his continual advise for this paper and its basis, which is partially contained in [21].

The paper has benefited from careful readings by two anonymous referees and Vladimir Norkin (associated editor).

The plots have been produced by use of Mathematica.

8. Bibliography

[1] C. Acerbi and P. Simonetti. Portfolio optimization with spectral measures of risk. *EconPapers*, 2002. 3

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9:203–228, 1999. 2

[3] A. Balbás, J. Garrido, and S. Mayoral. Properties of distortion risk measures. *Methodology and Computing in Applied Probability*, 11:385–399, 2009. 3

[4] A. S. Cherny. Weighted V@R and its properties. *Finance and Stochastics*, 10:367–393, 2006. 3

[5] F. Delbaen. Coherent risk measures on general probability spaces. In *Essays in Honour of Dieter Sondermann*, pages 1–37. Springer-Verlag, Berlin, 2002. 2

[6] T. S. Ferguson. *Mathematical Statistics. A Decision Theoretic Approach.* Academic Press, New York, 1967. 7

[7] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time.* de Gruyter Studies in Mathematics 27. de Gruyter, 2004. 2, 4

[8] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002. 6

[9] W. Hoeffding. Maßstabinvariante Korrelationstheorie. *Schriften Math. Inst. Univ. Berlin*, 5:181–233, 1940. In German. 7

[10] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics*, 9:49–71, 2006. 3

[11] L. Kantorovich. On the translocation of masses. *C.R. Acad. Sci. URSS*, 37:199–201, 1942. 6

[12] S. Kusuoka. On law invariant coherent risk measures. *Advances in Mathematical Economics*, 3:83–95, 2001. 3

[13] G. Monge. Mémoire sue la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704, 1781. 6

[14] R. B. Nelsen. *An Introduction to Copulas.* Number 139 in Lecture Notes in Statistics. Springer, New York, 1998. 7

[15] G. Ch. Pflug. On distortion functionals. *Statistics and Risk Modeling (formerly: Statistics and Decisions)*, 24:45–60, 2006. 3, 8

[16] G. Ch. Pflug and A. Pichler. *Approximations for Probability Distributions and Stochastic Optimization Problems*, volume 163 of *International Series in Operations Research & Management Science*, chapter 15, pages 343–387. Springer, New York, 2011. 6

[17] G. Ch. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012. 10

[18] G. Ch. Pflug, A. Pichler, and D. Wozabal. The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36:410–417, 2012. 19

[19] G. Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk.* World Scientific, River Edge, NJ, 2007. 2, 8

[20] G. Ch. Pflug and D. Wozabal. *Quantitative Fund Management.* Financial Mathematics Series. Chapman & Hall/ CRC, 2009. 8

[21] A. Pichler. *Distance Of Probability Measures And Respective Continuity Properties Of Acceptability Functionals.* PhD thesis, University of Vienna, Vienna, Austria, 2010. 20

[22] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems Vol. I: Theory, Vol. II: Applications*, volume XXV of *Probability and its applications.* Springer, New York, 1998. 6

[23] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970. 5

[24] R. T. Rockafellar and S. Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–41, 2000. 2

[25] R. T. Rockafellar, S. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74, 2006. 2

[26] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31:433–452, 2006. 5, 7

[27] A. Shapiro. On Kusuoka representations of law invariant risk measures. *Mathematics of Operations Research*, November 2012. 7

[28] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming.* MQS-SIAM Series on Optimization 9, 2009. 2, 4, 5

[29] H. Tsukahara. One-parameter families of distortion risk measures. *Mathematical Finance*, 19:691–705, 2009. 3

[30] A. W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1998. 17

[31] A. M. Vershik. Kantorovich metric: Initial history and little-known applications. *Journal of Mathematical Sciences*, 133(4):1410–1417, 2006. 1

[32] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003. 6

[33] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften.* Springer, Berlin, 2009. 6

[34] D. Williams. *Probability with Martingales.* Cambridge University Press, Cambridge, 1991. 12

[35] P. Wojtaszczyk. *Banach Spaces for Analysts.* Cambridge University Press, Cambridge, UK, 1991. 15