Distributional transforms, probability distortions, and their applications

Peng Liu* Alexander Schied† Ruodu Wang‡

Department of Statistics and Actuarial Science, University of Waterloo, Canada

March 19, 2020

Abstract

In this paper we provide a general mathematical framework for distributional transforms, which allows for many examples that are used extensively in the literature of finance, economics and optimization. We put a special focus on the class of probability distortions, which is a fundamental tool in decision theory. As our main results, we characterize distributional transforms satisfying various properties and this includes a set of conditions which forces a distributional transform to be a probability distortion. As the first application, we construct new risk measures using distributional transforms. Sufficient and necessary conditions are given to ensure the convexity or coherence of the generated risk measures. In the second application, we introduce a new method for sensitivity analysis of risk measures based on composition groups of probability distortions. Finally, we construct probability distortions describing a change of measures with an example in option pricing.

Key-words: distributional transforms, probability distortions, risk measures, option pricing, sensitivity analysis, change of measures, Value-at-Risk, Expected Shortfall, composition of groups

*Email: peng.liu1@uwaterloo.ca

†Email: aschied@uwaterloo.ca

‡Email: wang@uwaterloo.ca

The authors thank the Editor, an Associate Editor and two anonymous referees for helpful comments. PL, AS and RW gratefully acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC) through grants RGPIN-2017-04054, RGPIN-2018-03823 and RGPAS-2018-522590. RW also acknowledges support from the Society of Actuaries CAE grant at the University of Waterloo.

1 Introduction

Distributional transforms are mappings between sets of distribution functions. They have been extensively used in finance, risk analysis and economics. For instance, in asset pricing theory, calculating the no-arbitrage price of the European call/put options requires a change of measure, which can be viewed as a distributional transform (Section 7). In risk analysis, when it comes to the sensitivity of risk measures, the first step is to add some shock to the concerned loss distribution, resulting in a convolutional transform (Example 7); see Tsanakas and Millossovich (2016) and the reference therein. In economics, the Gini coefficient, the most commonly used measurement of inequality for wealth distribution, is based on the Lorenz curve (Lorenz (1905) and Gastwirth (1971); see Example 6). In the expected utility theory, utility functions change the values of random variables, which is a shape distribution transform (Example 5).

As a special class of distributional transforms, probability distortions have played an important role in decision theory, finance and risk optimization. In particular, the expected value of a distorted distribution is known as Yaari's dual utility (Yaari (1987)) or a distortion risk measure. Distortion risk measures include the Value-at-Risk (VaR) and the Expected Shortfall (ES, or CVaR) which are widely used in risk management and operations research (see e.g. McNeil et al. (2015), Föllmer and Schied (2016), and the references therein). In the rank-dependent expected utility (Quiggin (1982, 1993)), probability distortions are used to represent individual perception of unlikely extreme risk outcomes. Cumulative prospect theory (Tversky and Kahneman (1992)) is proposed based on the probability distortions representing the attitude towards relative loss and gain. Expected value of the distorted distribution is also used in calculation of risk-adjusted premium and option prices (Wang (2000)) and in performance measures (Cherny and Madan (2009)). Besides, one can find many other examples of probability distortions in the literature such as tail transform (Example 1) representing tail risk (Liu and Wang (2020)), and several probability distortions appeared in Wang (1996) describing the premium principle in insurance. In the above literature, the focus is usually put on the expected value of the distorted distribution, instead of the probability distortion itself.

These motivate us to develop a systematic theory for the general distributional transforms in this paper. An exact definition of distributional transforms is given in Section 2. It applies in particular to the class of probability distortions, and we will also provide several other examples. We also introduce some properties that are crucial for the characterization of distributional transforms.

Since general distributional transforms are abstract, they may be difficult to investigate. One of our main aims is to characterize explicit expression for distributional transforms satisfying some properties. This has been done in Section 3. With the aid of the properties listed in Section

2, we find an equivalent set of conditions to ensure that distributional transforms are probability distortions. This set of conditions includes monotonicity, commutation with shape transforms and lower-semicontinuity. In addition, we characterize the distributional transforms satisfying monotonicity and additivity by expressing its quantile through Choquet integrals.

Using distributional transforms, we can define new risk measures generated by a distributional transform and another risk measure (such as the mean). In Section 5, we find conditions of distributional transforms (probability distortions) under which the generated risk measures are convex/coherent provided the original risk measures are convex/coherent.

A new method for sensitivity analysis based on probability distortions is introduced in Section 6. Instead of an additive shock to the original loss, we apply a composition group of convex/concave probability distortions (see Cherny and Filipović (2011)) to the loss distribution to get the sensitivity of risk measure. We obtain the explicit expressions of sensitivity of a class of risk measures with respect to general composition groups. In particular, we give the exact expressions for the sensitivity of the Expected Shortfall (ES) and the Value-at-Risk (VaR). Numerical analysis has also been carried out to the sensitivity analysis of ES and VaR with respect to three composition groups for Pareto and normal distributions.

Finally, in Section 7, we apply probability distortions to obtain formulas for change of measures. We construct probability distortions which are mappings from the distributions of a class of random variables under the original probability measure to the distributions of the same random variables under a new probability measure. As an example, this result recovers the Black-Scholes formula in a Black-Scholes market.

2 Distributional transforms and probability distortions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space. For $r \in [0, \infty]$, let $\mathcal{M}_r$ denote the set of all distributions of random variables in $L^r(\Omega, \mathcal{F}, \mathbb{P})$. We set $\mathcal{M}$ as a generic subset of $\mathcal{M}_0$.

A *distributional transform* T is a mapping from $\mathcal{M}$ to $\mathcal{M}_0$. In the majority of our results, we focus on distributional transforms from $\mathcal{M}_\infty$ to $\mathcal{M}_\infty$ or $\mathcal{M}_0$ to $\mathcal{M}_0$; However, conceptually, distributional transforms can be defined on any subset of $\mathcal{M}_0$. For simplicity, we identify a distribution measure in $\mathcal{M}_0$ with its cumulative distribution function (cdf), i.e. for $F \in \mathcal{M}_0$, $F(x)$ should be interpreted as $F((-\infty, x])$ for $x \in \mathbb{R}$. For a cdf F, we define its (left) quantile as

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geqslant t\}, \quad t \in [0, 1], \tag{1}$$

and its right quantile as

$$F^{-1}_+(t) = \inf\{x \in \mathbb{R} : F(x) > t\}, \quad t \in [0, 1],$$

with the convention $\inf \varnothing = \infty$.

We denote by μ the mean functional; that is, $\mu(F)$ is the mean of F for $F \in \mathcal{M}_1$. For a random variable X and a distribution F, we write $X \sim F$ if the distribution of X is F. The most important class of distributional transforms in this paper is the class of probability distortions, defined as follows. Let $\mathcal{H}_0$ denote the collection of all increasing and right-continuous functions $g : [0,1] \to [0,1]$ with $g(0) = 0$ and $g(1) = 1$.

Definition 1. For $g \in \mathcal{H}_0$, the *probability distortion generated by* g, denoted by $T_g : \mathcal{M} \to \mathcal{M}_0$, is defined as $T_g(F) = g \circ F$.

In all the following examples, we let U be a generic random variable with uniform distribution on $[0,1]$. Note that if $g \in \mathcal{H}_0$ then $g \circ F \in \mathcal{M}_\infty$ for any $F \in \mathcal{M}_\infty$. Thus, T_g is a well defined distributional transform on $\mathcal{M}_\infty$ for any $g \in \mathcal{H}_0$. However, for $F \in \mathcal{M}_0$, $g \circ F$ may not be in $\mathcal{M}_0$, unless we require additionally that $g(1-) = 1$. For an alternative formulation, we can take $h : (0,1) \to (0,1]$ as any increasing function and define $T(F)$ as the law of $F^{-1}(h(U))$ for $F \in \mathcal{M}_\infty$. Then it is easy to see that $T = T_g$ for $g(x) = \mathbb{P}(h(U) \leqslant x),\ x \in [0,1]$. If additionally h does not take the value 1, then T defines to a probability transform on all of $\mathcal{M}_0$.

We look at two simple examples below.

Example 1 (Tail transform)**.** For $p \in [0,1)$, define $T : \mathcal{M}_0 \to \mathcal{M}_0$ via letting

$$T(F)(x) = \frac{(F(x) - p)^+}{1-p},\ x \in \mathbb{R},$$

which is the distribution of $F^{-1}(p + (1-p)U)$. If X is continuously distributed, then $T(F)$ is also the conditional distribution of $X \sim F$ given $\{X > F^{-1}(p)\}$.

Example 2 (Distorted power transform)**.** For $\gamma > 0$, define $T : \mathcal{M}_0 \to \mathcal{M}_0$ via letting

$$T(F)(x) = (F(x))^{1/\gamma},\ x \in \mathbb{R},$$

which is the distribution of $F^{-1}(U^\gamma)$. If $\gamma = 1$, then $T(F) = F$.

There are many other commonly used distributional transforms in the literature for different applications. We will see some of them in the following examples.

Example 3 (Scale-location transform)**.** For $a \in \mathbb{R}_+$ and $b \in \mathbb{R}$, define $T_{a,b} : \mathcal{M}_0 \to \mathcal{M}_0$ by letting $T_{a,b}(F)$ be the distribution of $aX + b$, where $X \sim F$.

Example 4 (Super-quantile transform)**.** For $F \in \mathcal{M}_1$, its super-quantile transform $T(F)$ is defined as the cdf $T(F) = G$ with quantile function given by

$$G^{-1}(p) = \frac{1}{1-p} \int_p^1 F^{-1}(t) \mathrm{d}t, \qquad p \in (0,1).$$

Example 5 (Shape transform)**.** For a strictly increasing and continuous function ϕ on $\mathbb{R}$ satisfying $\phi(\mathbb{R}) = \mathbb{R}$, define $T^{[\phi]} : \mathcal{M} \to \mathcal{M}_0$ via letting $T^{[\phi]}(F) = F \circ \phi^{-1}$. In other words, $T^{[\phi]}(F)$ is the distribution of $\phi(X)$, where $X \sim F$.

Example 6 (Lorenz curve)**.** Motivated by the Lorenz curve introduced by Lorenz (1905), the Lorenz transform of $F \in \mathcal{M}_1$ supported on $\mathbb{R}_+$ is defined by (see e.g. Gastwirth (1971))

$$L(F)(p) = \frac{\int_0^p F^{-1}(t)\mathrm{d}t}{\mu(F)}, \quad p \in [0,1].$$

In economics, F represents the wealth distribution of the population and $L(F)(p)$ is the percentage of total wealth for the poorest $100p\%$ people. The curve $L(F)$ is used in economics to reflect the inequality of the wealth distribution. For instance, the *Gini coefficient* is twice the area of the difference between the curve $L(F)$ and the diagonal line on $[0,1]$; note that the diagonal line represents the case of equal wealth.

Example 7 (Convolution transform)**.** For a fixed $G \in \mathcal{M}_0$, the convolution transform of $F \in \mathcal{M}_0$ is given by

$$T(F)(x) = \int F(x-y)\, G(\mathrm{d}y), \qquad x \in \mathbb{R}.$$

The transform T can be interpreted as a perturbation to a distribution with an independent noise distributed according to G.

Example 8 (Weighted transform)**.** Let $\mathcal{M}$ be the set of distributions with densities and $w : \mathbb{R} \to [0,1]$ be a measurable function (called the weight function). The weighted transform T of a distribution $F \in \mathcal{M}$ with density f is the distribution $T(F)$ with density

$$\frac{w(x)f(x)}{\int_{\mathbb{R}} w(y)f(y)\mathrm{d}y}, \quad x \in \mathbb{R},$$

provided that $\int_{\mathbb{R}} w(y)f(y)\mathrm{d}y > 0$. The weighted transform is commonly used in statistical forecasting; see Gneiting (2011), Holzmann and Klar (2017), and Lerch et al. (2017).

The above examples will be referred to throughout the paper, and some of them share some common features, in the sense that they satisfy several technical properties which are natural in different applications. These properties will help us later to characterize special subclasses of distributional transforms. Let δ_x denote the point mass at x for $x \in \mathbb{R}$. In what follows, we say that $\mathcal{M}$ is closed under T if T maps $\mathcal{M}$ to $\mathcal{M}$. Recall that $F \in \mathcal{M}_0$ is said to be smaller than $G \in \mathcal{M}_0$ in *stochastic order*, denoted by $F \leqslant_{\mathrm{st}} G$, if for all bounded increasing function f, we have $\int_{\mathbb{R}} f(x)\mathrm{d}F(x) \leqslant \int_{\mathbb{R}} f(x)\mathrm{d}G(x)$. Note that $F \leqslant_{\mathrm{st}} G$ if and only if $F(x) \geqslant G(x)$ for all $x \in \mathbb{R}$.

Definition 2. For a distributional transform T, we use the following terminologies, where $T^{[\phi]}$ is defined in Example 5. T is *monotone* if $T(F) \leqslant_{\mathrm{st}} T(G)$ for $F \leqslant_{\mathrm{st}} G$, $F, G \in \mathcal{M}$; T is *constant-preserving* if $T(\delta_x) = \delta_x$ for $x \in \mathbb{R}$; T is *lower-semicontinuous* if $T(F) \leqslant_{\mathrm{st}} G \in \mathcal{M}$ whenever

$\{F_n\}_{n\in\mathbb{N}} \subset \mathcal{M}$ converges weakly to $F \in \mathcal{M}$ and satisfies $T(F_n) \leqslant_{\mathrm{st}} G$; T is *commuting with shape transforms* if T and $T^{[\phi]}$ commute, i.e., for all strictly increasing and continuous function ϕ satisfying $\phi(\mathbb{R}) = \mathbb{R}$, $T \circ T^{[\phi]} = T^{[\phi]} \circ T$ and $\mathcal{M}$ is closed under all $T^{[\phi]}$.

With the exception of the Lorenz transform (Example 6) and the weighted transform (Example 8), all distributional transforms defined in Examples 1 through 8 are monotone; Examples 1, 2, and 4 are constant-preserving, while Examples 3, 5, 6 and 7 are generally not (Example 8 is defined only on the set of absolutely continuous distributions); Examples 1-7 are all lower-semicontinuous, whereas Examples 8 is generally not; Examples 1 and 2 are commuting with shape transforms, while Examples 3, 4, 5, 6, 7 and 8 are not commuting with shape transforms.

3 Characterization of distributional transforms

In this section we study a characterization of distributional transforms $T : \mathcal{M} \to \mathcal{M}$ with $\mathcal{M} = \mathcal{M}_0$ and $\mathcal{M} = \mathcal{M}_\infty$, and this leads to two theorems. First, given the importance of probability distortions in economics, finance and other fields, we give an equivalent set of conditions which characterize a probability distortion.

Theorem 1. *Let $T : \mathcal{M} \to \mathcal{M}$ where $\mathcal{M} = \mathcal{M}_0$ or $\mathcal{M} = \mathcal{M}_\infty$. The distributional transform T is lower-semicontinuous, monotone and commuting with shape transforms if and only if it is a probability distortion.*

Proof. We first prove the "if" part. Let T_g be a probability distortion for some $g \in \mathcal{H}_0$. It is obvious that T_g is monotone. To prove that T_g is commuting with shape transforms, note that for any $F \in \mathcal{M}$ and for any continuous and strictly increasing function $\phi : \mathbb{R} \to \mathbb{R}$ with $\phi(\mathbb{R}) = \mathbb{R}$,

$$T_g \circ T^{[\phi]}(F) = g(F \circ \phi^{-1}) = g(F) \circ \phi^{-1} = T^{[\phi]} \circ T_g(F).$$

To show the lower-semicontinuity of T_g, let G and $\{F_n\}_{n\in\mathbb{N}}$ be as in Definition 2 for lower-semicontinuity. If x is a continuous point of F, then

$$T_g(F)(x) = g(F(x)) \geqslant \limsup_{n\to\infty} g(F_n(x)) \geqslant G(x).$$

Hence, by the fact that continuous points of F are dense, and cdfs are right continuous, we have

$$T_g(F)(x) \geqslant G(x), \ x \in \mathbb{R}.$$

Next, we show the "only if" part. For $p \in (0,1)$, let $q_p : \mathcal{M} \to \mathbb{R}$ be the p-quantile of a distribution defined in (1). Further, define $q_p^T : \mathcal{M} \to \mathbb{R}$ by $q_p^T(F) = q_p(T(F))$. For each $p \in (0,1)$, we claim that the functional q_p^T satisfies

(i) $q_p^T(F) \leqslant q_p^T(G)$ if $F \leqslant_{\mathrm{st}} G$;

(ii) for any strictly increasing and continuous function ϕ on $\mathbb{R}$ satisfying $\phi(\mathbb{R}) = \mathbb{R}$, we have $q_p^T(T^{[\phi]}(F)) = \phi(q_p^T(F))$.

Among the above properties, (i) follows directly from the monotonicity of T. (ii) follows from the fact that T is commuting with shape transforms, and noting that the quantile function (both left and right) commutes with shape transforms, we have

$$q_p^T(T^{[\phi]}(F)) = q_p(T \circ T^{[\phi]}(F)) = q_p(T^{[\phi]} \circ T(F)) = \phi(q_p(T(F))) = \phi(q_p^T(F)).$$

We first focus on $\mathcal{M} = \mathcal{M}_\infty$. Theorem 1 of Chambers (2009) (noting also the footnote in Theorem 1 of Chambers (2009)) states that, with the above properties (i)-(ii), q_p^T is a (left or right) quantile of F at a fixed level for all $F \in \mathcal{M}$. We denote this level by $h(p) \in [0,1]$. Hence for any $F \in \mathcal{M}$

$$q_p^T(F) = \begin{cases} F^{-1}(h(p)), & p \in E, \\ F_+^{-1}(h(p)), & p \in (0,1) \setminus E, \end{cases} \tag{2}$$

where $E \subset (0,1)$ is independent of F. Note that by definition if $h(p) = 1$, $p \in E$; if $h(p) = 0$, $p \notin E$. Clearly, $h(p)$ is an increasing function over $(0,1)$. We next show that $E = (0,1)$. To this end, we assume by way of contradiction that there is $p_0 \in (0,1) \setminus E$. Notice that $h(p_0) \neq 1$. Let $F_n \sim \text{Bernoulli}\left((1-h(p_0))\frac{n-1}{n}\right)$ and $F \sim \text{Bernoulli}(1-h(p_0))$. It is clear that F_n converges weakly to F and

$$g(p) := \sup_{n \geqslant 1} q_p^T(F_n) \leqslant 1, \ p \in (0,1),$$

where $g(p)$ is increasing over $(0,1)$. Let

$$G(x) = \mathbb{P}\left(g(U) \leqslant x\right), \ x \in \mathbb{R}.$$

The boundness of g indicates that G belongs to $\mathcal{M}$. Hence

$$T(F_n) \leqslant_{\mathrm{st}} G, \ n \geqslant 1.$$

Lower-semicontinuity leads to

$$T(F) \leqslant_{\mathrm{st}} G.$$

Noting that g is a quantile function of G, we have

$$q_p^T(F) \leqslant G^{-1}(p) \leqslant g(p) = \sup_{n \geqslant 1} q_p^T(F_n), \quad p \in (0,1). \tag{3}$$

However, it follows from (2) that

$$q_{p_0}^T(F) = 1 \ \text{ and } \ \sup_{n \geqslant 1} q_{p_0}^T(F_n) = 0,$$

which contradicts with (3). Hence, $E = (0,1)$ and this implies

$$q_p^T(F) = F^{-1}(h(p)),\ p \in (0,1), \tag{4}$$

where $h : (0,1) \to (0,1]$ is an increasing function. This proves that T is a probability distortion in the case $\mathcal{M} = \mathcal{M}_\infty$.

We next focus on $\mathcal{M} = \mathcal{M}_0$. For $F \in \mathcal{M}_\infty$, we have shown (4). We next extend (4) from $\mathcal{M}_\infty$ to $\mathcal{M}_0$. To this end, we split $\mathcal{M}_0 = \mathcal{M}_\infty \cup \mathcal{M}_- \cup \mathcal{M}_+ \cup \mathcal{M}_\pm$, where

$$\mathcal{M}_- := \{F \in \mathcal{M}_0 :\ \text{the support of } F \text{ has infinite lower bound and finite upper bound}\},$$
$$\mathcal{M}_+ := \{F \in \mathcal{M}_0 :\ \text{the support of } F \text{ has finite lower bound and infinite upper bound}\},$$
$$\mathcal{M}_\pm := \{F \in \mathcal{M}_0 :\ \text{the support of } F \text{ has both infinite lower and upper bounds}\}.$$

We first consider $F \in \mathcal{M}_-$. Let $F_n(x) = F(x)\mathbb{1}_{\{x \geqslant -n\}},\ n \geqslant 1$. Then $F_n \in \mathcal{M}_\infty$ and $F \leqslant_{\rm st} F_n$. Hence we have

$$q_p^T(F) \leqslant q_p^T(F_n) = F_n^{-1}(h(p)),\ p \in (0,1).$$

Note that $h(p) > 0$ for $p \in (0,1)$. Taking the limit with respect to n in the above inequality yields

$$q_p^T(F) \leqslant \lim_{n\to\infty} F_n^{-1}(h(p)) = F^{-1}(h(p)),\ p \in (0,1). \tag{5}$$

For any $x_0 \in \mathbb{R}$, define

$$\phi_n(x) = \begin{cases} x_0 + \frac{x - x_0}{n}, & x < x_0, \\ x, & x \geqslant x_0, \end{cases} \quad n \geqslant 1, \quad \phi_0(x) = x \wedge x_0,$$

Note that $\lim_{n\to\infty} \phi_n(x) = \phi_0(x),\ x \in \mathbb{R}$. Thus we have $F \circ \phi_n^{-1} \in \mathcal{M}_-$ and $F \circ \phi_n^{-1}$ converges weakly to $F_0(x) := F(x)\mathbb{1}_{\{x \geqslant x_0\}} \in \mathcal{M}_\infty$. By the property (ii), we have

$$q_p^T(F \circ \phi_n^{-1}) = \phi_n(q_p^T(F)) \leqslant \phi_0(q_p^T(F)),\ p \in (0,1),\ n \geqslant 1.$$

By lower-semicontinuity and (4) we obtain for $p \in (0,1)$

$$q_p^T(F_0) \leqslant \phi_0(q_p^T(F)),\ q_p^T(F_0) = F_0^{-1}(h(p)).$$

For $x_0 < q_p^T(F)$,

$$F^{-1}(h(p)) = F_0^{-1}(h(p)) = q_p^T(F_0) \leqslant \phi_0(q_p^T(F)) = q_p^T(F).$$

Together with (5) we arrive at

$$q_p^T(F) = F^{-1}(h(p)),\ p \in (0,1).$$

Next, we focus on $F \in \mathcal{M}_+$. Let $F_n(x) = F(x)\mathbb{1}_{\{x<n\}} + \mathbb{1}_{\{x\geqslant n\}},\ n \geqslant 1$. Then $F_n \in \mathcal{M}_\infty$ and $F_n \leqslant_{\mathrm{st}} F$. Hence we have

$$q_p^T(F) \geqslant q_p^T(F_n) = F_n^{-1}(h(p)),\ p \in (0,1).$$

Letting $n \to \infty$ in the above inequality yields

$$q_p^T(F) \geqslant F^{-1}(h(p)),\ p \in (0,1). \tag{6}$$

Note that $h(p) < 1$ for $p \in (0,1)$. Or else $q_p^T(F) = \infty$ for some $p \in (0,1)$. Define

$$\hat{F}(x) = \begin{cases} F(x), & x \leqslant 0, \\ F(\tan x), & 0 < x < \pi/2, \\ 1, & x \geqslant \pi/2. \end{cases}$$

Moreover, for $n \in \mathbb{N}$, let

$$\hat{\phi}_n(x) = \begin{cases} x, & x \leqslant 0, \\ \tan x, & 0 < x \leqslant \frac{\pi}{2} - \frac{1}{n+1}, \\ x - n + \tan\left(\frac{\pi}{2} - \frac{1}{n+1}\right), & x > n, \end{cases}$$

and

$$\hat{\phi}_0(x) = \begin{cases} x, & x \leqslant 0, \\ \tan x, & 0 < x < \frac{\pi}{2}. \end{cases}$$

We note that $\hat{F} \in \mathcal{M}_\infty$, $\hat{F} \circ \hat{\phi}_n^{-1} \in \mathcal{M}_\infty$ and $\hat{F} \circ \hat{\phi}_n^{-1}$ converges weakly to F. In addition, noting that $q_p^T(\hat{F}) = \hat{F}^{-1}(h(p)) < \pi/2,\ p \in (0,1)$, we have, for large n,

$$q_p^T(\hat{F} \circ \hat{\phi}_n^{-1}) = \hat{\phi}_n(q_p^T(\hat{F})) \leqslant \hat{\phi}_0(q_p^T(\hat{F})).$$

Lower-semicontinuity leads to

$$q_p^T(F) \leqslant \hat{\phi}_0(q_p^T(\hat{F})) = \hat{\phi}_0(\hat{F}^{-1}(h(p))) = F^{-1}(h(p)),\ p \in (0,1).$$

Together with (6) we arrive at

$$q_p^T(F) = F^{-1}(h(p)),\ p \in (0,1).$$

Finally, we focus on $F \in \mathcal{M}_\pm$. Let $F_n(x) = F(x)\mathbb{1}_{\{x<n\}} + \mathbb{1}_{\{x\geqslant n\}},\ n \geqslant 1$, and $G_n = F(x)\mathbb{1}_{\{x\geqslant -n\}},\ n \geqslant 1$. Then $F_n \in \mathcal{M}_-$ and $G_n \in \mathcal{M}_+$. Monotonicity indicates

$$F_n^{-1}(h(p)) = q_p^T(F_n) \leqslant q_p^T(F) \leqslant q_p^T(G_n) = G_n^{-1}(h(p)),\ p \in (0,1).$$

Letting $n \to \infty$, we obtain

$$q_p^T(F) = F^{-1}(h(p)),\ p \in (0,1).$$

Therefore, we conclude that there exists an increasing function $h : (0,1) \to (0,1)$ such that for any $F \in \mathcal{M}_0$

$$q_p^T(F) = F^{-1}(h(p)),\ p \in (0,1).$$

This proves the case $\mathcal{M} = \mathcal{M}_0$. The proof is complete. □

It turns out from the proof of Theorem 1 that lower-semicontinuity plays an important role in showing that distributional transforms are probability distortions. If we relax the assumption of lower-semicontinuity in Theorem 1, the distributional transforms are not probability distortions. Below are two examples.

Example 9. 1. For $i = 1,2,3,4$, take $g_i \in \mathcal{H}_0$ with $g_i(1-) = 1$ such that $g_1 \geqslant g_2 \geqslant g_4$ and $g_1 \geqslant g_3 \geqslant g_4$. Define $T : \mathcal{M}_0 \to \mathcal{M}_0$ by

$$T(F) = \begin{cases} g_1 \circ F, & F \in \mathcal{M}_-, \\ g_2 \circ F, & F \in \mathcal{M}_\infty, \\ g_3 \circ F, & F \in \mathcal{M}_\pm, \\ g_4 \circ F, & F \in \mathcal{M}_+. \end{cases}$$

One can check similarly to the proof of Theorem 1 that T satisfies monotonicity and commutes with shape transforms. However, T is not a probability distortion unless $g_1 = g_2 = g_3 = g_4$.

2. Define $T : \mathcal{M} \to \mathcal{M}$ with $\mathcal{M} = \mathcal{M}_0$ or $\mathcal{M} = \mathcal{M}_\infty$ as $(T(F))^{-1}(p) = F_+^{-1}(1/2),\ p \in (0,1)$. One can easily check that T is monotone and commuting with shape transforms. However, T is not a probability distortion since it does not satisfy lower-semicontinuity.

The next result characterizes additive and monotone distributional transforms. Let $F \oplus G$ be the distribution with quantile function $F^{-1} + G^{-1}$. A distributional transform T is called *additive* if $T(F) \oplus T(G) = T(F \oplus G)$ for any $F, G \in \mathcal{M}$.

First, we note that all probability distortions T_g are monotone and additive. Indeed, note that for a.e. $p \in (0,1)$ and all $F \in \mathcal{M}$,

$$(T_g(F \oplus G))^{-1}(p) = (F^{-1} + G^{-1})(g^{-1}(p)) = F^{-1}(g^{-1}(p)) + G^{-1}(g^{-1}(p)) = (T_g(F) \oplus T_g(G))^{-1}(p).$$

This implies that $T_g(F \oplus G) = T_g(F) \oplus T_g(G)$. However, probability distortions are not the only distributional transforms having these two properties. For instance, the distributional transform $T : \mathcal{M}_\infty \to \mathcal{M}_\infty$ given by $T(F) = \delta_{\mu(F)}$ is also monotone and additive, but it is not a probability distortion. As another example, the super-quantile transform in Example 4 is also monotone and additive.

To characterize additive distributional transforms, we focus on the case $\mathcal{M} = \mathcal{M}_\infty$ for technical convenience; see Remark 1 below for the case $\mathcal{M} = \mathcal{M}_0$. Let $\mathcal{H}$ be the set of increasing functions h on $[0,1]$ with $h(0) = 0$ and $h(1) = 1$. For $h \in \mathcal{H}$, define the *Choquet integral* $I_h : \mathcal{M} \to \mathbb{R}$ as

$$I_h(F) = \int_{-\infty}^{0} (h(1 - F(x)) - 1)\mathrm{d}x + \int_{0}^{\infty} h(1 - F(x))\mathrm{d}x. \tag{7}$$

Theorem 2. *A distributional transform $T : \mathcal{M}_\infty \to \mathcal{M}_\infty$ is monotone, constant-preserving and additive if and only if it has the following form: there exists a point-wise increasing family of functions $h_p \in \mathcal{H}$, $p \in (0,1)$ such that for $F \in \mathcal{M}_\infty$, $T(F)$ is uniquely determined by*

$$(T(F))^{-1}(p) = I_{h_p}(F) \quad \textit{for} \quad \textit{a.e.} \ p \in (0,1). \tag{8}$$

Proof. "$\Rightarrow$": Note that for $p \in (0,1)$, $F \mapsto (T(F))^{-1}(p)$ is a functional from $\mathcal{M}_\infty$ to $\mathbb{R}$. Monotonicity and additivity of T imply the monotonicity and additivity of this functional. Moreover, constant-preservation indicates $(T(\delta_1))^{-1}(p) = 1$. By the main result of Schmeidler (1986) (see e.g. Theorem 1 of Wang et al. (2019) for an exact form), there exists $h_p \in \mathcal{H}$ such that

$$(T(F))^{-1}(p) = I_{h_p}(F).$$

For $0 < p_1 < p_2 < 1$, we have for any $F \in \mathcal{M}_\infty$

$$I_{h_{p_1}}(F) = (T(F))^{-1}(p_1) \leqslant (T(F))^{-1}(p_2) = I_{h_{p_2}}(F).$$

Taking $F(x) = (1-t)\mathbb{1}_{\{0 \leqslant x < 1\}} + \mathbb{1}_{\{x \geqslant 1\}}$ for some $t \in [0,1]$ yields that

$$h_{p_1}(t) = I_{h_{p_1}}(F) \leqslant I_{h_{p_2}}(F) = h_{p_2}(t), \tag{9}$$

giving the monotonicity of h_p with respect to p.

"$\Leftarrow$": A direct calculation shows that for $F \leqslant_{\mathrm{st}} G$, $F, G \in \mathcal{M}_\infty$,

$$I_{h_p}(F) \leqslant I_{h_p}(G), \ p \in (0,1).$$

This implies that

$$(T(F))^{-1}(p) \leqslant (T(G))^{-1}(p), \ p \in (0,1),$$

which means that T is monotone. It is well known that the Choquet integral is comonotonically additive. Hence, for $F, G \in \mathcal{M}_\infty$,

$$I_{h_p}(F \oplus G) = I_{h_p}(F) + I_{h_p}(G), \ p \in (0,1).$$

It follows that

$$(T(F \oplus G))^{-1}(p) = (T(F))^{-1}(p) + (T(G))^{-1}(p) = (T(F) \oplus T(G))^{-1}(p), \ p \in (0,1).$$

The additivity of T is obtained. Moreover, a direct calculation shows

$$(T(\delta_x))^{-1}(p) = I_{h_p}(\delta_x) = x,\ p \in (0,1),$$

implying that $T(\delta_x) = \delta_x$. Hence T is constant-preserving. This completes the proof. □

Remark 1. To generalize Theorem 2 to distributional transforms on $\mathcal{M}_0$, one needs some additional assumptions. This is because the integral in (7) may not be finite or well-defined for all $F \in \mathcal{M}_0$. To guarantee the finiteness of (7), we may replace $\mathcal{H}$ by the set of all increasing functions h on $[0,1]$ for which there exists $0 < \varepsilon < 1/2$ such that $h(x) = 0$ for $x \in [0,\varepsilon]$ and $h(x) = 1$ for $x \in [1-\varepsilon, 1]$. In addition, we need to assume that $\lim_{n\to\infty} T(F_n) = T(\delta_0)$ for any sequences $\{F_n\}_{n\in\mathbb{N}}$ converging weakly to δ_0. Under these additional assumptions, we can generalize the representation (8) to $\mathcal{M} = \mathcal{M}_0$.

We shall see in what follows that Theorem 2 is in fact very general including many interesting examples.

Example 10. (i) If $h_p(t) = \mathbb{1}_{\{t>1-p\}}$, then

$$I_{h_p}(F) = \int_{-\infty}^{0} \left(\mathbb{1}_{\{F(x)<p\}} - 1\right) \mathrm{d}x + \int_{0}^{\infty} \mathbb{1}_{\{F(x)<p\}} \mathrm{d}x = F^{-1}(p),\ p \in (0,1)$$

and thus T is the identity.

(ii) If $h_p(t) = t$, then I_{h_p} is the mean, and $T(F) = \delta_{\mu(F)}$.

(iii) If $h_p(t) = (\frac{1}{1-p}t) \wedge 1$, then

$$I_{h_p}(F) = \frac{1}{1-p}\int_{p}^{1} F^{-1}(t)\mathrm{d}t,$$

and T is the super-quantile transform in Example 4.

(iv) If $h_p(t) = \mathbb{1}_{\{t>1-g(p)\}}$ for an increasing function $g : (0,1) \to (0,1]$, then

$$I_{h_p}(F) = \int_{-\infty}^{0} \left(\mathbb{1}_{\{F(x)<g(p)\}} - 1\right) \mathrm{d}x + \int_{0}^{\infty} \mathbb{1}_{\{F(x)<g(p)\}} \mathrm{d}x = F^{-1}(g(p)),\ p \in (0,1)$$

is the left $g(p)$-quantile, and T is a probability distortion generated by

$$G(x) = \mathbb{P}\left(g(U) \leqslant x\right),\ x \in [0,1].$$

(v) If $h_p(t) = \mathbb{1}_{\{t\geqslant 1-g(p)\}}$ for an increasing function $g : (0,1) \to [0,1)$, then

$$I_{h_p}(F) = \int_{-\infty}^{0} \left(\mathbb{1}_{\{F(x)\leqslant g(p)\}} - 1\right) \mathrm{d}x + \int_{0}^{\infty} \mathbb{1}_{\{F(x)\leqslant g(p)\}} \mathrm{d}x = F_{+}^{-1}(g(p)),\ p \in (0,1)$$

is the right $g(p)$-quantile, and

$$G(F(x)-) \leqslant T(F)(x) \leqslant G(F(x)), \quad \text{where} \quad G(x) = \mathbb{P}\left(g(U) \leqslant x\right),\ x \in [0,1].$$

4 Further properties

In this section we study several other useful properties for commonly used distributional transforms, including the examples in Section 2. Recall that $F \in \mathcal{M}_1$ is smaller than $G \in \mathcal{M}_1$ in *increasing convex order*, denoted by $F \leqslant_{\mathrm{icx}} G$, if for all increasing convex function f, we have $\int_{\mathbb{R}} f(x)\mathrm{d}F(x) \leqslant \int_{\mathbb{R}} f(x)\mathrm{d}G(x)$.

Definition 3. T is *scale-location-invariant* if for all $a \in \mathbb{R}_+$ and $b \in \mathbb{R}$, $T_{a,b} \circ T = T \circ T_{a,b}$ and $\mathcal{M}$ is closed under $T_{a,b}$; T is $\leqslant_{\mathrm{icx}}$-*monotone* if $T(F) \leqslant_{\mathrm{icx}} T(G)$ for any $F \leqslant_{\mathrm{icx}} G$ with $F, G \in \mathcal{M} \subset \mathcal{M}_1$; T is *ampliative* if $T(F) \geqslant_{\mathrm{st}} F$ for all $F \in \mathcal{M}$.

Among examples in Section 2, Example 1, Example 2 (if $\gamma \leqslant 1$), Examples 3 and 4, Example 6 (if $\mu(F) = \mu(G)$) and Example 7 are $\leqslant_{\mathrm{icx}}$-monotone, whereas Examples 5 and 8 (for general w) are not $\leqslant_{\mathrm{icx}}$-monotone; Examples 1, 2, and 4 are scale-location-invariant, whereas Examples 3, 5, 6, 7 and 8 are not scale-location-invariant; Example 1, Example 2 (if $\gamma \leqslant 1$), Example 3 (if $a = 1$, $b \geqslant 0$), Example 4 and Example 5 (if $\phi(x) \geqslant x,\ x \in \mathbb{R}$) are ampliative whereas Example 6, Example 7 (for general G) and Example 8 (for general w) are not ampliative.

Some of the above properties imply other properties used in the paper. For instance, if T is scale-location-invariant, then T is constant-preserving. Note that $T_{0,x}(F) = \delta_x$ for all $x \in \mathbb{R}$ and $F \in \mathcal{M}_0$. Hence

$$T(\delta_x) = T \circ T_{0,x}(\delta_0) = T_{0,x} \circ T(\delta_0) = \delta_x.$$

On the other hand, if T is constant-preserving and commuting with shape transforms, then T is scale-location-invariant. To see this, note that $T_{a,b}, a > 0$ are shape transforms, which indicates that $T \circ T_{a,b} = T_{a,b} \circ T,\ a > 0$. Moreover, constant preservation guarantees that $T \circ T_{0,b} = T_{0,b} \circ T$. The next proposition shows that monotonicity and commutation with shape transform imply scale-location-invariance and additivity.

Proposition 1. *If a distributional transform $T : \mathcal{M}_\infty \to \mathcal{M}_\infty$ is monotone and commuting with shape transforms, then it is scale-location-invariant and additive.*

Proof. We first show that monotonicity and commutation with shape transforms imply additivity. To see this, we use (2) to represent $T(F)$ in the form

$$(T(F))^{-1}(p) = \begin{cases} F^{-1}(g(p)), & p \in E, \\ F_+^{-1}(g(p)), & p \in (0,1) \setminus E, \end{cases}$$

for some set $E \subset (0,1)$ and an increasing function $g : (0,1) \to [0,1]$. It follows from (iv) and (v) of Example 10 that the right-hand side of above equation can also be represented by the

Choquet integrals $I_{h_p}(F)$, where

$$h_p(t) = \begin{cases} \mathbb{1}_{\{t>1-g(p)\}}, & p \in E, \\ \mathbb{1}_{\{t\geqslant 1-g(p)\}}, & p \in (0,1) \setminus E. \end{cases}$$

Hence, it follows from Theorem 2 that T is additive.

Now we show that monotonicity and commutation with shape also imply scale-location-invariance. To see this, let $q_p^T(F) = (T(F))^{-1}(p),\ p \in E$ and we have

$$q_p^{T_{a,b}\circ T}(F) = aq_p^T(F) + b = aF^{-1}(h(p)) + b = q_p^T(T_{a,b}(F)) = q_p^{T\circ T_{a,b}}(F).$$

For $p \in (0,1) \setminus E$, the analogous identity follows by a similar argument. This implies that T is scale-location-invariant. □

Proposition 2. *Let $T_g : \mathcal{M}_\infty \to \mathcal{M}_\infty$ be a probability distortion. We have (i) T_g is ampliative if and only if $g(t) \leqslant t,\ t \in [0,1]$; (ii) T_g is $\leqslant_{\mathrm{icx}}$-monotone if and only if g is convex.*

Proof. (i): Sufficiency is obvious. To show necessity, let F be the uniform distribution on $[0,1]$. Then we have

$$g(x) = g(F(x)) \leqslant F(x) = x,\ x \in [0,1].$$

(ii): By Corollary 2 of Chew et al. (1987), g is convex if and only if for all $F \leqslant_{\mathrm{icx}} G$ and all strictly increasing and convex function u over $\mathbb{R}$, we have

$$\int_{\mathbb{R}} u(x)\mathrm{d}g(F(x)) \leqslant \int_{\mathbb{R}} u(x)\mathrm{d}g(G(x)).$$

This proves (ii). □

From the proof, it is clear that (i) of Proposition 2 remains valid for $T_g : \mathcal{M}_0 \to \mathcal{M}_0$. Let $\mathcal{T}$ be the set of distributional transforms, and $\mathcal{T}_S$ be the set of monotone and scale-location-invariant distributional transforms. We define the scale multiplication $\otimes$ as $a \otimes F = T_{a,0}(F)$ for $a \in \mathbb{R}_+$ and $F \in \mathcal{M}$. Below we give simple algebraic properties of $\mathcal{T}$ and $\mathcal{M}$ which can be checked in a straightforward manner.

Proposition 3. *(i) $(\mathcal{T}, \circ)$ is a monoid with the identity element $T_{1,0}$;*

(ii) $(\mathcal{T}_S, \circ)$ is a submonoid of $(\mathcal{T}, \circ)$;

(iii) $(\mathcal{M}, \oplus)$ is a monoid with the identity element δ_0;

(iv) $\mathcal{M}$ equipped with $(\oplus, \otimes)$ is a convex cone.

Next, we state some forms of additivity of distributional transforms, which will be useful for the study of risk measures generated by distributional transforms in Section 5. For a distributional transform T:

(i) T is $\leqslant_{\mathrm{st}}$-*subadditive* (resp. $\leqslant_{\mathrm{st}}$-*supperadditive*) if $T(F \oplus G) \leqslant_{\mathrm{st}} T(F) \oplus T(G)$ (resp. $T(F \oplus G) \geqslant_{\mathrm{st}} T(F) \oplus T(G)$) for all $F, G \in \mathcal{M}$;

(ii) T is $\leqslant_{\mathrm{icx}}$-*subadditive* (resp. $\leqslant_{\mathrm{icx}}$-*supperadditive*) if $T(F \oplus G) \leqslant_{\mathrm{icx}} T(F) \oplus T(G)$ (resp. $T(F \oplus G) \geqslant_{\mathrm{icx}} T(F) \oplus T(G)$) for all $F, G \in \mathcal{M}$.

Remark 2. (i) Clearly, additivity implies $\leqslant_{\mathrm{st}}$-subadditivity, which in turn implies $\leqslant_{\mathrm{icx}}$-subadditivity. (ii) It is easy to see that the following statements are equivalent: T is additive; T is both $\leqslant_{\mathrm{st}}$-subadditive and $\leqslant_{\mathrm{st}}$-superadditive; T is both $\leqslant_{\mathrm{icx}}$-subadditive and $\leqslant_{\mathrm{icx}}$-superadditive. (iii) Let $T_{a,b}$ be the scale-location transform in Example 3. Then, it is easy to verify that $T_{1,a}$ is additive if and only if $a = 0$, it is $\leqslant_{\mathrm{st}}$-subadditive if and only if $a \geqslant 0$, and it is $\leqslant_{\mathrm{st}}$-supperadditive if and only if $a \leqslant 0$.

5 Distributional transforms and risk measures

5.1 Risk measures

In this section we consider distributional transforms $T : \mathcal{M} \to \mathcal{M}$ with $\mathcal{M} = \mathcal{M}_\infty$ and its relation to risk measures. A *risk measure* ρ is a mapping from L^∞ to $\mathbb{R}$. We use the sign convention that a positive value of a random variable X represents a random loss faced by a financial institution and $\rho(X)$ is the capital charge for bearing the potential loss X. Some properties of risk measures are listed below; for their interpretations see Föllmer and Schied (2016).

Definition 4. For a risk measure ρ, we use the following terminologies:

(i) ρ is *cash-invariant* if $\rho(X + c) = \rho(X) + c$ for $c \in \mathbb{R}$ and $X \in L^\infty$;

(ii) ρ is *monotone* if $\rho(X) \leqslant \rho(Y)$ for $X, Y \in L^\infty$, $X \leqslant Y$;

(iii) ρ is *positively homogeneous* if $\rho(\lambda X) = \lambda \rho(X)$ for $\lambda \in (0, \infty)$ and $X \in L^\infty$;

(iv) ρ is *convex* if $\rho(\lambda X + (1 - \lambda)Y) \leqslant \lambda \rho(X) + (1 - \lambda)\rho(Y)$ for $X, Y \in L^\infty$ and $\lambda \in [0, 1]$;

(v) ρ is *subadditive* if $\rho(X + Y) \leqslant \rho(X) + \rho(Y)$ for $X, Y \in L^\infty$;

(vi) ρ is *law-invariant* if $\rho(X) = \rho(Y)$ for $X, Y \in L^\infty$, $X \overset{\mathrm{d}}{=} Y$;

(vii) ρ is a *monetary risk measure* if it is monotone and cash-invariant;

(viii) ρ is a *convex risk measure* if it is monetary and convex;

(ix) ρ is a *coherent risk measure* if it is monetary, convex, and positively homogeneous.

For $X \in L^\infty$, we denote its distribution by F_X. Let $\mathcal{R}_{\mathrm{ch}}$ and $\mathcal{R}_{\mathrm{cx}}$ denote the collections of all coherent and convex risk measures defined on L^∞, respectively. For any law-invariant risk measure ρ, define $\widehat{\rho}(F) = \rho(X)$, where $X \sim F$. For any distributional transform T and $X \in L^\infty$, let $\rho_T(X) = \widehat{\rho}(T(F_X))$ and for any probability distortion T_g, denote $\rho_g(X) = \rho_{T_g}(X)$. Two important risk measures are the Value-at-Risk, VaR_p, and Expected Shortfall, ES_p, $p \in (0,1)$, which are respectively given by

$$\mathrm{VaR}_p(X) = F_X^{-1}(p), \text{ and } \mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 F_X^{-1}(t)\mathrm{d}t.$$

See, e.g., McNeil et al. (2015) and Föllmer and Schied (2016) for the definitions, properties and applications of VaR and ES.

5.2 Risk measures generated by distributional transforms

We will use the classic result on convex order (see e.g. Theorem 3.5 of Rüschendorf (2013)), which says that $F_{X+Y} \leqslant_{\mathrm{icx}} F_X \oplus F_Y$ for all $X, Y \in L^\infty$. The following lemma gives the convexity of $\widehat{\rho}$ for a convex risk measure ρ.

Lemma 1. *If ρ is a law-invariant convex risk measure, then $\widehat{\rho}$ is convex on $\mathcal{M}$ equipped with $(\oplus, \otimes)$, i.e. $\widehat{\rho}((\lambda \otimes F) \oplus ((1-\lambda) \otimes G)) \leqslant \lambda\widehat{\rho}(F) + (1-\lambda)\widehat{\rho}(G)$ for all $F, G \in \mathcal{M}$ and $\lambda \in [0,1]$.*

Proof. Let U be a uniform random variable on $[0,1]$. Note that for any $F, G \in \mathcal{M}$, $F^{-1}(U) + G^{-1}(U)$ has the distribution $F \oplus G$. Therefore, for $\lambda \in [0,1]$,

$$\begin{aligned}\widehat{\rho}((\lambda \otimes F) \oplus ((1-\lambda) \otimes G)) &= \rho(\lambda F^{-1}(U) + (1-\lambda)G^{-1}(U)) \\ &\leqslant \lambda\rho(F^{-1}(U)) + (1-\lambda)\rho(G^{-1}(U)) = \lambda\widehat{\rho}(F) + (1-\lambda)\widehat{\rho}(G).\end{aligned}$$

Hence $\widehat{\rho}$ is convex. □

The main result of this section is the following theorem, which connects distributional transforms and risk measures.

Theorem 3. *For a monotone and scale-location invariant distributional transform T, the following statements are equivalent:*

(i) T is $\leqslant_{\mathrm{icx}}$-monotone and $\leqslant_{\mathrm{icx}}$-subadditive;

(ii) $\rho_T \in \mathcal{R}_{\mathrm{cx}}$ for all $\rho \in \mathcal{R}_{\mathrm{cx}}$;

(iii) $\rho_T \in \mathcal{R}_{\mathrm{ch}}$ for all $\rho \in \mathcal{R}_{\mathrm{ch}}$;

(iv) $(\mathrm{ES}_p)_T \in \mathcal{R}_{\mathrm{ch}}$ for all $p \in (0,1)$.

Proof. (i)⇒(ii): For $\lambda \in [0,1]$ and $X, Y \in L^\infty$, note that $F_{\lambda X+(1-\lambda)Y} \leqslant_{\mathrm{icx}} F_{\lambda X} \oplus F_{(1-\lambda)Y}$. Also note that law-invariant convex risk measures are monotone with respect to $\leqslant_{\mathrm{icx}}$. By the above facts, the $\leqslant_{\mathrm{icx}}$-monotonicity and the $\leqslant_{\mathrm{icx}}$-subadditivity of T, and Lemma 1, we have

$$\begin{aligned}
\rho_T(\lambda X + (1-\lambda)Y) &= \widehat{\rho}(T(F_{\lambda X+(1-\lambda)Y})) \\
&\leqslant \widehat{\rho}(T(F_{\lambda X} \oplus F_{(1-\lambda)Y})) \\
&\leqslant \widehat{\rho}(T(F_{\lambda X}) \oplus T(F_{(1-\lambda)Y})) \\
&= \widehat{\rho}((T \circ T_{\lambda,0}(F_X)) \oplus (T \circ T_{1-\lambda,0}(F_Y))) \\
&= \widehat{\rho}((T_{\lambda,0} \circ T(F_X)) \oplus (T_{1-\lambda,0} \circ T(F_Y))) \\
&\leqslant \lambda\widehat{\rho}(T(F_X)) + (1-\lambda)\widehat{\rho}(T(F_Y)) = \lambda\rho_T(X) + (1-\lambda)\rho_T(Y).
\end{aligned}$$

Hence, ρ_T is convex. The other properties are easy to check.

(ii)⇒(iii): From (ii), we know that ρ_T is convex for $\rho \in \mathcal{R}_{\mathrm{ch}}$. It suffices to show that ρ_T is positively homogeneous, which is implied by the argument that, for $\lambda > 0$ and $X \in L^\infty$,

$$\rho_T(\lambda X) = \widehat{\rho}(T \circ T_{\lambda,0}(F_X)) = \widehat{\rho}(T_{\lambda,0} \circ T(F_X)) = \lambda\rho_T(X).$$

(iii)⇒(iv): Obvious since ES_p is a coherent risk measure for all $p \in (0,1)$.

(iv)⇒(i): Let $\rho = \mathrm{ES}_p$, $p \in (0,1)$, and U be a uniform random variable on $[0,1]$. Since ρ_T is a coherent risk measure, we have, for all $F, G \in \mathcal{M}$,

$$\rho_T(F^{-1}(U) + G^{-1}(U)) \leqslant \rho_T(F^{-1}(U)) + \rho_T(G^{-1}(U)).$$

As a consequence,

$$\widehat{\rho}(T(F \oplus G)) \leqslant \widehat{\rho}(T(F)) + \widehat{\rho}(T(G)).$$

By comonotonic-additivity of ES_p, we know that $\widehat{\rho}$ is additive with respect to $\oplus$, and hence

$$\widehat{\rho}(T(F \oplus G)) \leqslant \widehat{\rho}(T(F)) + \widehat{\rho}(T(G)) = \widehat{\rho}(T(F) \oplus T(G)).$$

Since $p \in (0,1)$ is arbitrary, this implies $T(F \oplus G) \leqslant_{\mathrm{icx}} T(F) \oplus T(G)$ (e.g. Theorem 3.A.5 of Shaked and Shanthikumar (2007)).

On the other hand, since ρ_T is a law-invariant coherent risk measure, it is monotone with respect to $\leqslant_{\mathrm{icx}}$. For $F \leqslant_{\mathrm{icx}} G$, we have $\widehat{\rho}_T(F) \leqslant \widehat{\rho}_T(G)$. Since $p \in (0,1)$ is arbitrary, this implies $T(F) \leqslant_{\mathrm{icx}} T(G)$. □

We make the following simple but still interesting observation.

Proposition 4. *A mapping $\rho : L^\infty \to \mathbb{R}$ is a law-invariant coherent risk measure if and only if it is the mean of a distributional transform that is monotone, scale-location-invariant, $\leqslant_{\mathrm{icx}}$-subadditive and $\leqslant_{\mathrm{icx}}$-monotone.*

Proof. The "if" part is given by Theorem 3. For the "only-if" part, define the distributional transform $T : F \mapsto \delta_{\widehat{\rho}(F)}$. Obviously, $\mu(T(F)) = \widehat{\rho}(F)$, and the coherence of ρ implies that T is monotone and scale-location-invariant. Note that for $X, Y \in L^\infty$,

$$T(F_X \oplus F_Y) = \delta_{\rho(X+Y)} \leqslant_{\mathrm{icx}} \delta_{\rho(X)+\rho(Y)} = \delta_{\rho(X)} \oplus \delta_{\rho(Y)} = T(F_X) \oplus T(F_Y).$$

Hence, T is also $\leqslant_{\mathrm{icx}}$-subadditive. The $\leqslant_{\mathrm{icx}}$-monotonicity is similar. □

Remark 3. The above proof shows that the distributional transform $T : F \mapsto \delta_{\widehat{\rho}(F)}$ is even $\leqslant_{\mathrm{st}}$-subadditive, a stronger property than $\leqslant_{\mathrm{icx}}$-subadditivity.

Next, we present the relationship between probability distortions and risk measures.

Proposition 5. *Suppose T_g is a probability distortion generated by $g \in \mathcal{H}_0$. Then the following statements are equivalent:*

(i) g is convex;

(ii) $\rho_g \in \mathcal{R}_{\mathrm{cx}}$ for all $\rho \in \mathcal{R}_{\mathrm{cx}}$;

(iii) $\rho_g \in \mathcal{R}_{\mathrm{ch}}$ for all $\rho \in \mathcal{R}_{\mathrm{ch}}$;

(iv) $\mathbb{E}_g \in \mathcal{R}_{\mathrm{ch}}$.

Proof. (i)⇒ (ii) ⇒ (iii) ⇒ (iv) follow by a direct application of Theorem 3 and Proposition 2. Next we focus on (iv) ⇒ (i). Note that

$$\begin{aligned}\mathbb{E}_g(X) &= \int_0^\infty (1 - g(F_X(x)))\mathrm{d}x - \int_{-\infty}^0 g(F_X(x))\mathrm{d}x \\ &= \int_0^\infty \tilde{g}(1 - F_X(x))\mathrm{d}x + \int_{-\infty}^0 (\tilde{g}(1 - F_X(x)) - 1)\mathrm{d}x,\end{aligned}$$

where $\tilde{g}(x) = 1 - g(1 - x),\ x \in [0, 1]$. Theorem 4.94 and Proposition 4.75 in Föllmer and Schied (2016) shows that convexity of $\mathbb{E}_g$ implies the concavity of $\tilde{g}$, which further indicates the convexity of g. This completes the proof. □

Remark 4. (i) As we can see from Proposition 5, convexity of g is actually a necessary and sufficient condition for $\mathbb{E}_g \in \mathcal{R}_{ch}$. For another risk measure instead of $\mathbb{E}$, it is sufficient but not necessary. For example, let ρ be a distortion risk measure given by

$$\rho(X) = \int_{\mathbb{R}} x\mathrm{d}F_X^\beta(x)$$

with distortion function $x^\beta,\ \beta > 1$. Note that $\rho \in \mathcal{R}_{ch}$. Moreover, let $g(x) = x^\alpha,\ x \in [0, 1]$ with $\alpha < 1$ which is a concave function. Then

$$\rho_g(X) = \int_{\mathbb{R}} x\mathrm{d}F_X^{\alpha\beta}(x)$$

is coherent if $1/\beta \leqslant \alpha < 1$.

(ii) A relevant property for risk measures used in insurance premium pricing is called loadedness. The risk measure ρ is *loaded* if

$$\rho(F) \geqslant \mu(F), \quad F \in \mathcal{M}.$$

See, e.g., Wang (2016) for properties and applications of loadedness. The ampliativity of T_g in (i) of Proposition 2 coincides with the loadedness of $\mathbb{E}_g$, i.e., for $g \in \mathcal{H}_0$, the condition $g(t) \leqslant t,\ t \in [0,1]$ is equivalent to

$$\mathbb{E}_g[X] \geqslant \mathbb{E}[X], \quad X \in L^\infty. \tag{10}$$

A short proof is as follows. If T_g is ampliative, it is obvious that $\mathbb{E}_g$ is loaded. Conversely, we take $X \sim \text{Bernoulli}(p)$ with $0 < p < 1$. Then (10) implies that $1 - g(1-p) \geqslant p$ for $0 < p < 1$. This means ampliativity by Proposition 2.

5.3 Examples

Next we present some risk measures generated by probability distortions. Let $\mathcal{H}_{\mathrm{cx}}$ denote the collection of convex functions in $\mathcal{H}_0$.

Example 11. Let $u : \mathbb{R} \to \mathbb{R}$ be a concave and strictly increasing function. The convex *utility-based shortfall risk measure*, introduced by Föllmer and Schied (2002), is given by

$$\rho(X) = \inf\{m \in \mathbb{R} : \mathbb{E}\left[u(-X+m)\right] \geqslant u(0)\}, \quad X \in L^\infty.$$

One can also refer to Weber (2006), Delbaen et al. (2016) and Föllmer and Schied (2016) for more results on the utility-based shortfall risk measures.

For $g \in \mathcal{H}_0$, we have

$$\rho_g(X) = \inf\left\{m \in \mathbb{R} : \int_{\mathbb{R}} u(-x+m)\mathrm{d}g(F_X(x)) \geqslant u(0)\right\}.$$

Note that

$$\int_{\mathbb{R}} u(-x)\mathrm{d}g(F_{-X}(x)) = \int_{\mathbb{R}} u(x)\mathrm{d}(1-g(F_{-X}(-x))) = \int_{\mathbb{R}} u(x)\mathrm{d}h(F_X(x)) := H_{u,h}(X)$$

with $h(x) = 1 - g(1-x),\ x \in [0,1]$, which is called a rank-dependent expected utility (see Quiggin (1982, 1993)). Then

$$\rho_g(X) = \inf\{m \in \mathbb{R} : H_{u,h}(-X+m) \geqslant u(0)\}$$

is a risk measure induced by a rank-dependent expected utility $H_{u,h}$. Proposition 5 implies that $\rho_g \in \mathcal{R}_{\mathrm{cx}}$ if $g \in \mathcal{H}_{\mathrm{cx}}$.

Example 12. Let ρ be the coherent distortion risk measure given by

$$\rho(X) = \int_{\mathbb{R}} x \mathrm{d}h(F_X(x)), \quad X \in L^\infty,$$

where $h \in \mathcal{H}_{\mathrm{cx}}$ is a distortion function. For $g \in \mathcal{H}_0$,

$$\rho_g(X) = \int_{\mathbb{R}} x \mathrm{d}h(g(F_X(x)))$$

is also a distortion risk measure with the distortion function $h \circ g$. By Proposition 5, $g \in \mathcal{H}_{\mathrm{cx}}$ implies the coherence of ρ_g.

Example 13. Let $h : [0, \infty) \to \mathbb{R} \cup \{+\infty\}$ be a lower-semicontinuous convex function satisfying $h(1) = 1$ and $\lim_{x\to\infty} \frac{h(x)}{x} = \infty$. The divergence risk measure with respect to h has the following expression (see, e.g., Theorem 4.122 in Föllmer and Schied (2016))

$$\rho(X) = \inf_{z\in\mathbb{R}} \left(\mathbb{E}[h^*(X - z)] + z\right), \quad X \in L^\infty,$$

where $h^*(y) = \sup_{x>0}(xy - h(x))$ is the Fenchel-Legendre transform of h. Note that if $h = \infty \cdot \mathbb{1}(1/\lambda, \infty)$ with $\lambda > 0$ and $\mathbb{1}$ representing the indicator function, then

$$\rho(X) = \inf_{z\in\mathbb{R}} \left(\mathbb{E}\left[\frac{(X - z)^+}{\lambda}\right] + z\right)$$

is the Expected Shortfall. For $g \in \mathcal{H}_{\mathrm{cx}}$, by Proposition 5,

$$\rho_g(X) = \inf_{z\in\mathbb{R}} \left(\int_{\mathbb{R}} h^*(x - z) \mathrm{d}g(F_X(x)) + z\right)$$

is a convex risk measure.

6 Sensitivity analysis

In this section we propose a new method for sensitivity analysis using distributional transforms. For this purpose, we consider distributional transforms $T : L^1 \to L^1$, and we naturally extend risk measures generated by probability distortions to L^1. Let $\rho : L^1 \to \mathbb{R}$ be a law-invariant risk measure and for $g \in \mathcal{H}_0$ satisfying $g(1-) = 1$, define $\rho_g : L^1 \to \mathbb{R} \cup \{\infty\}$ by

$$\rho_g(X) = \begin{cases} \rho((g \circ F_X)^{-1}(U)), & \text{if } (g \circ F_X)^{-1}(U) \in L^1, \\ \infty, & \text{if } (g \circ F_X)^{-1}(U) \notin L^1. \end{cases}$$

Let us first consider a model consisting of three elements:

$$Y = h(X),$$

where $X \in L^1$ is the input, $h : \mathbb{R} \to \mathbb{R}$ is the model function and $Y \in L^1$ is the output. We typically measure the sensitivity of the output risk to an individual model input by stressing

that model input with a random shock. Specifically, for $Z \in L^1$, the sensitivity of $\rho(h(X))$ to the shock Z can be defined as

$$\frac{\partial}{\partial \varepsilon} \rho(h(X + \varepsilon Z))\Big|_{\varepsilon=0}. \tag{11}$$

We refer to Tsanakas and Millossovich (2016) and the references therein for the definitions, results and applications of sensitivity of risk measures.

For the case that h is the identity, the sensitivity of $\rho(X)$ to the shock Z can be reduced to

$$\frac{\partial}{\partial \varepsilon} \rho(X + \varepsilon Z)\Big|_{\varepsilon=0}.$$

If we assume that X and Z are independent, then

$$\rho(X + \varepsilon Z) = \rho_{T_\varepsilon}(X), \text{ where } T_\varepsilon \circ F_X(x) = \int_{\mathbb{R}} F_X(x - y) F_{\varepsilon Z}(\mathrm{d}y).$$

This means that the sensitivity of risk measures can be derived from a family of risk measures generated by convolution transforms in Example 7.

Inspired by this fact, we consider the sensitivity of risk measures with respect to distributional transforms. The main idea here is to replace $h(X + \varepsilon Z)$ in (11) by $T_\varepsilon(h(X))$ where T_ε is a distributional transform arbitrarily close to identity as $\varepsilon \downarrow 0$ or $\varepsilon \uparrow 0$, and then we evaluate the limit of

$$\frac{\rho(T_\varepsilon(h(X))) - \rho(h(X))}{\varepsilon} \text{ as } \varepsilon \to 0. \tag{12}$$

This formulation allows for the analysis of sensitivity for more general types of distributional perturbation, uncertainty, and shocks, not limited to the additive shock in (11). We put a focus on probability distortions, and study the sensitivity defined in (12),

$$\frac{\rho(T_\varepsilon(W)) - \rho(W)}{\varepsilon} \text{ as } \varepsilon \to 0,$$

where $W = h(X)$ is a generic random variable. To realize the above idea, we need to formally construct sequences of probability distortions that converge to the identity in a meaningful way. For this purpose, we borrow from the notion of composition (semi-)groups.

Let $\{\Psi_t\}_{t\in\mathbb{R}}$ be a composition group of strictly increasing and continuous functions $\Psi_t : [0,1] \to [0,1]$ such that the following conditions hold:

(i) $\Psi_t(0) = 0$ and $\Psi_t(1) = 1$, $t \in \mathbb{R}$.

(ii) $\Psi_0(x) = x$, $x \in [0,1]$.

(iii) $\Psi_s \circ \Psi_t = \Psi_{s+t}$, $s, t \in \mathbb{R}$.

(iv) Ψ_t is convex for any fixed $t \geqslant 0$.

Note that (iii) indicates the group $\{\Psi_t\}_{t\in\mathbb{R}}$ is commutative, which is called abelian group. The study of composition groups or semigroups of analytic or holomorphic functions and their infinitesimal generators has a long history; see, e.g., Shapiro (1993), Shoikhet (2001) and the references therein. Here, we will consider composition groups of convex and concave functions Ψ_t, where the trivial case $\Psi_t(x) = x$ for all $t \in \mathbb{R}$ and all $x \in [0,1]$ is excluded. The group $\{\Psi_t\}_{t\in\mathbb{R}}$ admits the infinitesimal generator

$$G(x) = \lim_{t\to 0} \frac{x - \Psi_t(x)}{t}, \tag{13}$$

defined as the pointwise limit for $x \in [0,1]$, where $G : [0,1] \to [0,\infty)$ is concave and positive on $(0,1)$. Note that $G(0) = 0$ and $G(1) = 0$ by (13). Moreover, Ψ_t can be recovered from G by means of the following formula:

$$\Psi_t(x) = \sup\left\{y \in [0,1] : \int_y^x \frac{1}{G(s)}\mathrm{d}s = t\right\},\ t \in \mathbb{R},\ x \in (0,1). \tag{14}$$

This result follows by applying Theorem 2.1 of Cherny and Filipović (2011) to the associated composition group $\Psi_t^*(x) = 1 - \Psi_t(1-x), t \geqslant 0$ of concave functions on $[0,1]$. It is clear that $\Psi_t \in \mathcal{H}_0$. Moreover, note that the strict monotonicity and continuity of Ψ_t indicates

$$\int_0^{1/2} \frac{1}{G(s)}\mathrm{d}s = \infty;\ \int_{1/2}^1 \frac{1}{G(s)}\mathrm{d}s = \infty. \tag{15}$$

Below, we present some examples for convex composition groups that are adapted from Cherny and Filipović (2011).

Example 14. 1. $\Psi_t(x) = x^{e^t},\ t \in \mathbb{R},\ x \in [0,1]$ with $G(x) = -x\ln x,\ x \in (0,1)$. A direct calculation shows that if e^t is an integer, then

$$\mathbb{E}_{\Psi_t}[X] = \mathbb{E}[\max\{X_1, \ldots, X_{e^t}\}],$$

where $X_1, \ldots, X_{e^t}$ are iid copies of X.

2. $\Psi_t(x) = 1 - (1-x)^{e^{-t}},\ t \in \mathbb{R},\ x \in [0,1]$ with $G(x) = -(1-x)\ln(1-x),\ x \in (0,1)$. A direct calculation gives that if e^t is an integer, then

$$\mathbb{E}_{\Psi_{-t}}[X] = \mathbb{E}[\min\{X_1, \ldots, X_{e^t}\}],$$

where $X_1, \ldots, X_{e^t}$ are iid copies of X.

3. $\Psi_t(x) = \Phi(\Phi^{-1}(x) - t),\ t \in \mathbb{R},\ x \in [0,1]$ with $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\Phi^{-1}(x))^2}{2}},\ x \in (0,1)$, which is the convex version of Wang's transform (Wang (2000)). Here, Φ denotes again the cdf of the standard normal distribution.

Remark 5. Suppose that $\Psi \in \mathcal{H}_0$ is a given convex function. It is an interesting question whether Ψ can be embedded into a composition group $\{\Psi_t\}_{t\in\mathbb{R}}$. That is, whether there exists a composition group $\{\Psi_t\}_{t\in\mathbb{R}}$ such that $\Psi = \Psi_1$. According to Smajdor (1975), this holds if Ψ is strictly increasing, differentiable on $(0,1)$, and satisfies $\Psi'(0+) > 0$. Moreover, under these conditions, the infinitesimal generator G can be defined by means of (13). Finally, the main result in Smajdor (1975) states that Ψ_t will be a convex function for each $t > 0$ if and only if the infinitesimal generator G is concave.

We will measure the sensitivity of the risk measure ρ with respect to $\{\Psi_t\}_{t\in\mathbb{R}}$ with generator G satisfying (15) by means of the following two quantities,

$$S^+_{\rho,G}(X) = \lim_{t\downarrow 0} \frac{\rho_{\Psi_t}(X) - \rho(X)}{t}, \quad S^-_{\rho,G}(X) = \lim_{t\uparrow 0} \frac{\rho_{\Psi_t}(X) - \rho(X)}{t}, \quad X \in L^1.$$

If $S^+_{\rho,G}(X) = S^-_{\rho,G}(X)$, we say that the sensitivity of ρ with respect to Ψ_t at X is

$$S_{\rho,G}(X) = S^{\pm}_{\rho,G}(X).$$

Note that if $t > 0$, as $\Psi_t(x) \leqslant x$ for all $x \in [0,1]$, using Proposition 2, we have $\rho_{\Psi_t} \geqslant \rho$ for all monotone law-invariant risk measures ρ. Similarly, $\rho_{\Psi_t} \leqslant \rho$ for $t < 0$. Therefore, $S^{\pm}_{\rho,G}(X) \geqslant 0$.

Remark 6. In some situations, non-invertible convex or concave functions will only give rise to composition semigroups rather than groups. For instance,

$$\Psi_t(x) = 1 - (e^t(1-x)) \wedge 1, \ t \geqslant 0, x \in [0,1], \text{ or } \Psi_t(x) = (e^{-t}x) \wedge 1, \ t \leqslant 0, x \in [0,1].$$

In such a situation, the infinitesimal generator G of the semigroup can still be defined analogously as the right- or left-hand limit. Likewise, either $S^+_{\rho,G}$ or $S^-_{\rho,G}$ can still be defined.

In the next result, we consider distortion risk measures, which have the form $\rho = \mathbb{E}_g$ with $g \in \mathcal{H}_0$ satisfying $g(1-) = 1$, i.e.,

$$\rho(X) = \int_{\mathbb{R}} x \mathrm{d}g(F_X(x)) = \int_0^\infty (1 - g(F_X(x)))\mathrm{d}x - \int_{-\infty}^0 g(F_X(x))\mathrm{d}x.$$

The Hölder continuity of g can guarantee the finiteness of $\rho(X)$ for $X \in L^1$. For any function g, denote by $g'_-(t)$ its left derivative and $g'_+(t)$ its right derivative at t provided that the left and right derivatives exist.

Proposition 6. *Suppose that $g \in \mathcal{H}_0$ is Hölder continuous on $[0,1]$ with the left and right derivatives of g existing on $(0,1)$, and $\{\Psi_t\}_{t\in\mathbb{R}}$ is a composition group satisfying (i)-(iv) with its infinitesimal generator G satisfying (15). For $X \in L^1$, if*

$$\int_{\mathbb{R}} |x| \mathrm{d}\Psi_t(F_X(x)) < \infty \quad \text{and} \quad \int_{\mathbb{R}} \max\{G(F_X(x)), G(\Psi_t(F_X(x)))\}\mathrm{d}x < \infty \tag{16}$$

for some $t > 0$, then

$$S^+_{\rho,G}(X) = \int_{\mathbb{R}} g'_-(F_X(x))G(F_X(x))\mathrm{d}x.$$

If (16) holds for some $t < 0$, then

$$S^-_{\rho,G}(X) = \int_{\mathbb{R}} g'_+(F_X(x))G(F_X(x))\mathrm{d}x.$$

Proof. We first consider $S^+_{\rho,G}(X)$. Note that for $X \in L^1$, Hölder continuity of g and (16) for some $t > 0$ lead to the finiteness of $\rho(X)$ and $\rho_{\Psi_t}(X)$ for $t > 0$ sufficiently small. By definition,

$$\begin{aligned}\frac{\rho_{\Psi_t}(X) - \rho(X)}{t} &= \int_{\mathbb{R}} \frac{g(F_X(x)) - g(\Psi_t(F_X(x)))}{t}\mathrm{d}x \\ &= \int_{F^{-1}_{X+}(0)}^{F_X^{-1}(1)} \frac{g(F_X(x)) - g(\Psi_t(F_X(x)))}{t}\mathrm{d}x,\end{aligned}$$

where F_{X+} is the right quantile function of F_X. If $F^{-1}_{X+}(0) = F_X^{-1}(1)$, the claim holds trivially. We next consider the case $F^{-1}_{X+}(0) < F_X^{-1}(1)$. Without loss of generality, we assume that $F^{-1}_{X+}(0) > -\infty$. Note that if there exists some $t > 0$ and $x \in (0,1)$ such that $\Psi_t(x) = x$, then $\Psi_t(x) = x$ for all $t \in \mathbb{R}$ and for all $x \in [0,1]$, which has been excluded. Hence $\Psi_t(F_X(x)) < F_X(x)$ for all $t > 0$ and $x \in (F^{-1}_{X+}(0), F_X^{-1}(1))$, which implies that

$$\lim_{t\downarrow 0} \frac{g(F_X(x)) - g(\Psi_t(F_X(x)))}{t} = g'_-(F_X(x))G(F_X(x)),\ x \in (F^{-1}_{X+}(0), F_X^{-1}(1)).$$

Note that (15) implies $G(0+) = G(1-) = 0$. By the concavity of G, there exists $\varepsilon \in (0, 1/2)$ such that G is increasing over $(0,\varepsilon)$ and decreasing over $(1-\varepsilon, 1)$. Hence, there exists $\delta \in (0,\varepsilon)$ such that for all $x \in (0,\delta)$ and $t \in (0,\delta)$, we have $\Psi_t(x) \in (0,\varepsilon)$, and for all $x \in (1-\delta, 1)$ and $t \in (0,\delta)$, we have $\Psi_t(x) \in (1-\varepsilon, 1)$. For $x \in (0,\delta) \cup (1-\delta, 1)$, using (14) and the fact that $G(s)$ is monotone for $s \in [\Psi_t(x), x]$, we have

$$0 \leqslant \frac{x - \Psi_t(x)}{t} = \frac{x - \Psi_t(x)}{\int_{\Psi_t(x)}^x \frac{1}{G(s)}\mathrm{d}s} \leqslant \max_{\Psi_t(x) < s \leqslant x} G(s) \leqslant \max(G(x), G(\Psi_t(x))).$$

Consequently, for $x > F^{-1}_{X+}(1-\delta)$,

$$\sup_{t\in(0,\delta)} \left| \frac{g(x) - g(\Psi_t(x))}{t} \right| \leqslant \max(G(x), G(\Psi_t(x))) \sup_{t\in(0,1)} g'_-(t) < \infty.$$

Observe that $\max(G(x), G(\Psi_t(x)))$ is an increasing function with respect to t for x sufficiently close to 1 and $t > 0$ sufficiently close to 0, and note that $\max_{s\in(0,1)} G(s) < \infty$. Hence the dominated convergence theorem leads to

$$S^+_{\rho,G}(X) = \lim_{t\downarrow 0} \frac{\rho_{\Psi_t}(X) - \rho(X)}{t} = \int_{F^{-1}_{X+}(0)}^{F_X^{-1}(1)} g'_-(F_X(x))G(F_X(x))\mathrm{d}x.$$

We can establish the claim for $S^-_{\rho,G}$ using the same approach as above. This completes the proof. □

The assumptions of Proposition 6 are somewhat complicated and this is partially due to the unboundedness of $X \in L^1$. If we focus on bounded random variables, then the conditions related to (16) can be removed, as illustrated in the following corollary.

Corollary 1. *Under the conditions of Proposition 6, for* $X \in L^\infty$

$$S^+_{\rho,G}(X) = \int_{\mathbb{R}} g'_-(F_X(x))G(F_X(x))\mathrm{d}x \quad \textit{and} \quad S^-_{\rho,G}(X) = \int_{\mathbb{R}} g'_+(F_X(x))G(F_X(x))\mathrm{d}x.$$

Refining the proof of Proposition 6 for $\rho = \mathrm{ES}_p$, we get the sensitivity of ES_p with simpler assumptions.

Corollary 2. *Assume that* $\rho = \mathrm{ES}_p$ *with* $p \in (0,1)$. *Under the conditions of Proposition 6, for* $X \in L^1$, *if*

$$\int_{\mathbb{R}} |x| \mathrm{d}\Psi_t(F_X(x)) < \infty \tag{17}$$

and $\int_{y_p}^\infty G(\Psi_t(F_X(x)))\mathrm{d}x < \infty$ *for some* $t > 0$, *where* $y_p = F^{-1}_{X+}(p)$, *then*

$$S^+_{\rho,G}(X) = \frac{1}{1-p}\int_{y_p}^\infty G(F_X(x))\mathrm{d}x;$$

if (17) *holds for some* $t < 0$, *then*

$$S^-_{\rho,G}(X) = \frac{1}{1-p}\int_{z_p}^\infty G(F_X(x))\mathrm{d}x, \tag{18}$$

where $z_p = F^{-1}_X(p)$.

Proof. Note that $g(x) = \frac{(x-p)^+}{1-p}$ for $\rho = \mathrm{ES}_p$. The case of $S^+_{\rho,G}$ follows directly from Proposition 6, noting that for $t > 0$ sufficiently close to 0 and x sufficiently large

$$\max\{G(F_X(x)), G(\Psi_t(F_X(x)))\} = G(\Psi_t(F_X(x))).$$

Next we analyze $S^-_{\rho,G}$, which is slightly different. For $t < 0$ sufficiently close to 0 and x sufficiently large

$$\max\{G(F_X(x)), G(\Psi_t(F_X(x)))\} = G(F_X(x)).$$

Hence, for $t < 0$ where $|t|$ is small enough,

$$\int_{z_p}^\infty \max\{G(F_X(x)), G(\Psi_t(F_X(x)))\}\mathrm{d}x < \infty$$

is equivalent to

$$\int_{z_p}^\infty G(F_X(x))\mathrm{d}x < \infty.$$

Therefore, if $\int_{z_p}^\infty G(F_X(x))\mathrm{d}x < \infty$, then Proposition 6 yields (18).

It remains to verify that $S^-_{\rho,G}(X)=\infty$ whenever $\frac{1}{1-p}\int_{z_p}^{\infty}G(F_X(x))\mathrm{d}x=\infty$. Note that by definition,

$$\frac{\rho_{\Psi_t}(X)-\rho(X)}{t}=\int_{\mathbb{R}}\frac{g(F_X(x))-g(\Psi_t(F_X(x)))}{t}\mathrm{d}x.$$

If $\int_{z_p}^{\infty}G(F_X(x))\mathrm{d}x=\infty$,

$$\begin{aligned}\lim_{t\uparrow 0}\frac{\rho_{\Psi_t}(X)-\rho(X)}{t}&\geqslant\lim_{n\to\infty}\lim_{t\uparrow 0}\int_{z_p}^{n}\frac{g(F_X(x))-g(\Psi_t(F_X(x)))}{t}\mathrm{d}x\\&=\lim_{n\to\infty}\frac{1}{1-p}\int_{z_p}^{n}G(F_X(x))\mathrm{d}x=\infty.\end{aligned}$$

Hence, (18) holds. □

Next we discuss the sensitivity of $\mathrm{VaR}_p,\ p\in(0,1)$.

Proposition 7. *Assume that $\rho=\mathrm{VaR}_p$ for some $p\in(0,1)$ and X is a continuous random variable. If the right derivative of F_X at $z_p=F_X^{-1}(p)$ exists, then*

$$S^+_{\rho,G}(X)=\begin{cases}\frac{G(p)}{(F_X)'_+(z_p)}, & (F_X)'_+(z_p)>0,\\ \infty, & (F_X)'_+(z_p)=0.\end{cases}$$

If the left derivative of F_X at z_p exists, then

$$S^-_{\rho,G}(X)=\begin{cases}\frac{G(p)}{(F_X)'_-(z_p)}, & (F_X)'_-(z_p)>0,\\ \infty, & (F_X)'_-(z_p)=0.\end{cases}$$

Proof. Using the fact that Ψ_t is strictly increasing, we have

$$\begin{aligned}\rho_{\Psi_t}(X)&=\inf\{x\in\mathbb{R}:\Psi_t(F_X(x))\geqslant p\}\\&=\inf\{x\in\mathbb{R}:F_X(x)\geqslant\Psi_t^{-1}(p)\}=\mathrm{VaR}_{\Psi^{-1}(p)}(X).\end{aligned}$$

Hence

$$\begin{aligned}\frac{\rho_{\Psi_t}(X)-\rho(X)}{t}&=\frac{\mathrm{VaR}_{\Psi_t^{-1}(p)}(X)-\mathrm{VaR}_p(X)}{t}\\&=\frac{\mathrm{VaR}_{\Psi_t^{-1}(p)}(X)-\mathrm{VaR}_p(X)}{\Psi_t^{-1}(p)-p}\frac{\Psi_t^{-1}(p)-p}{t}.\end{aligned}$$

Noting that $\Psi^{-1}(p)>p$ for $t>0$ and by the fact that $(F_X)'_+(z_p)>0$, we have

$$\lim_{t\downarrow 0}\frac{\mathrm{VaR}_{\Psi_t^{-1}(p)}(X)-\mathrm{VaR}_p(X)}{\Psi_t^{-1}(p)-p}=\frac{1}{(F_X)'_+(z_p)}.$$

Moreover, a direct calculation shows

$$\lim_{t\downarrow 0}\frac{\Psi_t^{-1}(p)-p}{t}=G(p).$$

Consequently,

$$\lim_{t\downarrow 0}\frac{\rho_{\Psi_t}(X)-\rho(X)}{t}=\frac{G(p)}{(F_X)'_+(z_p)}.$$

For $(F_X)'_+(z_p)=0$, we have

$$\lim_{t\downarrow 0}\frac{\mathrm{VaR}_{\Psi_t^{-1}(p)}(X)-\mathrm{VaR}_p(X)}{\Psi_t^{-1}(p)-p}=\infty.$$

Hence

$$S^+_{\rho,G}(X)=\infty.$$

Using same approach as above, we can get $S^-_{\rho,G}(X)$. This completes the proof. □

Next we give a numerical analysis of sensitivity of VaR_p and ES_p for $X\sim\mathrm{Pareto}(1,\alpha),\ \alpha>1$, and $X\sim\mathrm{N}(0,\sigma^2)$ with composition groups in Example 14. One can easily check that the assumptions of Corollary 2 are all satisfied for $X\sim\mathrm{Pareto}(1,\alpha),\ \alpha>1$ and $X\sim\mathrm{N}(0,\sigma^2)$ with composition groups $G=-x\ln x,\ G(x)=-(1-x)\ln(1-x)$, and $G(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(\Phi^{-1}(x))^2}{2}},\ x\in[0,1]$.

For $\rho=\mathrm{ES}_p(X)$ with $X\sim\mathrm{Pareto}(1,\alpha),\ \alpha>1$, it follows from Corollary 2 that

$$S_{\rho,G}(X)=\frac{1}{1-p}\int_{(1-p)^{-1/\alpha}-1}^{\infty}G(1-(1+x)^{-\alpha})\mathrm{d}x.$$

For $\rho=\mathrm{VaR}_p(X)$ with $X\sim\mathrm{Pareto}(1,\alpha)$, by Proposition 7 it follows that

$$S_{\rho,G}(X)=\frac{G(p)}{\alpha(1-p)^{\frac{\alpha+1}{\alpha}}}.$$

In Tables 1-6, we report the sensitivity of ES_p and VaR_p for Pareto and normally distributed X, $p=0.9,0,975,0.99$, and three different composition groups in Example 14. Note that $\mathrm{ES}_{0.975}$ is recently proposed to replace $\mathrm{VaR}_{0.99}$ in Basel III/IV (BCBS (2016)), and hence the comparison between sensitivity of $\mathrm{ES}_{0.975}$ and that of $\mathrm{VaR}_{0.99}$ is the most important. From the numerical results, we can see that the sensitivity of $\mathrm{ES}_{0.975}$ is very close to that of $\mathrm{VaR}_{0.99}$ in examples of normal distributions and Pareto distributions with large α, showing that both risk measures exhibit similar sensitivity to model uncertainty and random perturbation in these cases. On the other hand, for $X\sim\mathrm{Pareto}(1,2)$ or $X\sim\mathrm{Pareto}(1,4)$, the sensitivity of $\mathrm{ES}_{0.975}$ is visibly higher than that of $\mathrm{VaR}_{0.99}$. This observation is in line with the conclusion of Cont et al. (2010) regarding the comparison on the robustness of ES and VaR.

7 Change of measures

In this section we discuss distributional transforms arising from a change of measures. Let $(\Omega,\mathcal{F},\mathbb{P})$ be an atomless probability space as before. For illustration, let $\mathbb{Q}$ denote another

$S_{\rho,G}(X)$ ρ / α	ES_p			VaR_p		
	$p=0.9$	$p=0.975$	$p=0.99$	$p=0.9$	$p=0.975$	$p=0.99$
2	3.108479	6.298070	9.983300	1.499306	3.122416	4.974916
4	0.579778	0.833774	1.051829	0.421561	0.620791	0.786603
10	0.136493	0.159724	0.175681	0.119377	0.142790	0.157694
20	0.057582	0.062905	0.066097	0.053197	0.059370	0.062630

Table 1: $G(x) = -x\ln x,\ X \sim \mathrm{Pareto}(1,\alpha)$

$S_{\rho,G}(X)$ ρ / α	ES_p			VaR_p		
	$p=0.9$	$p=0.975$	$p=0.99$	$p=0.9$	$p=0.975$	$p=0.99$
2	13.60597	35.97963	66.05170	3.640707	11.66526	23.02585
4	2.155226	4.210065	6.259732	1.023660	2.319260	3.640707
10	0.477510	0.771265	1.006633	0.289878	0.533458	0.729870
20	0.198138	0.300100	0.374882	0.129177	0.221803	0.289878

Table 2: $G(x) = -(1-x)\ln(1-x),\ X \sim \mathrm{Pareto}(1,\alpha)$

$S_{\rho,G}(X)$ ρ / α	ES_p			VaR_p		
	$p=0.9$	$p=0.975$	$p=0.99$	$p=0.9$	$p=0.975$	$p=0.99$
2	7.894700	18.77852	32.46422	2.774872	7.392782	13.32607
4	1.351766	2.329875	3.235342	0.780213	1.469816	2.107037
10	0.308187	0.435825	0.529572	0.220939	0.338076	0.422408
20	0.128877	0.170534	0.198156	0.098456	0.140566	0.167765

Table 3: $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\Phi^{-1}(x))^2}{2}},\ X \sim \mathrm{Pareto}(1,\alpha)$

$S_{\rho,G}(X)$ ρ / σ	ES_p			VaR_p		
	$p=0.9$	$p=0.975$	$p=0.99$	$p=0.9$	$p=0.975$	$p=0.99$
1	0.460335	0.375330	0.337978	0.540316	0.422360	0.373322
5	2.301676	1.876651	1.689887	2.701577	2.111800	1.866610
10	4.603352	3.753303	3.379775	5.403155	4.223601	3.733221
20	9.206705	7.506606	6.759550	10.80631	8.447201	7.466441

Table 4: $G(x) = -x\ln x,\ X \sim \mathrm{N}(0,\sigma^2)$

probability defined on $(\Omega,\mathcal{F})$ satisfying $\mathbb{Q} \ll \mathbb{P}$ and denote its Radon-Nikodym derivative by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. Write $F_X^{\mathbb{P}}$ and $F_X^{\mathbb{Q}}$ as the distributions of $X \in L^0$ under $\mathbb{P}$ and $\mathbb{Q}$ respectively. As the main

$S_{\rho,G}(X)$ ρ / σ	ES_p			VaR_p		
	$p=0.9$	$p=0.975$	$p=0.99$	$p=0.9$	$p=0.975$	$p=0.99$
1	1.498554	1.734090	1.870863	1.312027	1.577926	1.727880
5	7.492768	8.670451	9.354314	6.560134	7.889629	8.639400
10	14.98554	17.34090	18.70863	13.12027	15.77926	17.27880
20	29.97107	34.68180	37.41725	26.24054	31.55852	34.55760

Table 5: $G(x) = -(1-x)\ln(1-x),\ X \sim \mathrm{N}(0,\sigma^2)$

$S_{\rho,G}(X)$ ρ / σ	ES_p			VaR_p		
	$p=0.9$	$p=0.975$	$p=0.99$	$p=0.9$	$p=0.975$	$p=0.99$
1	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
5	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000
10	10.00000	10.00000	10.00000	10.00000	10.00000	10.00000
20	20.00000	20.00000	20.00000	20.00000	20.00000	20.00000

Table 6: $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\Phi^{-1}(x))^2}{2}},\ X \sim \mathrm{N}(0,\sigma^2)$

result of this section, we identify two probability distortions, which transform $F_X^{\mathbb{P}}$ to $F_X^{\mathbb{Q}}$, given that the risk is a monotone function of $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P}$.

Proposition 8. *Denote by $X = f\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)$ for some $f : \mathbb{R} \to \mathbb{R}$.*

(i) If f is increasing,

$$F_X^{\mathbb{Q}} = g_1 \circ F_X^{\mathbb{P}},\ \textit{with } g_1(x) = \int_0^x (F^{\mathbb{P}}_{\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}})^{-1}(t)\mathrm{d}t,\ x \in [0,1];$$

(ii) If f is decreasing,

$$F_X^{\mathbb{Q}} = g_2 \circ F_X^{\mathbb{P}},\ \textit{with } g_2(x) = \int_0^x (F^{\mathbb{P}}_{\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}})^{-1}(1-t)\mathrm{d}t,\ x \in [0,1].$$

Proof. (i): We denote $\eta = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. A direct calculation gives

$$F_X^{\mathbb{Q}}(x) = \mathbb{Q}\left(f(\eta) \leqslant x\right) = \mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{f(\eta)\leqslant x\}}\eta\right].$$

Noting that $(F_\eta^{\mathbb{P}})^{-1}(U)$ has the same distribution as η under $\mathbb{P}$ with U being a standard uniform random variable under $\mathbb{P}$, we have

$$\begin{aligned} F_X^{\mathbb{Q}}(x) &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{f((F_\eta^{\mathbb{P}})^{-1}(U))\leqslant x\}}(F_\eta^{\mathbb{P}})^{-1}(U)\right] \\ &= \int_0^1 \mathbb{1}_{\{f((F_\eta^{\mathbb{P}})^{-1}(t))\leqslant x\}}(F_\eta^{\mathbb{P}})^{-1}(t)\mathrm{d}t. \end{aligned}$$

Since f is increasing, $f((F_\eta^{\mathbb{P}})^{-1}(t)) = (F_X^{\mathbb{P}})^{-1}(t)$, for a.e. $t \in (0,1)$. Hence

$$\begin{aligned} F_X^{\mathbb{Q}}(x) &= \int_0^1 \mathbb{1}_{\{(F_X^{\mathbb{P}})^{-1}(t) \leqslant x\}} (F_\eta^{\mathbb{P}})^{-1}(t) \mathrm{d}t \\ &= \int_0^{F_X^{\mathbb{P}}(x)} (F_\eta^{\mathbb{P}})^{-1}(t) \mathrm{d}t = g_1(F_X^{\mathbb{P}}(x)). \end{aligned}$$

Observe that $g_1 \in \mathcal{H}_0$ is continuous and convex. Part (ii) follows from a similar argument as in the proof of (i). □

Let us now see a special case, where $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ has a lognormal distribution. Let $\{B_t\}_{t \geqslant 0}$ be a standard Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$, $t_0 > 0$ be a time horizon and

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = e^{wB_{t_0} - \frac{w^2}{2} t_0}, \ w \in \mathbb{R}. \tag{19}$$

The quantile of $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ under $\mathbb{P}$ can be expressed as

$$(F^{\mathbb{P}}_{\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}})^{-1}(p) = e^{-\frac{w^2 t_0}{2} + |w|\sqrt{t_0}\Phi^{-1}(p)}, \ p \in (0,1).$$

Hence

$$\begin{aligned} g_1(x) &= \int_0^x e^{-\frac{w^2 t_0}{2} + |w|\sqrt{t_0}\Phi^{-1}(t)} \mathrm{d}t = \Phi(\Phi^{-1}(x) - |w|\sqrt{t_0}), \\ g_2(x) &= \int_0^x e^{-\frac{w^2 t_0}{2} + |w|\sqrt{t_0}\Phi^{-1}(1-t)} \mathrm{d}t = \Phi(\Phi^{-1}(x) + |w|\sqrt{t_0}), \end{aligned}$$

where Φ is the standard normal distribution. Next we apply the above result to option pricing in a simplest model: the Black-Scholes market model. Recall that in a Black-Scholes market model, the stochastic process

$$S_t = S_0 e^{\sigma B_t + (\mu - \frac{\sigma^2}{2})t}, \ t \geqslant 0$$

represents the price of an underlying asset with volatility $\sigma > 0$ and average rate of return $\mu > 0$. The discounted payoffs of European call and put options with underlying stock S_t with no dividends, strike $K > 0$ and expiry date $t_0 > 0$ are

$$(S_{t_0} - K)^+ e^{-rt_0} \text{ and } (K - S_{t_0})^+ e^{-rt_0}$$

respectively, where $r > 0$ is the risk-free interest rate. We denote their distributions under $\mathbb{P}$ by F_c and F_p respectively. Define a probability $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by its Radon-Nikodym derivative (19).

Corollary 3. *Under the above assumptions, we have, for $x \in \mathbb{R}$,*

$$\mathbb{Q}\left((S_{t_0} - K)^+ e^{-rt_0} \leqslant x\right) = \Phi(\Phi^{-1}(F_c(x)) - |w|\sqrt{t_0}),$$

and

$$\mathbb{Q}\left((K - S_{t_0})^+ e^{-rt_0} \leqslant x\right) = \Phi(\Phi^{-1}(F_p(x)) + |w|\sqrt{t_0}).$$

If we take $w = \frac{r-\mu}{\sigma}$, then $\mathbb{Q}$ is the risk-neutral probability measure in the Black-Scholes market. This gives the price of European call and put options,

$$\mathbb{E}^{\mathbb{Q}}\left[(S_{t_0} - K)^+ e^{-rt_0}\right] = \int_{\mathbb{R}} x \mathrm{d}\Phi(\Phi^{-1}(F_c(x)) - \sigma^{-1}|r-\mu|\sqrt{t_0}),$$
$$\mathbb{E}^{\mathbb{Q}}\left[(K - S_{t_0})^+ e^{-rt_0}\right] = \int_{\mathbb{R}} x \mathrm{d}\Phi(\Phi^{-1}(F_p(x)) + \sigma^{-1}|r-\mu|\sqrt{t_0}),$$

which coincide with the Black-Scholes formula.

In Proposition 8, probability distortions are used to describe the distribution changes for a class of random variables under different probability measures. It can be applied to other problems related to changes of measures, as long as they are monotone functions of the Radon-Nikodym derivative.

8 Concluding remarks

In this paper, we formally introduce the mathematical framework of distributional transforms. Via several technical properties, we characterize probability distortions and the class of distributional transforms generated by Choquet integrals. The power of distributional transforms is illustrated by means of three applications: risk measures, sensitivity analysis, and change of measures. In view of the fact that distributional transforms have appeared in many different areas such as finance, economics, risk analysis and optimization, our work serves as a (theoretical) building block for their future applications. Our approach is based on quantile techniques and is thus genuinely one-dimensional. Moreover, the notion of a probability distortion as used in this paper is limited to probability measures on $\mathbb{R}$. An analysis of distributional transforms for multivariate probability distributions will be developed in future research.

References

BCBS (2016). *Standards. Minimum Capital Requirements for Market Risk. January 2016.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

Chambers, C. P. (2009). An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance*, **19**(2), 335–342.

Cherny, A. S. and Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, **22**(7), 2571–2606.

Cherny, A. S. and Filipović, D. (2011). Concave distortion semigroups. *arXiv:* `1104.0508`.

Chew, S. H., Karni, E. and Safra, Z. (1987). Risk aversion in the theory of expected utility with rank dependent probabilities. *Journal of Economic Theory*, **42**(2), 370–381.

Cont, R., Deguest, R., and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**(6), 593–606.

Delbaen, F. (2012). *Monetary Utility Functions*. Osaka Univeristy Press.

Delbaen, F., Bellini, F., Bignozzi, V. and Ziegel, J. (2016). Risk measures with the CxLS property. *Finance and Stochastics*. **20**(2), 433–453.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**, 429–447.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Walter de Gruyter, Berlin, Fourth Edition.

Holzmann, H. and Klar, B. (2017). Focusing on regions of interest in forecast evaluation. *Annals of Applied Statistics*. **11**(4), 2404–2431.

Gastwirth, J.L. (1971). A general definition of the Lorenz curve. *Econometrica*. **39**(6), 1037–1039.

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

Lerch, S., Thorarinsdottir, T. L., Ravazzolo, F. and Gneiting, T. (2017). Forecaster's dilemma: Exreme events and forecast evaluation. *Statistical Science*. **32**(1), 106–127.

Liu, F. and Wang, R. (2020). A theory for measures of tail risk. *Mathematics of Operations Research*, forthcoming.

Lorenz, M.O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**(70), 209–219.

Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, **3**(4), 323–343.

Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model*. Kluwer, the Netherlands.

Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.

Schmeidler, D. (1986). *Integral representation without additivity. Proceedings of the Americian Mathematical Society*, **97**(2), 255–261.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics.

Shapiro, J. H. (1993). *Composition Operators and Classical Function Theory*. Universitext: Tracts in Mathematics. Springer-Verlag, New York.

Shoikhet, D. (2001). *Semigroups in Geometrical Function Theory*. Kluwer Academic Publishers, Dordrecht.

Smajdor, A. (1975). Note on the existence of convex iteration groups. *Fundamenta Mathematicae*, **87**(3), 213–218.

Tsanakas, A. and Millossovich, P. (2016). Sensitivity analysis using risk measures. *Risk Analysis*, **36**(1), 30–48.

Tversky, A. and Kahneman, E. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**, 297–323.

Wang, R. (2016). Regulartory arbitrage of risk measures. *Quantitative Finance*, **16**(3), 337–347.

Wang, R. and Wei, Y. and Willmot, G. E. (2019). Characterization, robustness and aggregation of signed Choquet Integrals. *Mathematics of Operations Research*, forthcoming.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, **26**(1), 71–92.

Wang, S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance*, **67**, 15–36.

Weber, S. (2006). Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance*, **16**, 419–441.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.