Depend. Model. 2016; 4:348–367 DE GRUYTER OPEN

Research Article Open Access

Special Issue: Recent Developments in Quantitative Risk Management

Silvana M. Pesenti*, Pietro Millossovich, and Andreas Tsanakas

Robustness regions for measures of risk aggregation

DOI 10.1515/demo-2016-0020
Received September 30, 2016; accepted November 18, 2016

Abstract: One of risk measures' key purposes is to consistently rank and distinguish between different risk profiles. From a practical perspective, a risk measure should also be robust, that is, insensitive to small perturbations in input assumptions. It is known in the literature [14, 39], that strong assumptions on the risk measure's ability to distinguish between risks may lead to a lack of robustness. We address the trade-off between robustness and consistent risk ranking by specifying the regions in the space of distribution functions, where law-invariant convex risk measures are indeed robust. Examples include the set of random variables with bounded second moment and those that are less volatile (in convex order) than random variables in a given uniformly integrable set. Typically, a risk measure is evaluated on the output of an aggregation function defined on a set of random input vectors. Extending the definition of robustness to this setting, we find that law-invariant convex risk measures are robust for any aggregation function that satisfies a linear growth condition in the tail, provided that the set of possible marginals is uniformly integrable. Thus, we obtain that all law-invariant convex risk measures possess the aggregation-robustness property introduced by [26] and further studied by [40]. This is in contrast to the widely-used, non-convex, risk measure Value-at-Risk, whose robustness in a risk aggregation context requires restricting the possible dependence structures of the input vectors.

Keywords: Convex risk measures, Aggregation, Value-at-Risk, Robustness, Continuity

MSC: 62G35, 62H99, 62P05, 91B30, 91G99

1 Introduction

Since the wide-spread adoption of Value-at-Risk (VaR) frameworks in the 1990s, risk measures have constituted an integral part of financial risk management. The use of risk measures is prescribed by banking [6, 7] and insurance regulation [23] for calculating the capital requirements of portfolios of future losses. Furthermore, the use of risk measures, evaluated using internally developed statistical models, is increasingly embedded in the operations of insurance companies [45, 46].

As a consequence, the discussion of desirable properties of risk measures has been the focus of much academic and industry debate. A first set of considerations relates to risk measures' ability to reflect diversification appropriately, by the properties of subadditivity [4] and convexity [28, 30], and to order risk consistently [5, 18]. These issues are interrelated: law-invariant convex risk measures, introduced by [28, 30] and subsuming coherent risk measures [4], rank risks in a way that preserves first-order and second-order stochas-

***Corresponding Author: Silvana M. Pesenti:** Cass Business School, City, University of London, E-mail: Silvana.Pesenti@cass.city.ac.uk
Pietro Millossovich: Cass Business School, City, University of London, E-mail: Pietro.Millossovich.1@city.ac.uk
Andreas Tsanakas: Cass Business School, City, University of London, E-mail: A.Tsanakas.1@city.ac.uk

tic dominance [5]. The risk measure Expected Shortfall (ES) is the convex risk measure used most widely in the practice of risk management.

A second set of considerations acknowledges that risk measures need to be estimated from historical and/or simulated data and thus require reliable estimators. A fundamental concept is the question of robustness, that is, whether risk measure estimates remain relatively insensitive to small perturbations in the underlying distribution from which data are generated [33, 34]. A growing academic literature is concerned with robustness in the context of risk measurement [8, 14, 26, 38–40]. A key finding of this literature is that robustness is to an extent contradictory to a consistent ordering of risks. In particular, there does not exist a law-invariant convex risk measure that is robust (following the definition of [33, 34]) on the whole space of integrable random variables. This fact has been used as an argument against the use of convex risk measures such as ES and in favour of the non-convex risk measure VaR [14]. Such arguments have coloured much of the policy discussion surrounding the relative merits of ES and VaR for use in capital regulation [6, 7, 35].

One way to address the apparent conflict between consistency of risk ranking and robustness, is to consider alternative, less restrictive, definitions of robustness [38, 39]. Another approach also taken in [40], which we follow in this paper, is to relax the requirement that risk measures be robust on the whole space of integrable random variables, given that "... this case is not generally interesting in econometric or financial applications since requiring robustness against all perturbations of the model is quite restrictive..." [14]. This approach suggests an analysis of regions on which risk measures are robust. Consequently, since in different applications different regions of distributions may form plausible input spaces, selection of a risk measure for a particular application should reflect the extent to which the risk measure is robust on the region of interest.

In this paper, we study robustness regions for convex risk measures and show that they are characterised by the property of uniform integrability – through examples we demonstrate that this is not an excessively strong requirement on the input space. Furthermore, we consider the realistic case where risk measures are evaluated on (possibly non-linear) functions of random vectors of risk factors, such that the input space consists of multivariate distributions [46, 50]. This case, typical in the risk modelling performed by insurance companies, is generally not considered in the literature on robustness, with the exception of [26, 40] who focus on fixed marginals. However, robustness as defined in [33, 34], that is, insensitivity to small deviations from the underlying distribution, includes both perturbation in the marginals and the dependence structure of the random vector of input risk factors. Allowing for uncertainty in the marginal distributions, we show that weak restrictions on the marginals (uniform integrability) and the aggregation function (linear growth in the tail) ensure robustness of convex risk measures. Consequently, we argue that in applications where risk aggregation takes place and uncertainty around the dependence structure is high, convex risk measures such as ES have attractive robustness properties, compared to, say, VaR.

In Section 2 notation and mathematical preliminaries are stated. In Section 3, the robustness of convex risk measures is studied. First, in Section 3.1, robustness is formally defined and its relationship to continuity of risk measures (Hampel's theorem) is presented. A key result for the rest of the paper (that also follows from [40]) is then shown: convex risk measures are robust on uniformly integrable sets. Subsequently, in Section 3.2, examples of such uniformly integrable sets are given. Uniform integrability is a constraint on the tail behaviour of a set of distributions. Thus convex risk measures are robust on sets including parametric families with bounded second moment; sets of random variables that are less volatile (in convex order) than those in a given uniformly integrable set. Section 3.3 presents examples of sets on which convex risk measures are not robust and Section 3.4 points at possible extensions to risk measures defined on the set of random variables with finite p-th moment.

In Section 4, robustness is studied in the context of risk aggregation, where a risk measure is applied on real-valued *aggregation function* of a random vector of risk factors; we call the composition of the risk measure with the aggregation function an *aggregation measure*. In Section 4.1, robustness of aggregation measures is defined with respect to distributions of random vectors. A direct multivariate extension of Hampel's theorem is given, associating robustness with continuity of the aggregation measure. Consequently, if the risk measure is convex and the aggregation function continuous, the aggregation measure is robust as long as the aggregate risk position belongs to a uniformly integrable set. In Section 4.2 we show that for robustness of aggregation measures it is sufficient that the marginals of the vector of risk factors belong to uniformly integrable sets and that the aggregation function possesses a linear growth condition in the tail. Significantly, no constraints on the dependence structure of risk factors are placed. This includes, as a special case, aggregation via the ordinary sum and thus generalises the results on *aggregation robustness* in [26] to the class of law-invariant convex risk measures and the results in [40] to uncertainty in the marginal distributions. In Section 4.3 it is shown that robustness is also satisfied for aggregation via compound distributions, a typical setting in actuarial science, as long as the frequency and severity distributions are dominated (in first-order stochastic dominance) by integrable random variables.

Finally, in Section 5, a comparison with the robustness regions of the (non-convex) VaR measure is made. VaR is robust as long as the distribution function is strictly increasing. We argue that in applications, this can be a stronger requirement than the uniform integrability that is required when convex risk measures are used. Non-linear aggregation functions, such as the ones arising in the context of reinsurance, can lead to constant parts of the aggregate distribution function and thus to non-robustness. Furthermore, it is known from the literature on dependence uncertainty that dependence structures can be designed such that the distribution of the sum is not strictly monotonic in the tail, when the marginal distributions satisfy particular ('mixability') conditions [9, 25, 53–55]. Thus, robustness of VaR requires restrictions both in the aggregation function and the dependence structure. In applications such as the internal capital modelling performed by insures, we believe that such constraints are unrealistic, compared to those applying to convex risk measures. Thus our paper indicates that in applications where non-linear aggregations and high dependence uncertainty are present, convex risk measures such as ES, may be preferable to VaR.

2 Preliminaries

Throughout the paper, we consider an atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We denote the space of real-valued random variables by $\mathcal{L}^0 = \mathcal{L}^0(\Omega, \mathcal{A}, \mathbb{P})$, the subspace of integrable random variables by $\mathcal{L}^1 = \{X \in \mathcal{L}^0 \,|\, \|X\|_1 = \mathbb{E}(|X|) < +\infty\}$ and the subset of (essentially) bounded random variables by $\mathcal{L}^\infty$. For $\mathcal{X} \subset \mathcal{L}^0$ we define the corresponding set of distribution functions by $\mathfrak{D}(\mathcal{X}) = \{\mathbb{P} \circ X^{-1} \,|\, X \in \mathcal{X}\}$. We denote by $F_X(\cdot) = \mathbb{P}(X \le \cdot)$ the distribution function of X and write $X \sim F_X$, so that $\mathfrak{D}(\mathcal{X}) = \{F_X \,|\, X \in \mathcal{X}\}$. Note that we identify distribution functions on $\mathbb{R}$ with the corresponding probabilities on the Borel σ-field $\mathcal{B}(\mathbb{R})$. We write $\mathfrak{M} = \mathfrak{D}(\mathcal{L}^0)$ for the set of all distribution functions on $\mathbb{R}$, and $\mathfrak{M}^1 = \{F \in \mathfrak{M} \,|\, \int_{\mathbb{R}} |x| \mathrm{d}F(x) < +\infty\} = \mathfrak{D}(\mathcal{L}^1)$.

On the space $\mathfrak{M}$ we consider the Prokhorov distance defined for $F, G \in \mathfrak{M}$ through

$$d_P(F, G) = \inf\{\varepsilon > 0 \,|\, F(B) \le G(B^\varepsilon) + \varepsilon, \text{ for all Borel sets } B \text{ on } \mathbb{R}\},$$

where $B^\varepsilon = \{x \in \mathbb{R} \,|\, \inf_{y \in B} |x - y| \le \varepsilon\}$.

The following definition is of central importance throughout the paper. A set of distribution functions $\mathfrak{U} \subset \mathfrak{M}^1$ is *uniformly integrable* if

$$\lim_{K \to +\infty} \sup_{F \in \mathfrak{U}} \int\limits_{|x|>K} |x| \mathrm{d}F(x) = 0.$$

We say a set of random variables $\mathcal{U} \subset \mathcal{L}^1$ is uniformly integrable if $\mathfrak{D}(\mathcal{U})$ is uniformly integrable, equivalently

$$\lim_{K \to +\infty} \sup_{X \in \mathcal{U}} \mathbb{E}\left(|X| \mathbb{1}_{\{|X|>K\}}\right) = 0.$$

Uniform integrability of a set posits that the contribution of the distributions' far tails can be uniformly controlled across the elements of the set. Thus, it is a stronger condition than requiring that all elements of a set are integrable.

A risk measure $\rho : \mathcal{L}^1 \to \mathbb{R}$ is a function that associates to every integrable random variable a real number. The argument of ρ is assumed throughout to represent a financial loss. Possible properties of a risk measure are:

i) Law-invariance: $\rho(X) = \rho(Y)$ for $X, Y \in \mathcal{L}^1$ with $F_X = F_Y$.
ii) Translation invariance: $\rho(X + m) = \rho(X) + m$ for $X \in \mathcal{L}^1$, $m \in \mathbb{R}$.
iii) Monotonicity: $\rho(X) \leq \rho(Y)$ for $X, Y \in \mathcal{L}^1$ with $X \leq Y$, $\mathbb{P}$-a.s.
iv) Convexity: $\rho\big((1-\lambda)X + \lambda Y\big) \leq (1-\lambda)\rho(X) + \lambda\rho(Y)$ for $X, Y \in \mathcal{L}^1$, $\lambda \in [0, 1]$.

A convex risk measure is a risk measure fulfilling ii), iii) and iv), see [29, 30] and references therein. A law-invariant risk measure $\rho(\cdot)\colon \mathcal{L}^1 \to \mathbb{R}$ induces a functional on the corresponding set of distribution functions, $\rho[\cdot]\colon \mathfrak{M}^1 \to \mathbb{R}$, through $\rho[F_X] = \rho(X)$ for $F_X \in \mathfrak{M}^1$. For instance, we write $\mathbb{E}(X) = \mathbb{E}[F_X]$. (Throughout the paper, we denote law invariant functionals using round brackets $(\cdot)$ when the argument is a random variable, and square brackets $[\cdot]$ when the argument is a distribution.) We say a risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is continuous on $\mathcal{X} \subset \mathcal{L}^1$ with respect to the Prokhorov distance if the restriction of the induced functional $\rho[\cdot]$ to $\mathfrak{D}(\mathcal{X})$ is continuous with respect to d_P. That is, for all $F_0 \in \mathfrak{D}(\mathcal{X})$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all $F \in \mathfrak{D}(\mathcal{X})$ we have $d_P(F_0, F) < \delta$ implies $|\rho[F_0] - \rho[F]| < \varepsilon$. The property of law invariance is standard in risk management applications, requiring that risk assessments only depend on the distribution of random losses. Therefore all risk measures in this paper are tacitly assumed to be law-invariant without this being explicitly stated in the sequel.

Remark 2.1. A substantial part of the early literature considers risk measures, axiomatically introduced in [3, 4], defined on $\mathcal{L}^\infty$; however, insurance and financial portfolios are primarily exposed to unbounded risks. Therefore we choose $\mathcal{L}^1$ as our model space. In fact, the natural model space for law-invariant convex risk measures is $\mathcal{L}^1$, since outside this space the risk measure can only take value $+\infty$ [27, 47]. Selected literature on risk measures defined on a broader space than $\mathcal{L}^\infty$ are [16] for general probability spaces, [37, 47] on sets of random variables with finite p-th moment, [13, 32, 39] on Orlicz spaces and [27] for extensions of risk measures from $\mathcal{L}^\infty$ to $\mathcal{L}^1$.

An example of a convex risk measure that is finite on $\mathcal{L}^1$ is Expected Shortfall (ES) at level $\alpha \in [0, 1)$, defined by

$$\mathrm{ES}_\alpha[F] = \frac{1}{1-\alpha}\int_\alpha^1 F^{-1}(u)\mathrm{d}u.$$

Expected Shortfall belongs to the class of spectral risk measures, introduced in [1, 56],

$$\rho(X) = \int_0^1 F_X^{-1}(u)\phi(u)\mathrm{d}u, \quad \text{for } X \in \mathcal{L}^1,$$

where $F_X^{-1}(u) = \inf\{y \in \mathbb{R} \mid F_X(y) \geq u\}$, $u \in (0, 1)$, is the generalised inverse and we identify $\inf\emptyset = -\infty$. The weight function $\phi\colon [0, 1] \to [0, +\infty)$ is non-decreasing and normalised, that is $\int_0^1 \phi(u)\mathrm{d}u = 1$. Spectral risk measures are generally not finite on $\mathcal{L}^1$. However, finiteness is guaranteed if the weight function ϕ is constant on $(1-\varepsilon, 1]$ for $\varepsilon > 0$, as is the case for Expected Shortfall, corresponding to $\phi(u) = \frac{1}{1-\alpha}\mathbb{1}_{\{u>\alpha\}}$.

3 Robustness

3.1 Robustness of convex risk measures

The classical definition of statistical robustness [33], considers estimators as functionals of empirical distribution functions. For a distribution function $F \in \mathfrak{M}^1$ and sample size $k \geq 1$ the empirical distribution function is defined by the random measure

$$\hat{F}_k(t, \omega) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{\{X_i(\omega) \leq t\}}, \quad (t, \omega) \in \mathbb{R} \times \Omega,$$

where $X_1, \dots, X_k \in \mathcal{L}^1$ are independent with common distribution function F. In the sequel we consider the sequence of estimators $\{\hat{\rho}_k\}_k$ of a risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ by evaluating the risk measure on the empirical distribution functions. That is, for $F \in \mathfrak{M}^1$ and $k \geq 1$, we define

$$\hat{\rho}_k[F](\omega) = \rho[\hat{F}_k(\cdot,\omega)],\ \omega \in \Omega. \tag{1}$$

Note that the estimator $\hat{\rho}_k[F]$ is a random variable. Ideally, the estimator $\{\hat{\rho}_k\}_k$ should be consistent and robust. The sequence of estimators is consistent if it converges to the true value, $\hat{\rho}_k[F] \to \rho[F]$ in probability. Robustness, according to Hampel [33, 34], is understood as insensitivity of estimators to small perturbations in the distribution F.

Definition 3.1. *([33])*
A risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is *robust* on $\mathcal{X} \subset \mathcal{L}^1$ (equivalently $\rho[\cdot]$ is robust on $\mathfrak{D}(\mathcal{X})$) if for any $F_0 \in \mathfrak{D}(\mathcal{X})$ the sequence of estimators $\{\hat{\rho}_k[F_0]\}_k$, as defined in (1), fulfils that for all $\varepsilon > 0$ there exists $\delta > 0$ and $k_0 \in \mathbb{N}$ such that, for all $F \in \mathfrak{D}(\mathcal{X})$ and $k \geq k_0$, we have

$$d_P(F_0, F) < \delta \quad \Rightarrow \quad d_P\Big(\mathfrak{D}\big(\hat{\rho}_k[F_0]\big), \mathfrak{D}\big(\hat{\rho}_k[F]\big)\Big) < \varepsilon.$$

By the celebrated theorem of Hampel [33], given consistency, robustness of a risk measure is equivalent to continuity with respect to the Prokhorov distance.

Theorem 3.2. *([33], Theorem 2.21 in [34])*
Let $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ be a risk measure and $\mathcal{X} \subset \mathcal{L}^1$. Assume that the sequence $\{\hat{\rho}_k\}_k$, as defined in (1), *is consistent for all $F_0 \in \mathfrak{D}(\mathcal{X})$. Then ρ is continuous on $\mathfrak{D}(\mathcal{X})$ with respect to the Prokhorov distance if and only if the risk measure is robust on $\mathfrak{D}(\mathcal{X})$.*

For convex risk measures we obtain a one-to-one correspondence between robustness and continuity, since they are consistent on $\mathfrak{M}^1$.

Proposition 3.3. Let $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ be a convex risk measure and $\mathcal{X} \subset \mathcal{L}^1$. Then, ρ is continuous with respect to the Prokhorov distance on $\mathfrak{D}(\mathcal{X})$ if and only if it is robust on $\mathfrak{D}(\mathcal{X})$.

Proof. We show strong consistency of convex risk measures, that is for $F_0 \in \mathfrak{M}^1$ we have $\hat{\rho}_k[F_0](\omega) \to \rho[F_0]$ for almost every $\omega \in \Omega$. Let $\{\hat{F_0}_k(\cdot,\omega)\}_k$, $\omega \in \Omega$, be the corresponding sequence of empirical distribution functions. By Glivenko-Cantelli $\{\hat{F_0}_k(\cdot,\omega)\}_k$ converges to $F_0(\cdot)$ for almost every $\omega \in \Omega$ in the Prokhorov distance. The strong law of large numbers implies that for $X_{0,i} \sim F_0,\ i = 1, \dots, k$ and almost every $\omega \in \Omega$

$$\int_{\mathbb{R}} |x| \mathrm{d}\hat{F_0}_k(x,\omega) = \frac{1}{k}\sum_{i=1}^{k} |X_{0,i}(\omega)| \quad \longrightarrow \quad \mathbb{E}(|X_0|) = \int_{\mathbb{R}} |x| \mathrm{d}F_0(x),\ \text{as } k \to +\infty.$$

Applying Lemma A.1 $\{\hat{F_0}_k(\cdot,\omega)\}_k$ converges to $F_0(\cdot)$ in the Wasserstein distance (see Appendix for the definition and properties of such distance) for almost every $\omega \in \Omega$. Since convex risk measures are continuous with respect to the Wasserstein distance, Theorem 2.8 in [39], $\hat{\rho}_k[F_0](\omega) = \rho[\hat{F_0}_k(\cdot,\omega)] \to \rho[F_0]$, as $k \to +\infty$, for almost every $\omega \in \Omega$. □

No convex risk measure is robust on the whole of $\mathcal{L}^1$, as shown in Lemma 3.4 below.

Lemma 3.4. There does not exist a convex risk measure that is robust on $\mathcal{L}^1$.

Proof. [5, 14, 39] show that there does not exist a convex risk measure that is continuous with respect to the Prokhorov distance on the whole space of integrable random variables. Applying Proposition 3.3 gives the claim. □

Given the importance of both convexity and robustness for risk management, the need emerges to study subsets of $\mathcal{L}^1$ on which convex risk measures become robust. Uniformly integrable sets are at the core of characterising robustness regions for convex risk measures.

Theorem 3.5. *A convex risk measure is robust on $\mathcal{X} \subset \mathcal{L}^1$ if the set $\mathcal{X}$ is uniformly integrable.*

Proof. Convex risk measures are continuous on $\mathfrak{M}^1$ with respect to the Wasserstein distance, Theorem 2.8 in [39]. On a uniformly integrable set the topology induced by the Wasserstein distance is equivalent to the topology induced by the Prokhorov distance, see Lemma A.1 or Theorem 2 in [20]. Hence, on $\mathcal{X}$ the risk measure is continuous with respect to the Prokhorov distance and we can apply Proposition 3.3. □

Alternatively, the proof of Theorem 3.5 follows from Theorem 2.6 in [40].

Remark 3.6. The general concept of robustness is based on continuity with respect to the weak topology on $\mathfrak{M}$ [34]. Due to its tractability, the Lévy distance is frequently used for defining robustness [14]. Since both the Prokhorov and the Lévy distance generate the weak topology on $\mathfrak{M}$, they give rise to the same notion of robustness [34]. We adopt the Prokhorov distance since it allows for a natural extension to multivariate distribution functions, see Section 4.

3.2 Robustness regions of convex risk measures

In this section, we provide some examples of classes of sets that are uniformly integrable and on which, by Theorem 3.5, convex risk measures are robust. It is seen throughout that uniform integrability puts a constraint on the tail behaviour of the risks considered.

First, we note that a convex risk measure is robust when evaluated on a set of empirical distribution functions.

Lemma 3.7. Let $F \in \mathfrak{M}^1$. A convex risk measure is robust on the sequence of empirical distribution functions $\left\{\hat{F}_k(\cdot, \omega) \mid k \geq 1\right\} \subset \mathfrak{M}^1$ for almost every $\omega \in \Omega$.

Proof. In the proof of Proposition 3.3 it was shown that the sequence $\hat{F}_k(\cdot, \omega)$ converges in the (Prokhorov and) Wasserstein distance to F for almost every $\omega \in \Omega$. By Lemma A.1 this implies that the sequence is, for almost every ω, uniformly integrable and we can apply Theorem 3.5. □

More generally, a convex risk measure is robust on sets of uniformly bounded random variables, that is $\{X \in \mathcal{L}^1 \mid |X| \leq M,\ \mathbb{P}\text{-a.s.}\}$ for $M > 0$, see [22, p. 220]. Instead of restricting the support of the random variables we could restrict their moments. A convex risk measure is robust on the set of distribution functions $\mathfrak{U} \subset \mathfrak{M}^1$ having uniformly bounded second moments or, more generally, satisfying [11, p. 218]

$$\sup_{F \in \mathfrak{U}} \int_{\mathbb{R}} |x|^{1+\varepsilon} \mathrm{d}F(x) < +\infty, \quad \text{for some } \varepsilon > 0.$$

Subsequently, a convex risk measure is robust on a family of parametric models, $\{F_\theta \mid \theta \in \Theta\}$, if the family fulfils $\int_{\mathbb{R}} |x|^2 \mathrm{d}F_\theta(x) < M$, for all $\theta \in \Theta$. For example, consider the exponential dispersion family, a parametric family of distribution functions with density

$$f(x; \theta, \phi) = \exp\left\{\frac{x\theta - b(\theta)}{\phi / w} + c(x, \phi, w)\right\}, \ x \in \mathbb{R}$$

with weight $w > 0$, dispersion parameter $\phi > 0$ and normalising function $c(\cdot, \cdot, \cdot)$. The canonical parameter of the exponential dispersion family is $\theta \in \Theta$, where $\Theta \subset \mathbb{R}$ and $b \colon \Theta \to \mathbb{R}$ is the cumulant function such that the density is well-defined and has identical support for all $\theta \in \Theta$, [43]. The exponential dispersion family includes the Poisson, Negative-Binomial, Gamma, Gaussian and Inverse Gaussian.

Lemma 3.8. A convex risk measure is robust on the exponential dispersion family if the parameter space Θ is compact and the function b twice continuously differentiable on Θ.

Proof. Let X follow a distribution that belongs to the exponential dispersion family. Then $\mathbb{E}(X) = b'(\theta)$ and $\mathrm{Var}(X) = \frac{\phi}{w} b''(\theta)$, [57]. Both the first and second derivative b', b'' are continuous and hence bounded on the compact set Θ. □

We refer to [40] for a broader discussion and examples involving parametric models such as the Normal, Pareto, Gamma and Gumbel distributions.

Now we consider the relationship between uniform integrability and stochastic orderings. A convex risk measure is robust on a set of non-negative random variables that are smaller (in first-order stochastic dominance) than those in a given uniformly integrable set.

Lemma 3.9. Let $\mathcal{U}$ be a uniformly integrable set of non-negative random variables. A convex risk measure is robust on the set

$$\mathcal{N} = \{Y \in \mathcal{L}^1 \,|\, Y \geq 0 \text{ and there exists } X \in \mathcal{U} \text{ such that } \mathbb{E}(f(Y)) \leq \mathbb{E}(f(X)) \text{ for all increasing } f\}.$$

Proof. For $K > 0$, the function $f(x) = x\mathbb{1}_{\{x>K\}}$ is increasing. Hence we have, by uniform integrability of $\mathcal{U}$,

$$\lim_{K\to+\infty} \sup_{Y\in\mathcal{N}} \mathbb{E}\big(Y\mathbb{1}_{\{Y>K\}}\big) \leq \lim_{K\to+\infty} \sup_{X\in\mathcal{U}} \mathbb{E}\big(X\mathbb{1}_{\{X>K\}}\big) = 0.$$

The conclusion follows by Theorem 3.5. □

An example of the application of Lemma 3.9 is the *Generalised Pareto Distribution* (GPD) denoted by $G_{\xi;\sigma}$, with shape and scale parameters, $\xi \in \mathbb{R}$ and $\sigma > 0$ respectively, defined through

$$G_{\xi;\sigma}(x) = \begin{cases} 1 - \left(1 + \xi\frac{x}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left\{-\frac{x}{\sigma}\right\} & \xi = 0, \end{cases}$$

where $x \geq 0$, if $\xi \geq 0$, and $0 \leq x \leq -\sigma/\xi$, if $\xi < 0$. The GPD is often used in insurance and operational risk management to model portfolios that can produce very large claims, since it is the limit distribution of conditional excesses over high thresholds [24]. The expectation of a GPD is finite if the shape parameter satisfies $\xi < 1$. For a set of GPDs to be uniformly integrable it is necessary that their shape parameters be bounded away from 1; see Proposition 3.14 for the necessity of this condition in the more general case of regularly varying distributions. A convex risk measure is robust on the set of distributions $\{G_{\xi;\sigma} \,|\, \sigma \leq \overline{\sigma},\ \xi \leq \overline{\xi}\}$, where $\overline{\xi} < 1$. This follows from Lemma 3.9 and the observation that, for fixed σ and $0 < \xi < 1$ the family $G_{\xi;\sigma}$ is first-order stochastically ordered in ξ (for fixed σ) and in σ (for fixed ξ).

Similarly, a convex risk measure is robust on a set of random variables that are less volatile (in convex order) than those in a given uniformly integrable set. An example is the set of conditional expectations $\{\mathbb{E}[X|\mathcal{G}] \,|\, \mathcal{G} \text{ sub-}\sigma\text{-algebra of } \mathcal{A}\}$ for $X \in \mathcal{L}^1$, see [11, p. 469].

Lemma 3.10. Let $\mathcal{U}$ be a uniformly integrable set. A convex risk measure is robust on the set

$$\mathcal{N} = \{Y \in \mathcal{L}^1 \,|\, \text{there exists } X \in \mathcal{U} \text{ such that } \mathbb{E}(f(Y)) \leq \mathbb{E}(f(X)) \text{ for all convex } f\}.$$

Proof. For $K > 0$, the function $f(x) = (|x| - K)\mathbb{1}_{\{|x|>K\}}$ is convex. Hence we have, for $Y \in \mathcal{N}$ and $X \in \mathcal{U}$ dominating Y in convex order,

$$\begin{aligned} \mathbb{E}\big(|Y|\mathbb{1}_{\{|Y|>K\}}\big) &= \mathbb{E}\big(\big(|Y| - K\big)\mathbb{1}_{\{|Y|>K\}}\big) + K\mathbb{P}(|Y| > K) \\ &\leq \mathbb{E}\big(\big(|X| - K\big)\mathbb{1}_{\{|X|>K\}}\big) + K\mathbb{P}(|Y| > K) \\ &\leq \mathbb{E}\big(|X|\mathbb{1}_{\{|X|>K\}}\big) + K\mathbb{P}(|Y| > K). \end{aligned}$$

By De la Vallée Poussin's Theorem [19], there exist a non-decreasing convex function $\psi\colon [0,+\infty) \to [0,+\infty)$ with $\psi(0) = 0$, such that $\lim_{x\to+\infty} \frac{\psi(x)}{x} = +\infty$ and $\sup_{X\in\mathcal{U}} \mathbb{E}\big(\psi(|X|)\big) < +\infty$. Applying Markov's inequality, we have

$$K\mathbb{P}(|Y| > K) \le \frac{K}{\psi(K)}\mathbb{E}\big(\psi(|Y|)\big) \le \frac{K}{\psi(K)}\mathbb{E}\big(\psi(|X|)\big).$$

By uniform integrability of $\mathcal{U}$,

$$\lim_{K\to+\infty}\sup_{Y\in\mathcal{N}} \mathbb{E}\big(|Y|\mathbb{1}_{\{|Y|>K\}}\big) \le \lim_{K\to+\infty}\sup_{X\in\mathcal{U}} \left(\mathbb{E}\big(|X|\mathbb{1}_{\{|X|>K\}}\big) + \frac{K}{\psi(K)}\mathbb{E}\big(\psi(|X|)\big)\right) = 0.$$

The conclusion follows by Theorem 3.5. □

Note that Lemma 3.10, in the special case when $\mathcal{U}$ is a singleton, follows from Proposition 3.3 in [42].

We now consider how larger uniformly integrable sets are constructed from other uniformly integrable sets. Finite unions of uniformly integrable sets are uniformly integrable, so that a convex risk measure that is robust on finitely many uniformly integrable sets is also robust on their union. Moreover, to any uniformly integrable set on which a convex risk measure is robust we can add finitely many distribution functions without losing robustness. The next proposition shows that a convex risk measure that is robust on a uniformly integrable set $\mathfrak{U} \subset \mathfrak{M}^1$ is also robust on the larger set of all possible mixtures of elements of $\mathfrak{U}$. Mixtures are used to model experimental error or contaminations, by assuming that the underlying distribution function F is contaminated with an error, with distribution G, that occurs with (small) probability $\lambda \in (0,1)$, so that the contaminated distribution is $(1-\lambda)F + \lambda G$.

Proposition 3.11. For a uniformly integrable set $\mathfrak{U} \subset \mathfrak{M}^1$, a convex risk measure is robust on the set of mixtures $\big\{(1-\lambda)F + \lambda G \mid F, G \in \mathfrak{U},\ \lambda \in [0,1]\big\}$.

Proof. By Theorem 3.5 it is enough to show that $\big\{(1-\lambda)F + \lambda G \mid F, G \in \mathfrak{U},\ \lambda \in [0,1]\big\}$ is uniformly integrable. For $\lambda \in [0,1]$ and $F, G \in \mathfrak{U}$ we calculate

$$\begin{aligned}
&\sup_{F,G\in\mathfrak{U},\lambda\in[0,1]} \int_{|x|>K} |x| \mathrm{d}[(1-\lambda)F(x) + \lambda G(x)] \\
&\le \sup_{F,G\in\mathfrak{U},\lambda\in[0,1]} (1-\lambda) \int_{|x|>K} |x|\mathrm{d}F(x) + \sup_{F,G\in\mathfrak{U},\lambda\in[0,1]} \lambda \int_{|x|>K} |x|\mathrm{d}G(x) \\
&= \sup_{F\in\mathfrak{U}} \int_{|x|>K} |x|\mathrm{d}F(x),
\end{aligned}$$

which goes to zero, as $K \to +\infty$, by uniform integrability of $\mathfrak{U}$. □

Let $\{F_\theta \mid \theta \in \Theta\}$ describe possible model inputs and assume that the set is uniformly integrable, for example a parametric family with bounded second moment. By Theorem 3.5, any convex risk measure is robust on $\{F_\theta \mid \theta \in \Theta\}$. Assume however, that the data is contaminated, through measurement errors or the parametric family does not fit sufficiently, and the risk measure is evaluated on the mixture

$$(1-\lambda)F_\theta + \lambda G, \ \text{for small } \lambda \in (0,1),\ \theta \in \Theta,\ G \in \mathfrak{N},$$

where $\mathfrak{N} \subset \mathfrak{M}^1$ denotes the collection of possible error distributions. If we have additional knowledge on the elements of $\mathfrak{N}$, such as bounded support or (uniformly) bounded mean and variance, then the convex risk measure is robust on the set of all possible mixtures, see Proposition 3.11.

3.3 Non-robustness of convex risk measures

In this section we present examples of sets on which convex risk measures fail to be robust. Such situations can emerge when the set is closed under mixtures and positive shifts. These conditions allow the construction of convergent sequences of distributions with divergent means. Thus situations arise where small changes in distribution can result in huge variations in the value of the risk measure.

Proposition 3.12. No spectral risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is robust on $\mathcal{X} \subset \mathcal{L}^1$, whenever $\mathfrak{D}(\mathcal{X})$ is closed under mixtures and contains a sequence of distribution functions whose means diverge to $+\infty$. Then, spectral risk measures are not robust at any distribution function $F \in \mathfrak{D}(\mathcal{X})$.

Proof. Let $F \in \mathfrak{D}(\mathcal{X})$ and denote by $G_k \in \mathfrak{D}(\mathcal{X})$ the sequence of distribution functions with $\lim_{k\to+\infty} \mathbb{E}[G_k] = +\infty$. Choose $C > 0$ and define the mixture

$$F^{(k)} = (1-\lambda_k)F + \lambda_k G_k, \text{ where } \lambda_k = \min\left\{\frac{C}{|\mathbb{E}[G_k]|}, 1\right\}.$$

Note that $\lambda_k \in [0, 1]$ converges to 0, as $k \to +\infty$, hence $F^{(k)}$ converges in the Prokhorov distance to F. Spectral risk measures are concave with respect to mixtures, [52], and exceed the expectation, [17], so that

$$\begin{aligned}\liminf_{n\to+\infty} \rho[F^{(k)}] &\geq \liminf_{k\to+\infty} \left((1-\lambda_k)\rho[F] + \lambda_k \rho[G_k]\right)\\ &\geq \liminf_{k\to+\infty} \left((1-\lambda_k)\rho[F] + \lambda_k \mathbb{E}[G_k]\right)\\ &= \lim_{k\to+\infty} \left((1-\lambda_k)\rho[F] + \lambda_k \mathbb{E}[G_k]\right)\\ &= \rho[F] + C.\end{aligned}$$

□

A similar result is now proved for general convex risk measures. For this, we need the additional assumption that the set $\mathfrak{D}(\mathcal{X})$ is closed under positive shifts, that is $F(\cdot - c) \in \mathfrak{D}(\mathcal{X})$ for all $c > 0$, and $F \in \mathfrak{D}(\mathcal{X})$. Note this is stronger than assuming the existence of a sequence of distribution functions with divergent mean. This additional assumption was not needed in the proof of Proposition 3.12, where instead the property of concavity with respect to mixtures of spectral risk measures [52] was used.

Proposition 3.13. No convex risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ is robust on $\mathcal{X} \subset \mathcal{L}^1$, whenever the set of distribution functions $\mathfrak{D}(\mathcal{X})$ is closed under mixtures and positive shifts. In this case, the risk measure is not robust at any distribution function $F \in \mathfrak{D}(\mathcal{X})$.

Proof. By Proposition 6.8 in [47] the risk measure is continuous with respect to $\|\cdot\|_1$. Therefore the risk measure admits the Kusuoka representation, Theorem 6.44 in [47], that is there exists a set of probability measures $\mathfrak{P}$ on $[0, 1)$ such that the risk measure can be written as

$$\rho[G] = \sup_{\mu\in\mathfrak{P}} \left(\int_0^1 \mathrm{ES}_\alpha[G]\mathrm{d}\mu(\alpha) - \beta(\mu)\right), \text{ for } G \in \mathfrak{M}^1,$$

where $\beta(\cdot)$ is a penalty function on $\mathfrak{P}$, see [47] for the definition. For $C > 0$, define the mixture $F^{(k)} = (1 - \lambda_k)F + \lambda_k G_k$, where $\lambda_k = \min\{C/k, 1\}$ and $G_k(\cdot) = F(\cdot - k)$, $k \geq 1$. Note that the mixture $F^{(k)}$ converges in the Prokhorov distance to F. Since ES_α is concave with respect to mixtures [52], we obtain for $k \geq 1$,

$$\begin{aligned}\liminf_{k\to+\infty} \rho[F^{(k)}] &= \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{\int_0^1 \mathrm{ES}_\alpha\left[(1-\lambda_k)F + \lambda_k G_k\right]\mathrm{d}\mu(\alpha) - \beta(\mu)\right\}\\ &\geq \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{\int_0^1 \left((1-\lambda_k)\mathrm{ES}_\alpha[F] + \lambda_k \mathrm{ES}_\alpha[G_k]\right)\mathrm{d}\mu(\alpha) - \beta(\mu)\right\}\\ &= \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{\int_0^1 \mathrm{ES}_\alpha[F]\mathrm{d}\mu(\alpha) - \beta(\mu) + \lambda_k \int_0^1 \left(\mathrm{ES}_\alpha[G_k] - \mathrm{ES}_\alpha[F]\right)\mathrm{d}\mu(\alpha)\right\}\end{aligned}$$

$$= \liminf_{k\to+\infty} \sup_{\mu\in\mathfrak{P}} \left\{ \int_0^1 \mathrm{ES}_\alpha[F]\mathrm{d}\mu(\alpha) - \beta(\mu) \right\} + C$$
$$= \rho[F] + C. \qquad \square$$

In Section 3.2 we have seen that for robustness of convex risk measures on the space of heavy tailed distribution functions, in particular GPDs, it is necessary that the shape parameter be bounded away from 1. The following proposition considers the case of regularly varying distribution functions. A distribution function $F \in \mathfrak{M}$ on $(0, +\infty)$ is regularly varying with tail index $\alpha > 0$, if for all $t > 0$ it holds that

$$\lim_{x\to+\infty} \frac{1 - F(xt)}{1 - F(x)} = t^{-\alpha}. \tag{2}$$

Note that, for $\xi > 0$, the GPD $G_{\xi;\sigma}$ is regularly varying with tail index $1/\xi$. The next proposition sheds some light on the trade-off between robustness of risk measures and their sensitivity to the tail of distribution functions, see also the discussion in [39].

Proposition 3.14. No convex risk measure is robust on the set of regularly varying distribution functions with tail index $\alpha > 1$.

Proof. Let $F_{\alpha_1}, F_{\alpha_2} \in \mathfrak{M}^1$ be regularly varying with indexes $\alpha_1 > 1$, respectively $\alpha_2 > 1$. We first show that the set of regularly varying distribution functions is closed under mixtures, that is

$$F = (1 - \lambda)F_{\alpha_1} + \lambda F_{\alpha_2}$$

for $\lambda \in [0, 1]$, is regularly varying. Note that $1-F = (1-\lambda)(1-F_{\alpha_1})+\lambda(1-F_{\alpha_2})$. It is clear that both $(1-\lambda)(1-F_{\alpha_1})$ and $\lambda(1-F_{\alpha_2})$ satisfy the limit in (2). Proposition 1.5.7 in [12] implies then that the sum $1-F$ of these two functions satisfies again the limit in (2) with tail index equal to $\min\{\alpha_1, \alpha_2\}$. Hence, F is a regularly varying distribution function with tail index $\min\{\alpha_1, \alpha_2\} > 1$. Clearly, any shifted regularly varying distribution function is regularly varying with the same tail index. The sequence of Pareto distributions with shape parameter $1+\frac{1}{k}$ and scale 1, that is $F_k(x) = 1 - x^{-(1+1/k)},\ x \geq 1$, belongs to the class of regularly varying distribution functions and its mean, $\mathbb{E}[F_k] = \frac{1+1/k}{1+1/k-1} = k + 1$, diverges to $+\infty$. Applying Proposition 3.13 yield the assertion. $\square$

Remark 3.15. In this paper, we consider the classical notion of robustness, defined via continuity with respect to the Prokhorov distance. A spectrum of different types of robustness, defined using alternative distances on $\mathfrak{M}$, are introduced by [39]. If a weaker notion of robustness were defined through the Wasserstein distance, see Appendix, the constructed sequence of mixtures appearing in the proof of Proposition 3.13, $(1-\lambda_k)F + \lambda_k G_k$, with $F \in \mathfrak{M}^1$ and $G_k(\cdot) = F(\cdot - k)$ would not generate a discontinuity. The mixture converges in the Prokhorov distance to F, however, its mean diverges, hence it does not converge in the Wasserstein distance, see Lemma A.1.

3.4 Generalisation to risk measures defined on $\mathcal{L}^p$

Let $p \in [1, +\infty)$ and define the space of random variables with finite p-th moment by $\mathcal{L}^p = \{X \in \mathcal{L}^0 \mid \mathbb{E}(|X|^p) < +\infty\}$. Requiring a risk measure to be real-valued on the entire space of integrable random variables excludes interesting examples such the *mean-deviation risk measures* defined by

$$\rho(X) = \mathbb{E}(X) + c\mathbb{E}\left(|X - \mathbb{E}(X)|^p\right)^{1/p}, \quad X \in \mathcal{L}^p,\ c \geq 0.$$

Note that, for every $p \in [1, +\infty)$, the mean-deviation risk measure is convex and finite on $\mathcal{L}^p$ but not on the larger space $\mathcal{L}^r$, $1 \leq r < p$ [47].

The Definition 3.1 of robustness can be generalised straightforwardly by replacing the space $\mathcal{L}^1$ with $\mathcal{L}^p$. Then, Theorem 3.5 generalises as follows.

Theorem 3.16. *Let $\rho\colon \mathcal{L}^p \to \mathbb{R}$ be a convex risk measure. Then ρ is robust on $\mathcal{X} \subset \mathcal{L}^p$ if $\mathcal{X}$ is uniformly p-integrable, that is*

$$\lim_{K\to+\infty} \sup_{X\in\mathcal{X}} \mathbb{E}\big(|X|^p \mathbb{1}_{\{|X|^p>K\}}\big) = 0.$$

The proof follows by reasoning similar to that in the proof of Theorem 3.5. Alternatively, it follows directly from [40]. We refer to [40] for a thorough study of robustness of risk measures defined on Orlicz hearts.

4 Aggregation

4.1 Robustness of aggregation measures

In risk management applications, risk measures are often evaluated on the output of a complex model, which generates portfolio losses through a non-linear function of a vector of risk factors. A typical example is the aggregated loss of an insurance portfolio, represented through the insurance company's internal model. We describe this setting through a (measurable) function $g\colon \mathbb{R}^n \to \mathbb{R}$, called *aggregation* function, that maps an n-dimensional vector into a real number. Applying the aggregation function to a random vector of input risk factors, $\boldsymbol{X} = (X_1, \dots, X_n)$ with (multivariate) cumulative distribution function $\boldsymbol{F_X}$, we can evaluate a risk measure at the (one-dimensional random) output $g(\boldsymbol{X})$. We denote the space of n-dimensional random vectors by $\boldsymbol{\mathcal{L}^0} = \boldsymbol{\mathcal{L}^0}(\Omega, \mathcal{A}, \mathbb{P})$ and the set of the corresponding (multivariate) distribution functions on $\mathbb{R}^n$ by $\boldsymbol{\mathfrak{M}} = \mathfrak{D}(\boldsymbol{\mathcal{L}^0})$. By equipping $\boldsymbol{\mathcal{L}^0}$ with the norm $\|\boldsymbol{X}\|_1 = \sum_{i=1}^n \mathbb{E}(|X_i|)$ we write $\boldsymbol{\mathcal{L}^1} = \{\boldsymbol{X} \in \boldsymbol{\mathcal{L}^0} \mid \|\boldsymbol{X}\|_1 < +\infty\}$ and $\boldsymbol{\mathfrak{M}^1} = \mathfrak{D}(\boldsymbol{\mathcal{L}^1})$.

Throughout this section, we restrict to aggregation functions g that satisfy $g(\boldsymbol{X}) \in \mathcal{L}^1$ whenever $\boldsymbol{X} \in \boldsymbol{\mathcal{L}^1}$. This is guaranteed by, for example, the linear growth condition of Definition 4.7; see also the discussion following Theorem 4.8. Weaker conditions on g could be required if more restrictions were placed on $\boldsymbol{X}$, consistently with the discussion of Section 3.4.

Definition 4.1. For an aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ and a risk measure $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ we define the *aggregation measure* $\rho_g(\cdot)\colon \boldsymbol{\mathcal{L}^1} \to \mathbb{R}$ by $\rho_g(\boldsymbol{X}) = \rho(g(\boldsymbol{X}))$.

Thus, an aggregation measure is a functional of the input vector of risk factors. An aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ induces a functional $T_g[\cdot]\colon \boldsymbol{\mathfrak{M}} \to \mathfrak{M}$ through $T_g[\boldsymbol{F_X}] = \mathfrak{D}\big(g(\boldsymbol{X})\big)$. The functional T_g takes the (multivariate) distribution functions $\boldsymbol{F_X} \in \boldsymbol{\mathfrak{M}}$ of the vector $\boldsymbol{X}$ and returns the (univariate) distribution function $T_g[\boldsymbol{F_X}] \in \mathfrak{M}$ of $g(\boldsymbol{X})$. Since risk measures are assumed to be law-invariant, all considered aggregation measures are law-invariant and can be described by a functional on the space of distribution functions $\rho_g[\cdot]\colon \boldsymbol{\mathfrak{M}^1} \to \mathbb{R}$ through

$$\rho_g[\boldsymbol{F_X}] = \rho_g(\boldsymbol{X}) = \rho\left[T_g[\boldsymbol{F_X}]\right], \text{ for } \boldsymbol{F_X} \in \boldsymbol{\mathfrak{M}^1}.$$

Note that a continuous aggregation function g induces, by the continuous mapping theorem, an aggregation functional $T_g\colon \boldsymbol{\mathfrak{M}} \to \mathfrak{M}$ that is continuous with respect to the Prokhorov distance, $\boldsymbol{\mathfrak{M}}$, $\mathfrak{M}$ both endowed with the Prokhorov distance. The Prokhorov distance on $\boldsymbol{\mathfrak{M}}$ is defined for $\boldsymbol{F}, \boldsymbol{G} \in \boldsymbol{\mathfrak{M}}$ through

$$d_P(\boldsymbol{F}, \boldsymbol{G}) = \inf\{\varepsilon > 0 \,|\, \boldsymbol{F}(\boldsymbol{B}) \le \boldsymbol{G}(\boldsymbol{B}^\varepsilon) + \varepsilon, \text{ for all Borel sets } \boldsymbol{B} \text{ on } \mathbb{R}^n\},$$

where $\boldsymbol{B}^\varepsilon = \{\boldsymbol{x} \in \mathbb{R}^n \mid \inf_{\boldsymbol{y}\in\boldsymbol{B}} |\boldsymbol{x} - \boldsymbol{y}| \le \varepsilon\}$ and, for a vector $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, we denote $|\boldsymbol{x}| = \sum_{i=1}^n |x_i|$. We say an aggregation measure $\rho_g\colon \boldsymbol{\mathcal{L}^1} \to \mathbb{R}$ is continuous on $\boldsymbol{\mathcal{X}} \subset \boldsymbol{\mathcal{L}^1}$ with respect to the Prokhorov distance if the restriction of the induced functional $\rho_g[\cdot]$ on $\mathfrak{D}(\boldsymbol{\mathcal{X}})$ is continuous with respect to d_P. That is, for all $\boldsymbol{F_0} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ and $\varepsilon > 0$ there exists $\delta > 0$ such that for all $\boldsymbol{F} \in \mathfrak{D}(\boldsymbol{\mathcal{X}})$ we have $d_P(\boldsymbol{F_0}, \boldsymbol{F}) < \delta$ implies $|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]| < \varepsilon$.

We extend Hampel's definition of robustness to aggregation measures, in order to reflect the sensitivity of the risk assessment to small perturbations in the distribution of the vector of risk factors. Clearly, for an aggregation measure $\rho_g\colon \mathcal{L}^1 \to \mathbb{R}$ a small deviation in the n-dimensional input vector includes both perturbations in the marginals and the dependence structure (copula). Analogously to the one-dimensional case, we consider estimators of risk measures evaluated at the multivariate empirical distribution function. For a distribution function $\boldsymbol{F} \in \mathfrak{M}^1$, sample size $k \geq 1$ and independent random variables $\boldsymbol{X}_1, \dots, \boldsymbol{X}_k$ with common distribution function $\boldsymbol{F}$, the multivariate empirical distribution function is given by the random measure

$$\hat{\boldsymbol{F}}_k(\boldsymbol{t}, \omega) = \frac{1}{k}\sum_{i=1}^{k} \mathbb{1}_{\{\boldsymbol{X}_i(\omega)\leq \boldsymbol{t}\}}, \ (\boldsymbol{t}, \omega) \in \mathbb{R}^n \times \Omega.$$

For an aggregation measure $\rho_g\colon \mathcal{L}^1 \to \mathbb{R}$ and a distribution function $\boldsymbol{F} \in \mathfrak{M}^1$ we define the sequence of estimators $\{\hat{\rho}_{g,k}\}_{k\geq 1}$ through its evaluation at the multivariate empirical distribution function. That is, for $k \geq 1$ we define

$$\hat{\rho}_{g,k}[\boldsymbol{F}](\omega) = \rho_g[\hat{\boldsymbol{F}}_k(\cdot, \omega)], \ \omega \in \Omega. \tag{3}$$

Note that for fixed $\boldsymbol{t} \in \mathbb{R}^n$ the multivariate empirical distribution function, $\hat{\boldsymbol{F}}_k(\boldsymbol{t}, \cdot)$, is a random variable and for fixed $\omega \in \Omega$ a distribution function. Hence, the estimator $\hat{\rho}_{g,k}[\boldsymbol{F}]$ is a random variable.

Definition 4.2. Let $\rho_g\colon \mathcal{L}^1 \to \mathbb{R}$ be an aggregation measure and $\{\hat{\rho}_{g,k}\}_k$ the sequence of estimators defined in (3). We say that the aggregation measure ρ_g is *robust* on $\mathcal{X} \subset \mathcal{L}^1$ (equivalently $\rho_g[\cdot]$ is robust on $\mathfrak{D}(\mathcal{X})$) if for any $\boldsymbol{F_0} \in \mathfrak{D}(\mathcal{X})$ it holds that for all $\varepsilon > 0$, there exists $\delta > 0$ and $k_0 \in \mathbb{N}$ such that for all $\boldsymbol{F} \in \mathfrak{D}(\mathcal{X})$ and $k \geq k_0$ we have

$$d_P(\boldsymbol{F_0}, \boldsymbol{F}) < \delta \quad \Rightarrow \quad d_P\Big(\mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F}]\big)\Big) < \varepsilon.$$

We obtain a generalisation of Hampel's theorem, Theorem 3.2, to the multivariate case. The proof follows mostly the steps of the proof of Hampel's theorem, Theorem 3.2, for distribution function on the real line [34].

Theorem 4.3. *Let $\rho_g\colon \mathcal{L}^1 \to \mathbb{R}$ be an aggregation measure and $\mathcal{X} \subset \mathcal{L}^1$. Assume that the sequence of estimators $\{\hat{\rho}_{g,k}\}_k$, defined in* (3)*, is consistent for all $\boldsymbol{F_0} \in \mathfrak{D}(\mathcal{X})$. Then, the aggregation measure ρ_g is continuous on $\mathfrak{D}(\mathcal{X})$ with respect to the Prokhorov distance if and only if it is robust on $\mathfrak{D}(\mathcal{X})$.*

Proof. Assume the aggregation measure ρ_g is continuous with respect to d_P on $\mathfrak{D}(\mathcal{X})$ and let $\boldsymbol{F_0} \in \mathfrak{D}(\mathcal{X})$. Let $\varepsilon > 0$ and $k \in \mathbb{N}$ then for all $\boldsymbol{F} \in \mathfrak{D}(\mathcal{X})$ it holds that

$$\begin{aligned} d_P\Big(\mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\hat{\rho}_{g,k}[\boldsymbol{F}]\big)\Big) &= d_P\Big(\mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}]\big), \mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_k]\big)\Big) \\ &\leq d_P\Big(\mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}]\big), \mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big)\Big) + d_P\Big(\mathfrak{D}\big(\rho_g[\boldsymbol{F_0}]\big), \mathfrak{D}\big(\rho_g[\hat{\boldsymbol{F}}_k]\big)\Big). \end{aligned} \tag{4}$$

Note that $\rho_g[\boldsymbol{F_0}]$ is a degenerate random variable. For all $\boldsymbol{F} \in \mathfrak{D}(\mathcal{X})$, the multivariate version of Glivenko-Cantelli states that the empirical distribution function $\hat{\boldsymbol{F}}_k(\cdot, \omega)$ converges for almost every ω to $\boldsymbol{F}$, as $k \to +\infty$, see [21, 48]. The first term on the right hand side in (4) can be made arbitrarily small (say $\varepsilon/2$) by choosing k large enough since the aggregation measure is consistent at $\boldsymbol{F_0}$, that is $\hat{\rho}_{g,k}[\boldsymbol{F_0}] = \rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}] \to \rho_g[\boldsymbol{F_0}]$ in probability. Next we show that the second term in (4) is smaller than $\varepsilon/2$. By continuity of the aggregation function at $\boldsymbol{F_0}$ there exists $\delta > 0$ such that, for any $\boldsymbol{F} \in \mathfrak{D}(\mathcal{X})$, $d_P(\boldsymbol{F_0}, \boldsymbol{F}) < \delta$ implies $|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]| < \frac{\varepsilon}{2}$. Thus, we obtain

$$\begin{aligned} \mathbb{P}\Big(\big|\rho_g[\boldsymbol{F_0}] - \rho_g[\hat{\boldsymbol{F}}_k]\big| \leq \frac{\varepsilon}{2}\Big) &\geq \mathbb{P}\Big(\big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\big| + \big|\rho_g[\boldsymbol{F}] - \rho_g[\hat{\boldsymbol{F}}_k]\big| \leq \frac{\varepsilon}{2}\Big) \\ &= \mathbb{P}\Big(\big|\rho_g[\boldsymbol{F}] - \rho_g[\hat{\boldsymbol{F}}_k]\big| \leq \frac{\varepsilon}{2} - \big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\big|\Big), \end{aligned}$$

where $\frac{\varepsilon}{2} - \big|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\big| > 0$. As the aggregation measure is consistent, for all $\gamma > 0$ we have $\mathbb{P}\big(\big|\rho_g[\boldsymbol{F}] - \rho_g[\hat{\boldsymbol{F}}_k]\big| \leq \gamma\big) \to 1$ as $k \to +\infty$. Hence, choosing k large enough, we obtain

$$\mathbb{P}\Big(\big|\rho_g[\boldsymbol{F_0}] - \rho_g[\hat{\boldsymbol{F}}_k]\big| \leq \frac{\varepsilon}{2}\Big) \geq 1 - \frac{\varepsilon}{2},$$

which, by Strassen's theorem [49], is equivalent to $d_P\left(\mathfrak{D}\left(\rho_g[\boldsymbol{F_0}]\right), \mathfrak{D}\left(\rho_g[\hat{\boldsymbol{F}}_k]\right)\right) < \frac{\varepsilon}{2}$.

For the converse assume that the aggregation measure is robust on $\mathfrak{D}(\mathcal{X})$. Note that for degenerate distribution functions on $\mathbb{R}$ the Prokhorov distance reduces to the absolute value. Let $\boldsymbol{F_0}, \boldsymbol{F} \in \mathfrak{D}(\mathcal{X})$ and interpreting $\rho_g[\boldsymbol{F}]$, $\rho_g[\boldsymbol{F_0}]$ as degenerate random variables we obtain for $k \in \mathbb{N}$

$$\begin{aligned}\left|\rho_g[\boldsymbol{F_0}] - \rho_g[\boldsymbol{F}]\right| &= d_P\left(\mathfrak{D}\left(\rho_g[\boldsymbol{F_0}]\right), \mathfrak{D}\left(\rho_g[\boldsymbol{F}]\right)\right) \\ &\le d_P\left(\mathfrak{D}\left(\rho_g[\boldsymbol{F_0}]\right), \mathfrak{D}\left(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\right)\right) + d_P\left(\mathfrak{D}\left(\hat{\rho}_{g,k}[\boldsymbol{F_0}]\right), \mathfrak{D}\left(\hat{\rho}_{g,k}[\boldsymbol{F}]\right)\right) \\ &\quad + d_P\left(\mathfrak{D}\left(\hat{\rho}_{g,k}[\boldsymbol{F}]\right), \mathfrak{D}\left(\rho_g[\boldsymbol{F}]\right)\right).\end{aligned}$$

The second term can be made small by robustness of the aggregation measures. The other two distances can be made arbitrarily small since the sequence of estimators is consistent for any $\boldsymbol{F} \in \mathfrak{D}(\mathcal{X})$. □

An aggregation measure composed by a continuous aggregation function and a convex risk measure is consistent at each $\boldsymbol{F} \in \mathfrak{M}^{\boldsymbol{1}}$, that is $\hat{\rho}_{g,k}[\boldsymbol{F}] \to \hat{\rho}_g[\boldsymbol{F}]$ in probability (even $\mathbb{P}$-a.s.). Hence, as a generalisation of Proposition 3.3 we obtain a one-to-one correspondence between robustness and continuity with respect to the Prokhorov distance.

Proposition 4.4. Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous aggregation function, $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ be a convex risk measure and $\mathcal{X} \subset \mathcal{L}^{\boldsymbol{1}}$. Then, the aggregation measure $\rho_g\colon \mathcal{L}^{\boldsymbol{1}} \to \mathbb{R}$ is continuous with respect to d_P on $\mathfrak{D}(\mathcal{X})$ if and only if it is robust on $\mathfrak{D}(\mathcal{X})$.

Proof. Let $\boldsymbol{F_0} \in \mathfrak{M}^1$. It is enough to show that for a continuous g and a convex risk measure ρ the aggregation measure $\rho_g = \rho \circ T_g$ is consistent. We even show strong consistency, that is $\rho_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}](\omega) \to \rho_g[\boldsymbol{F_0}]$ for almost every $\omega \in \Omega$. Since convex risk measures are continuous with respect to the Wasserstein distance, Proposition 6.8 in [47], we have to show that $d_W\left(T_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)], T_g[\boldsymbol{F_0}]\right) \to 0$ for almost every ω.

The multivariate empirical distribution function $\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)$ converges for almost every ω to $\boldsymbol{F_0}$, as $k \to +\infty$, see [21, 48]. In particular, for almost every ω, $d_P\left(\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega), \boldsymbol{F_0}\right) \to 0$, as $k \to +\infty$, and by continuity of the aggregation function, that is $T_g\colon \mathfrak{M} \to \mathfrak{M}$ is continuous w.r.t d_P, $d_P\left(T_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)], T_g[\boldsymbol{F_0}]\right) \to 0$, as $k \to +\infty$. For $k \in \mathbb{N}$ and $\omega \in \Omega$ denote by $\boldsymbol{X_0}_k^{\omega}$ a random variable that has distribution function $\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)$. Note that, by definition, $T_g[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)] = \mathfrak{D}(g(\boldsymbol{X_0}_k^{\omega})) \in \mathfrak{M}^1$ and we have

$$\int_{\mathbb{R}} |y|\, \mathrm{d}T_g\left[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)\right](y) = \int_{\mathbb{R}^n} |g(\boldsymbol{y})| \mathrm{d}\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\boldsymbol{y},\omega) = \frac{1}{k}\sum_{i=1}^{k}\left|g\left(\boldsymbol{X_0}_i(\omega)\right)\right|.$$

By the strong law of large numbers $\frac{1}{k}\sum_{i=1}^{k}\left|g\left(\boldsymbol{X_0}_i\right)\right| \to \mathbb{E}(\,|\,g(\boldsymbol{X_0})\,|\,) < +\infty$, $\mathbb{P}$-a.s. Hence for almost every $\omega \in \Omega$

$$\int_{\mathbb{R}} |y|\, \mathrm{d}T_g\left[\hat{\boldsymbol{F}}_{\boldsymbol{0}k}(\cdot,\omega)\right](y) \quad \to \quad \int_{\mathbb{R}} |y|\, \mathrm{d}T_g[\boldsymbol{F_0}](y), \text{ as } k \to +\infty.$$

The conclusion follows from Lemma A.1. □

Analogously to Theorem 3.5, robustness of the aggregation measure ρ_g depends on uniform integrability of the set of losses produced by the aggregation function g.

Theorem 4.5. *Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be a continuous aggregation function and $\rho\colon \mathcal{L}^1 \to \mathbb{R}$ a convex risk measure. Then the aggregation measure $\rho_g\colon \mathcal{L}^{\boldsymbol{1}} \to \mathbb{R}$ is robust on $\mathcal{X} \subset \mathcal{L}^{\boldsymbol{1}}$ if the set $g(\mathcal{X})$ is uniformly integrable.*

Proof. If $g(\mathcal{X})$ is uniformly integrable the risk measure is continuous with respect to d_P, see Theorem 3.5. Therefore the composition $\rho_g = \rho \circ T_g$ is continuous with respect to Prokhorov distance and by Proposition 4.4 the aggregation measure ρ_g is robust on $\mathcal{X}$. □

A similar problem is considered in [40], when the marginal distributions are fixed. Note that our extension of Hampel's classical definition of robustness to aggregation measures, Definition 4.2, requires the aggregation measure to be (relatively) insensitive to perturbations in the underlying distribution. Since the input of the aggregation measure is a random vector of risk factors, perturbation in the distribution can arise from changes in the marginals and/or the copula. Given Theorem 4.5, in order to characterise robustness of the aggregation measure ρ_g, it is necessary to study which properties of g and the set $\mathcal{X}$ produce a set of losses $g(\mathcal{X})$ that is uniformly integrable. The next section investigates this issue.

Remark 4.6. It is not necessarily the case in practice that the multivariate distribution function of $\boldsymbol{F}$ is estimated by the empirical distribution of historical data; parametric statistical methods are typically used instead. Nonetheless, the definition of robustness used here remains relevant when calculating $\rho_g[\boldsymbol{F}]$ by Monte-Carlo simulation. In that context, $\mathbf{X}$ is simulated from model $\boldsymbol{F}$ and $\hat{\boldsymbol{F}}_k$ is interpreted as the empirical distribution function of the simulated observations. Then $\rho_g[\boldsymbol{F}]$ is calculated via evaluation of $\rho_g[\hat{\boldsymbol{F}}_k]$, as is typically done in insurance internal models [46]. It is desirable that small changes in the assumed distribution $\boldsymbol{F}$ of risk factors does not produce excessive variation in the estimated aggregate risk.

4.2 Aggregation robustness and linear growth

A typical setup in risk management is linear risk aggregation, for example when aggregating different lines of business or positions in a portfolio, such that

$$\rho\big(g(\boldsymbol{X})\big) = \rho(X_1 + \cdots + X_n), \text{ for } \boldsymbol{X} \in \mathcal{L}^1. \tag{5}$$

By Sklar's theorem the distribution of vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is specified through its marginals and its dependence structure (copula). Statistically, estimating copulas can be very challenging and often relies on expert judgement. Since diverse dependence structures can lead to substantial differences in aggregate risk, risk management is especially concerned about misspecification in the copula. A substantial literature exists on *dependence uncertainty*, including calculations of upper and lower bounds for (5), for fixed marginals $X_i \sim F_i,\ i = 1, \ldots, n$ and an unspecified copula, see [9, 25, 55] and references therein.

Furthermore, [26] show that, when ρ is a spectral risk measure, the aggregation measure defined through (5) is robust on the set $\{(X_1, \ldots, X_n) \,|\, X_i \sim F_i,\ i = 1 \ldots, n\}$, where $F_1, \ldots F_n \in \mathfrak{M}^1$ are fixed marginal distributions. Taking a step further, [40] consider robustness of convex risk measures composed with non-linear aggregation functions for fixed marginals, see discussion after Theorem 4.8. Here, we build on [26, 40] by considering robustness in the more general case of uncertainty in both the dependence structure and the marginals of the model input $\boldsymbol{X}$. Theorem 4.8 below shows that robustness is guaranteed if the aggregation function satisfies a linear growth condition in the tail, similar to that of [40], and the marginals belong to uniformly integrable sets.

For sets of univariate distribution functions $\mathfrak{N}_i \subset \mathfrak{M}^1,\ i = 1, \ldots n$, we define the set of all possible random vectors $\boldsymbol{X} = (X_1, \ldots, X_n)$ with marginals F_{X_i} belonging to the corresponding sets $\mathfrak{N}_i,\ i = 1, \ldots, n$, through

$$\mathfrak{C}(\mathfrak{N}_1, \ldots \mathfrak{N}_n) = \left\{(X_1, \ldots, X_n) \,|\, F_{X_i} \in \mathfrak{N}_i \subset \mathfrak{M}^1,\ i = 1, \ldots n\right\} \subset \mathcal{L}^1.$$

Definition 4.7. We say an aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$ possesses the *linear growth condition in the tail*, if there exist $A, L, M > 0$ such that

$$|g(\boldsymbol{x})| \leq A + L|\boldsymbol{x}|, \text{ for all } |\boldsymbol{x}| > M.$$

Continuity of g combined with linear growth in the tail as in Definition 4.7 guarantee that $g(\boldsymbol{X}) \in \mathcal{L}^1$ for $\boldsymbol{X} \in \mathcal{L}^1$.

Theorem 4.8. *Let the sets $\mathfrak{U}_1, \ldots \mathfrak{U}_n \in \mathfrak{M}^1$ be uniformly integrable, the function g be continuous and satisfy the linear growth condition in the tail, and ρ be a convex risk measure. Then the aggregation measure $\rho_g\colon \mathcal{L}^1 \to \mathbb{R}$ is robust on $\mathfrak{C}(\mathfrak{U}_1, \ldots \mathfrak{U}_n)$.*

Proof. By Theorem 4.5 it is enough to show that $g\big(\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)\big)$ is uniformly integrable. The aggregation function g is continuous on the compact set $\{\boldsymbol{x}\in\mathbb{R}^n \,|\, |\boldsymbol{x}|\le M\}$, hence there exists $C>0$ such that $\sup_{|\boldsymbol{x}|\le M}|g(\boldsymbol{x})|\le C$.

Let $\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)$. For $K>\max\{A,C\}$ we have that $\{|\boldsymbol{X}|\le M\cap|g(\boldsymbol{X})|>K\}=\varnothing$, thus

$$\begin{aligned}\sup_{g(\boldsymbol{X})\in g(\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n))}\mathbb{E}\big(|g(\boldsymbol{X})|\mathbb{1}_{\{|g(\boldsymbol{X})|>K\}}\big)&=\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\mathbb{E}\big(|g(\boldsymbol{X})|\mathbb{1}_{\{|g(\boldsymbol{X})|>K\}}\mathbb{1}_{\{|\boldsymbol{X}|>M\}}\big)\\&\le L\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\mathbb{E}\Big(\sum_{i=1}^n|X_i|\mathbb{1}_{\{L\sum_{i=1}^n|X_i|>K-A\}}\Big)\\&\quad+A\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\mathbb{P}\Big(L\sum_{i=1}^n|X_i|>K-A\Big).\end{aligned}\tag{6}$$

The first term in (6) can be bounded as follows. Note that for $d\ge0$ and $x_1,\dots,x_n\in\mathbb{R}$, there exists j such that $\max_{i=1,\dots,n}|x_i|\mathbb{1}_{\{\max_{i=1,\dots,n}|x_i|>d\}}=|x_j|\mathbb{1}_{\{|x_j|>d\}}\le\sum_{i=1}^n|x_i|\mathbb{1}_{\{|x_i|>d\}}$. Therefore,

$$\begin{aligned}L\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\mathbb{E}\Big(\sum_{i=1}^n|X_i|\mathbb{1}_{\{L\sum_{i=1}^n|X_i|>K-A\}}\Big)&\le L\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\mathbb{E}\big(n\max_{i=1,\dots,n}|X_i|\mathbb{1}_{\{nL\max_{i=1,\dots,n}|X_i|>K-A\}}\big)\\&\le nL\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\sum_{i=1}^n\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>(K-A)/(nL)\}})\\&\le nL\sum_{i=1}^n\sup_{F_{X_i}\in\mathfrak{U}_i}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>(K-A)/(nL)\}})\to0,\end{aligned}$$

as $K\to+\infty$, by uniform integrability of each $\mathfrak{U}_i$. For the second term in (6) we use Markov's inequality

$$A\sup_{\boldsymbol{X}\in\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)}\mathbb{P}\Big(\sum_{i=1}^n|X_i|>\frac{K-A}{L}\Big)\le\frac{AL}{K-A}\sum_{i=1}^n\sup_{F_{X_i}\in\mathfrak{U}_i}\mathbb{E}(|X_i|),$$

which goes to zero as $K\to+\infty$. □

Note that Theorem 4.8 requires assumptions on the marginal distributions of $\boldsymbol{X}$, but not on its dependence structure. Hence robustness of convex risk measures holds even in the presence of complete dependence uncertainty, where no information on the copula exists. Theorem 4.8 offers a slight generalisation of Theorem 4.23 in [40] to the case of uncertainty in the marginal distributions. Also, [40] require a global linear growth condition, while we use linear growth in the tail combined with continuity of g.

An immediate consequence of Theorem 4.8 involves linear aggregation.

Corollary 4.9. Let the function g be given by $g(\boldsymbol{x})=\sum_{i=1}^n x_i$, $\boldsymbol{x}\in\mathbb{R}^n$. For a convex risk measure ρ, the aggregation measure ρ_g is robust on $\mathcal{C}(\mathfrak{U}_1,\dots\mathfrak{U}_n)$, with $\mathfrak{U}_i$, $i=1,\dots,n$, uniformly integrable.

There also exist many relevant continuous non-linear aggregation functions that satisfy the linear growth condition of Definition 4.7. It is easiest to envisage such situations arising in the context of reinsurance, with the elements of the random vector $\boldsymbol{X}$ representing insurance liabilities (losses from lines of business or individual policies). Then, by standard considerations of insurable interest and moral hazard, it is not plausible to have (re)insurance portfolio losses that increase in $\boldsymbol{X}$ faster than linearly. Note that optimal Pareto reinsurance contracts are typically Lipschitz continuous [15] and hence possess the linear growth condition. For example, a reinsurance company faces the aggregate risk of excess-of-loss reinsurance contracts on individual risks $X_1,\dots,X_n$ with deductibles d_i and limits $c_i>d_i$, $i=1,\dots,n$, such that

$$g(\boldsymbol{X})=\sum_{i=1}^n(X_i-d_i)_+-(X_i-c_i-d_i)_+,$$

where $(x)_+ = \max\{x, 0\}$. Alternatively, a reinsurance company taking the risk that an aggregated portfolio exceeds $c > 0$, faces claim

$$g(\boldsymbol{X}) = \Big(\sum_{i=1}^{n} X_i - c\Big)_+.$$

Note that in the first example g is constant for large $\boldsymbol{x}$ and in the second case it is linear in its marginals, hence fulfilling in both cases the linear growth condition in the tail.

Alternatively, one could view $g(\boldsymbol{X})$ as a portfolio of financial derivatives with underlyings $\boldsymbol{X}$, such that $g(\boldsymbol{X}) = \sum_{i=1}^{n} h_i(X_i)$. Standard options, even leveraged ones with pay-offs of the form $h_i(x) = (\lambda x - c)_+$, $\lambda > 1$, satisfy the linear growth condition. However, note that other exotic options, such as powered options of the form $h_i(x) = \big((x - c)_+\big)^p$, with $p > 1$, do not satisfy the linear growth condition. To achieve robustness for such pay-offs, one would need to restrict $\boldsymbol{X}$ to $\mathcal{L}^p$, see Section 3.4. For details on such derivatives see [36], pp. 168-169.

4.3 Aggregation through compound distributions

A common form of aggregation in insurance (as well as operational and credit risk modelling), takes place via compound distributions that model the future total claim amount as a random sum of individual claims. Within a specific (homogeneous) line of business, individual claims are modelled as independent and identically distributed positive random variables X_i and the (unknown) number of claims through a (discrete and random) count variable N independent of the X_i. The total claim amount $X_1 + \cdots + X_N$ cannot be readily expressed via an aggregation function $g\colon \mathbb{R}^n \to \mathbb{R}$. However, the distribution function of the random sum can be straightforwardly defined through an aggregation operator T acting on distributions, namely

$$T[\cdot,\cdot]\colon \mathfrak{M}^1 \times \mathfrak{M}^1 \to \mathfrak{M}^1; \quad T[F, G] = \mathfrak{D}\Big(\sum_{i=1}^{N} X_i\Big),\ X_i \sim_{i.i.d.} F \text{ independent of } N \sim G. \tag{7}$$

Therefore, $T[F, G] = \int_0^{+\infty} F^{\star(n)}(\cdot)\mathrm{d}G(n)$, where $F^{\star(n)}$ is the n-th convolution of F.

Theorem 4.10. *Let $\mathfrak{U}$ be a uniformly integrable set of distribution functions on $[0, +\infty)$ and $\mathfrak{N}$ a uniformly integrable set of distributions on the non-negative integers, such that*

$$\int_0^{+\infty} x\,\mathrm{d}F^\star(x) < +\infty \text{ and } \int_0^{+\infty} x\,\mathrm{d}G^\star(x) < +\infty,$$

where $F^\star$ and $G^\star$ are distribution functions given by $F^\star = \inf_{F\in\mathfrak{U}} F$ and $G^\star = \inf_{G\in\mathfrak{N}} G$ respectively. Let the operator T be defined by (7) and ρ be a convex risk measure. Then, the aggregation measure defined by $\rho \circ T\colon \mathfrak{M}^1 \times \mathfrak{M}^1 \to \mathbb{R}$ is robust on $\mathfrak{U} \times \mathfrak{N}$.

Proof. By Theorem 4.5 it is enough to show that the set $\{T[F, G] \mid F \in \mathfrak{U},\ G \in \mathfrak{N}\}$ is uniformly integrable. Note that $F^\star$ is a distribution function. Indeed, $F^\star$ is non-decreasing, right continuous (the infimum of a family of right continuous non-decreasing, hence upper semi-continuous, functions is right continuous, see Lemma 2.39 [2]) and

$$\lim_{x\to+\infty}(1 - F^\star(x)) \le \lim_{x\to+\infty}\sup_{F\in\mathfrak{U}} x(1 - F(x)) \le \lim_{x\to+\infty}\sup_{F\in\mathfrak{U}} \int_{y>x} y\mathrm{d}F(y) = 0,$$

by uniform integrability. Analogously, $\inf_{G\in\mathfrak{N}} G$ is a distribution function on the non-negative integers. Choose $F \in \mathfrak{U}$ and $G \in \mathfrak{N}$ and denote $X_i \sim_{i.i.d.} F$ and $N \sim G$ independent of the X_i. Similarly, denote $X_i^\star \sim_{i.i.d.} F^\star$ and $N^\star \sim G^\star$ independent of the $X_i^\star$ and note that $X_i^\star$ and $N^\star$ first-order stochastically dominate X_i and N, respectively. As first-order stochastic dominance is preserved under compounding, see Proposition 3.3.31 in [18], $\sum_{i=1}^{N} X_i$ is lower than $\sum_{i=1}^{N^\star} X_i^\star$ in first-order stochastic dominance. The result follows from Lemma 3.9 and the fact that the compound sum $\sum_{i=1}^{N^\star} X_i^\star$ is integrable, given the integrability of $X_i^\star$ and $N^\star$. □

Examples of sets of distribution functions on the non-negative integers fulfilling the assumptions of Theorem 4.10 include the Poisson distribution with parameter $0 < \lambda \le \overline{\lambda}$ and the Geometric with $p \ge \underline{p} > 0$, see Table 3.1 in [18]. For the claim size distribution, an example is the family of Pareto distributions $F(x) = 1 - x_m^\alpha x^{-\alpha}$ with parameters $0 < x_m \le \overline{x}_m$ and $\alpha \ge \underline{\alpha} > 1$ or, more generally, the set of GPDs, $\{G_{\xi;\sigma} \mid \sigma \le \overline{\sigma},\ 0 < \xi \le \overline{\xi}\}$, $\overline{\xi} < 1$.

5 Comparison to robustness regions of Value-at-Risk

In this section we compare the robustness properties of the popular non-convex risk measure VaR to those of the convex risk measures studied in this paper. Since different risk measures are robust on different sets, the choice of risk measure should also reflect information on the plausible sets of distribution functions expected to be encountered in particular applications.

VaR at level $\alpha \in (0, 1)$ is defined as the left-sided α-quantile, $\mathrm{VaR}_\alpha[F] = F^{-1}(\alpha) = \inf\{y \in \mathbb{R} \mid F(y) \ge \alpha\}$. It is known that VaR_α is not robust on the whole of $\mathfrak{M}^1$; however, it is robust on the set of distribution functions that are strictly increasing in a neighbourhood of their α-quantile [14, 33]. In particular, VaR is not robust on discrete random variables and hence the set where VaR is not robust is dense in $\mathfrak{M}^1$.

The following insurance example, where strict increasingness is not satisfied, leads to non-robustness of VaR. Consider the risk exposure $Y = \min\{X, d\}$, $X \in \mathcal{L}^1$, that occurs when an insurer with exposure X buys reinsurance protection with deductible $d \ge 0$. The distribution of Y, $F_Y(x) = F_X(x)\mathbb{1}_{\{x<d\}} + \mathbb{1}_{\{x\ge d\}}$, is flat for all $x > d$, hence VaR_α is not robust at F_Y whenever $\alpha \ge F_X(d)$.

Thus, neither convex risk measures such as ES nor VaR, are robust on $\mathcal{L}^1$. VaR requires strictly increasing distribution functions. Convex risk measures like ES place requirements on the tail of the underlying distribution functions via the uniform integrability condition, see Theorem 3.5. A comparative assessment of those two risk measures thus relies on whether strict increasingness or uniform integrability is a more realistic constraint on the set of distributions on which the risk measure is to be evaluated. This depends on the context of the application. For example, in reinsurance problems where distributions with constant parts can occur, uniform integrability may be a more suitable assumption. On the other hand, when dealing with an asset return with an approximately bell-shaped density but arbitrarily heavy tails, strict increasingness of the distribution appears to be a more appropriate condition.

Turning now to the case of risk aggregation, consider the aggregation measure defined by $\mathrm{VaR}_{\alpha,g} \colon \boldsymbol{\mathcal{L}^1} \to \mathbb{R}$, where $g \colon \mathbb{R}^n \to \mathbb{R}$ is an aggregation function. The aggregation measure $\mathrm{VaR}_{\alpha,g}$ will not be robust if the distribution of $g(\boldsymbol{X})$ is constant in a neighbourhood of $F^{-1}_{g(\boldsymbol{X})}(\alpha)$. Such flat regions can emerge due to the nature of function g. For instance, in a slight generalisation of the previous example, for an insurance company that buys an unlimited layer of reinsurance protection for its portfolio, we have $g(\boldsymbol{X}) = \min\{\sum_{i=1}^n X_i, d\}$.

Flat regions in the distribution of $g(\boldsymbol{X})$ can also appear through the effect of the dependence structure of $\boldsymbol{X}$. This is exemplified by the special case of linear portfolio aggregation, $g(\mathbf{x}) = \sum_{i=1}^n x_i$. Then the aggregation measure $\mathrm{VaR}_{\alpha,g}$ is not robust on a set $\mathcal{X} \subset \boldsymbol{\mathcal{L}^1}$ if there exists an input vector $\boldsymbol{X} \in \mathcal{X}$ such that $X_1 + \cdots + X_n$ is discrete for large values. Example 2.2 in [26] provides explicit choices of marginals and copulas that lead to non-robustness of the aggregation measure $\mathrm{VaR}_{\alpha,g}$ through the construction of a degenerate aggregate risk. The problem of the existence of a dependence structure of random variables $X_1, \dots, X_n$, such that the aggregated risk $X_1 + \cdots + X_n$ is almost surely constant, is extensively studied in probability theory and risk management [41, 44]. Examples of distribution functions include $F_1 = \cdots = F_n$ being Gaussian or Cauchy; we refer the reader to [55] and references therein in the context of risk management.

In quantitative risk management applications, one is often concerned about aggregate risks. Seldom is a risk measure evaluated on a random loss that does not in turn depend on further risk factors. A particular example is the use of internal models in insurance for calculating capital requirements across the portfolio. Compared to evaluating a risk measure on a real-valued random variable, in risk aggregation, there is the additional complication of the dependence structure of the input vector. Thus, there are two sources of uncertainty, in the marginal distributions and in the dependence structure. Modelling accurately the dependence structure is usually more challenging than modelling marginals, due to a lack of extensive multivariate datasets. Therefore, it is critical that the risk measure is robust to changes in the dependence structure.

We have seen that robustness of aggregation measures derived from convex risk measures, such as ES, depends on weak assumptions on the aggregation function g and the marginals, while no requirements are placed on the dependence structure. On the other hand, robustness of VaR requires restricting both the form of the aggregation function g and the possible dependence structures of the input vector. In applications such as the internal modelling performed by insurers, such constraints are not necessarily realistic. Thus our paper indicates that in applications where (non-linear) aggregations are present and high dependence uncertainty persists, the use of convex risk measures may be preferable to that of VaR.

A Wasserstein space

For $F, G \in \mathfrak{M}^1$, the Wasserstein distance [20, 31] is given by

$$d_W(F, G) = \int_{\mathbb{R}} |F(x) - G(x)| \mathrm{d}x = \int_0^1 |F^{-1}(u) - G^{-1}(u)| \mathrm{d}u,$$

where $F^{-1}(u) = \inf\{y \in \mathbb{R} \mid F(y) \geq u\}$, $u \in [0, 1)$, is the generalised inverse and we identify $\inf \varnothing = -\infty$.

Lemma A.1. (Lemma 8.3 in [10])
For $F, F_k \in \mathfrak{M}^1$, $k \geq 1$ the following are equivalent

i) $d_W(F_k, F) \to 0$, as $k \to +\infty$.
ii) $d_P(F_k, F) \to 0$ and $\int_{\mathbb{R}} |x| \mathrm{d}F_k(x) \to \int_{\mathbb{R}} |x| \mathrm{d}F(x)$, as $k \to +\infty$.
iii) $d_P(F_k, F) \to 0$ and the set $\{F_k \mid k \geq 1\}$ is uniformly integrable.

Lemma A.2. A risk measure $\rho \colon \mathcal{L}^1 \to \mathbb{R}$ is continuous with respect to the norm $\|\cdot\|_1$ on $\mathcal{L}^1$ if and only if it is continuous with respect to the Wasserstein distance on $\mathcal{L}^1$.

Proof. Assume that the risk measure is continuous with respect to $\|\cdot\|_1$. On $\mathcal{L}^1$ a sequence of random variables X_n converges in the Wasserstein distance to X if and only if there exist random variables $\tilde{X}_n$ on $\mathcal{L}^1$ with the same distribution as X_n and $\tilde{X}$ with the same distribution as X such that $\|\tilde{X}_n - \tilde{X}\|_1 \to 0$, see Theorem 3.5 in [39]. Hence by law-invariance of the risk measure

$$\rho(X_n) = \rho(\tilde{X}_n) \to \rho(\tilde{X}) = \rho(X), \text{ as } n \to +\infty.$$

For $X, Y \in \mathcal{L}^1$ the inequality $d_W(X, Y) \leq \|X - Y\|_1$ implies that a sequence converging in $\|\cdot\|_1$ also converges in the Wasserstein distance. Hence continuity with respect to d_W implies continuity with respect to $\|\cdot\|_1$. □

On the set of integrable distribution functions over $\mathbb{R}^n$, that is $\boldsymbol{\mathfrak{M}^1} = \mathfrak{D}(\boldsymbol{\mathcal{L}^1})$, the Wasserstein distance is defined for $\boldsymbol{F}, \boldsymbol{G} \in \boldsymbol{\mathfrak{M}^1}$ by

$$d_W(\boldsymbol{F}, \boldsymbol{G}) = \inf\left\{\mathbb{E}\Big(\|\boldsymbol{X} - \boldsymbol{Y}\|_1\Big) \,\Big|\, \boldsymbol{X} \sim \boldsymbol{F},\ \boldsymbol{Y} \sim \boldsymbol{G}\right\},$$

where the infimum is taken over all joint distribution functions of dimension $2n$ with marginals $\boldsymbol{F}$ and $\boldsymbol{G}$ of size n. Note that on the real line we have the dual representation $d_W(F, G) = \inf\{\mathbb{E}(|X - Y|) \mid X \sim F,\ Y \sim G\} = \int_{\mathbb{R}} |F(x) - G(x)| \mathrm{d}x$, $F, G \in \mathfrak{M}^1$ [51].

Acknowledgement: The authors are grateful to two referees, whose feedback significantly improved the paper. Furthermore, we thank Alfred Müller and Ruodu Wang for helpful comments.

Previous versions of this paper have been presented at the *3rd European Actuarial Journal Conference* (Lyon) and the conference on *Model Uncertainty and Robust Finance* (Milan).

References

[1] Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *J. Bank. Financ. 26*(7), 1505–1518.
[2] Aliprantis, C. D. and K. Border (2006). *Infinite Dimensional Analysis: a Hitchhiker's Guide*. Springer, Berlin.
[3] Artzner, P., F. Delbaen, J.-M. Eber, and E. Heath (1997). Thinking coherently. *RISK 10*, 68–71.
[4] Artzner, P., F. Delbaen, J.-M. Eber, and E. Heath (1999). Coherent measures of risk. *Math. Finance 9*(3), 203–228.
[5] Bäuerle, N. and A. Müller (2006). Stochastic orders and risk measures: consistency and bounds. *Insurance Math. Econom. 38*(1), 132–148.
[6] BCBS (2012). Consultative Document May 2012. Fundamental review of the trading book. Technical report, Basel Committee on Banking Supervision, Bank for International Settlements.
[7] BCBS (2013). Consultative Document October 2013. Fundamental review of the trading book: A revised market risk framework. Technical report, Basel Committee on Banking Supervision, Bank for International Settlements.
[8] Bellini, F., B. Klar, A. Müller, and E. Rosazza Gianin (2014). Generalized quantiles as risk measures. *Insurance Math. Econom. 54*, 41–48.
[9] Bernard, C., X. Jiang, and R. Wang (2014). Risk aggregation with dependence uncertainty. *Insurance Math. Econom. 54*, 93–108.
[10] Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *Ann. Statist. 9*(6), 1196–1217.
[11] Billingsley, P. (2008). *Probability and Measure*. Third edition. John Wiley & Sons, New York.
[12] Bingham, N. H., C. M. Goldie, and J. L. Teugels (1989). *Regular Variation*. Cambridge University Press.
[13] Cheridito, P. and T. Li (2009). Risk measures on Orlicz hearts. *Math. Finance 19*(2), 189–214.
[14] Cont, R., R. Deguest, and G. Scandolo (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance 10*(6), 593–606.
[15] Dana, R.-A. and M. Scarsini (2007). Optimal risk sharing with background risk. *J. Econom. Theory 133*(1), 152–176.
[16] Delbaen, F. (2002). Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics*, pp. 1–37. Springer, Berlin.
[17] Denneberg, D. (1990). Premium calculation: why standard deviation should be replaced by absolute deviation. *Astin Bull. 20*(2), 181–190.
[18] Denuit, M., J. Dhaene, M. Goovaerts, and R. Kaas (2006). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons, Chichester.
[19] Diestel, J. (1991). Uniform integrability: an introduction. *Rendiconti dell'Istituto di Matematica dell'Universitá di Trieste 23*, 41–80.
[20] Dobrushin, R. (1970). Prescribing a system of random variables by conditional distributions. *Theory Probab. Appl. 15*(3), 458–486.
[21] Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press.
[22] Durrett, R. (2013). *Probability: Theory and Examples*. Edition 4.1. 4th edition published by Cambridge University Press in 2010.
[23] EIOPA (2009). Directive 2009/138/EC of the European Parliament and of the Council. Technical report, European Insurance and Occupational Pensions Authority.
[24] Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
[25] Embrechts, P., G. Puccetti, and L. Rüschendorf (2013). Model uncertainty and VaR aggregation. *J. Bank. Financ. 37*(8), 2750–2764.
[26] Embrechts, P., B. Wang, and R. Wang (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stoch. 19*(4), 763–790.
[27] Filipović, D. and G. Svindland (2012). The canonical model space for law-invariant convex risk measures is L^1. *Math. Finance 22*(3), 585–589.
[28] Föllmer, H. and A. Schied (2002). Convex measures of risk and trading constraints. *Finance Stoch. 6*(4), 429–447.
[29] Föllmer, H. and A. Schied (2011). *Stochastic Finance: An Introduction in Discrete Time*. Third edition. Walter de Gruyter, Berlin.
[30] Frittelli, M. and E. Rosazza Gianin (2002). Putting order in risk measures. *J. Bank. Financ. 26*(7), 1473–1486.
[31] Givens, C. R. and S. M. Rae (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J. 31*(2), 231–240.
[32] Haezendonck, J. and M. Goovaerts (1982). A new premium calculation principle based on Orlicz norms. *Insurance Math. Econom. 1*(1), 41–53.

[33] Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat. 42*(6), 1887–1896.
[34] Huber, P. and E. M. Ronchetti (2009). *Robust Statistics*. Second edition. Wiley, New York.
[35] IAIS (2014). Consultation Document December 2014. Risk-based global insurance capital standard. Technical report, International Association of Insurance Supervisors.
[36] Jarrow, R. A. and S. M. Turnbull (2000). *Derivative Securities*. Second Edition. South-Western College Publishing, Cincinnati OH.
[37] Kaina, M. and L. Rüschendorf (2009). On convex risk measures on L^p-spaces. *Math. Methods Oper. Res. 69*(3), 475–495.
[38] Kiesel, R., R. Rühlicke, G. Stahl, and J. Zheng (2016). The Wasserstein metric and robustness in risk management. *Risks 4*(3), 32.
[39] Krätschmer, V., A. Schied, and H. Zähle (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch. 18*(2), 271–295.
[40] Krätschmer, V., A. Schied, and H. Zähle (2015). Domains of weak continuity of statistical functionals with a view toward robust statistics. arXiv preprint arXiv:1511.08677.
[41] Makarov, G. D. (1982). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory Probab. Appl. 26*(4), 803–806.
[42] Mao, T. and R. Wang (2015). On aggregation sets and lower-convex sets. *J. Multivariate Anal. 138*, 170–181.
[43] Nelder, J. A. and R. W. M. Wedderburn (1973). Generalized linear models. *J. R. Stat. Soc. Ser. A 135*(3), 370–384.
[44] Rüschendorf, L. (1982). Random variables with maximum sums. *Adv. Appl. Probab. 14*(3), 623–632.
[45] Sandström, A. (2016). *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*. CRC Press, Boca Raton FL.
[46] SCOR (2008). From principle-based risk management to solvency requirements, 2nd edition. Technical report, SCOR Switzerland AG.
[47] Shapiro, A., D. Dentcheva, and A. Ruszczyński (2009). *Lectures on Stochastic Programming: Modeling and Theory* (Second ed.). SIAM, Philadelphia PA.
[48] Shorack, G. R. and J. A. Wellner (2009). *Empirical Processes with Application to Statistics*. SIAM.
[49] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist. 36*(2), 423–439.
[50] Tsanakas, A. and P. Millossovich (2016). Sensitivity analysis using risk measures. *Risk Analysis 36*(1), 30–48.
[51] Vallender, S. (1974). Calculation of the Wasserstein distance between probability distributions on the line. *Theory Probab. Appl. 18*(4), 784–786.
[52] Wakker, P. (1994). Separating marginal utility and probabilistic risk aversion. *Theory and Decision 36*(1), 1–44.
[53] Wang, B. and R. Wang (2011). The complete mixability and convex minimization problems with monotone marginal densities. *J. Multivariate Anal. 102*(10), 1344 – 1360.
[54] Wang, B. and R. Wang (2016). Joint mixability. *Math. Oper. Res. 41*(3), 808–826.
[55] Wang, R., L. Peng, and J. Yang (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance Stoch. 17*(2), 395–417.
[56] Wang, S. (1996). Premium calculation by transforming the layer premium density. *Astin Bull. 26*(l), 71–92.
[57] Wüthrich, M. V. (2016). Non-Life Insurance: Mathematics & Statistics. *Available at SSRN: 2319328*.