Dual utilities under dependence uncertainty*

Ruodu Wang† Zuo Quan Xu‡ Xun Yu Zhou§

5th July 2018

Abstract

Finding the worst-case value of a preference over a set of plausible models is a well-established approach to address the issue of model uncertainty or ambiguity. In this paper, we study the worst-case evaluation of Yaari's dual utility functionals of an aggregate risk under dependence uncertainty along with its decision-theoretic implications. To arrive at our main findings, we introduce a technical notion of conditional joint mixability. Lower and upper bounds on dual utilities with dependence uncertainty are established and, in the presence of conditional joint mixability, they are shown to be exact bounds. Moreover, conditional joint mixability is indeed necessary for attaining these exact bounds when the distortion functions are strictly inverse-S-shaped. A particular economic implication of our results is what we call the pessimism effect. We show that a (generally non-convex/non-concave) dual utility-based decision maker under dependence uncertainty behaves as if she had a more pessimistic risk-averse dual utility but free of dependence uncertainty.

Keywords: dual utility; conditional joint mixability; risk aggregation; dependence uncertainty; pessimism effect

---

*Wang acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590). Xu acknowledges financial supports from NSFC (No. 11471276), Hong Kong GRF (No. 15204216 and No. 15202817), and the Hong Kong Polytechnic University. Zhou acknowledges financial supports through a start-up grant at Columbia University, and from Columbia's FDT Center for Intelligent Asset Management.

†Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada. Email: wang@uwaterloo.ca

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong. Email: maxu@polyu.edu.hk

§Department of Industrial Engineering and Operations Research, Columbia University, New York, USA. Email: xz2574@columbia.edu

1 Introduction

Expected utility theory (EUT) has been well studied in both the static and dynamic settings, ranging from investment problems to corporate finance. However, behavioral patterns that are inconsistent with EUT have been observed. For instance, EUT fails to explain many paradoxes and puzzles such as the Allais paradox, the Ellesberg paradox,the Friedman and Savage puzzle, and the equity premium puzzle. Considerable attempts and efforts have been made to overcome the drawbacks of EUT; for example, Yaari's dual theory of choice (DT; Yaari (1987)), Quiggin's rank-dependent utility theory (RDUT; Quiggin (1982); Schmeidler (1989)), and Lopes' SP/A theory (Lopes (1987)). Among many others, Kahneman and Tversky's cumulated prospect theory (CPT; Kahneman and Tversky (1979); Tversky and Kahneman (1992)) is arguably the most prominent alternative to EUT. Probability weighting (or distortion) – the assumption that individuals overweight unlikely and extreme events, as a key ingredient of the theory, is proposed to capture a different dimension of risk preference over the tail events. It can in turn explain many phenomena; e.g. how hope and fear, properly defined through probability weighting, are present simultaneously for the same individual (see He and Zhou (2016)).

In decision theory, a decision maker's preference, denoted by a binary relationship $\prec$, is often modeled by a preference functional $\mathcal{U}$. That is, for any two random outcomes X and Y,

$$X \prec Y \;\Leftrightarrow\; \mathcal{U}(X) \leqslant \mathcal{U}(Y), \tag{1.1}$$

where $X \prec Y$ means that the decision maker prefers Y at least as much as X. A decision maker whose preference satisfies (1.1) will be referred to as a $\mathcal{U}$-decision maker in this paper.

Typically, the evaluation of $\mathcal{U}(X)$, such as in EUT, DT, RDUT or CPT, requires that the underlying probability measure (or the distribution of X) to be specified. In practice, though, it is more likely that a decision maker is unable to know the exact probability measure governing the random outcome of a decision (as in the Ellsberg paradox). In such cases the decision maker may be concerned about the utility of her decisions under the "worst-case". This leads to the theory of ambiguity/uncertainty aversion (or robust theory) of Gilboa and Schmeidler (1989) and Hansen and Sargent (2001). In mathematical terms, this means

$$X \prec Y \;\Leftrightarrow\; \inf_{Q\in\mathcal{Q}} \mathcal{U}^Q(X) \leqslant \inf_{Q\in\mathcal{Q}} \mathcal{U}^Q(Y),$$

where $\mathcal{Q}$ is a collection of possible probability measures, and $\mathcal{U}^Q$ is the utility of X evaluated under $Q \in \mathcal{Q}$.

In this paper we address a particular type of uncertainty: the uncertainty in the dependence structure among random variables. This setting, termed *dependence uncertainty* by Bernard et al. (2014), is relevant in many practical cases in which a precise knowledge or modeling for dependence among many random variables is unavailable or simply impossible. Under this setting, the preference of the decision maker is dictated by the worst-case (or best-case) aggregate value of her preference functional. The mathematical formulation is the following: given several distributions $F_1, \dots, F_n$, one needs to evaluate

$$\text{inf or sup of} \quad \mathcal{U}\left(\sum_{i=1}^{n} X_i\right) \text{ subject to } X_i \sim F_i \;, i = 1, \dots, n. \tag{1.2}$$

As a concrete example, an investor (a $\mathcal{U}$-decision maker), on top of her financial investment outcome X, has an additional income Y whose distribution F is known but its dependence with the financial market is unspecified[1]. In this case, the worst-case value of her utility for a financial investment outcome X is

$$\inf \left\{\mathcal{U}(X+Y) : Y \sim F\right\}, \tag{1.3}$$

which is a special case of (1.2). More examples of (1.2) in decision analysis will be given in Section 5.

The values of (1.2) can be very difficult to evaluate, especially in the case when $\mathcal{U}$ is neither concave or convex. In this paper we focus on the case in which $\mathcal{U}$ in (1.2) is Yaari's dual utility (see (2.1) below for definition). When $\mathcal{U}$ is a quantile (i.e. VaR), problem (1.2) has recently led to an active stream of research in the risk management literature; see e.g. Embrechts et al. (2013, 2015) and McNeil et al. (2015, Section 8.4). The paper by Embrechts et al. (2014) contains a survey on risk aggregation with dependence uncertainty in the context of regulatory capital calculation. Unfortunately,

[1] Such a Y is termed *intractable claim* in Hou and Xu (2016).

the mathematical results from risk management cannot be applied directly to general dual utilities. A major difficulty associated with dual utilities is that commonly used dual utilities are neither convex nor concave functionals, whereas in risk management typically one considers either convex functionals (called convex risk measures in finance) or quantile functionals (which have nice properties, even if not convex or concave).

In this paper, we focus on the calculation of (1.2) and its economic implications. Our main contribution is threefold. First we introduce a notion of conditional joint mixability, a generalization of joint mixability by Wang et al. (2013), which itself is of mathematical interest. Second, we establish bounds on the worst- and best-case values of dual utilities and show that, under an assumption of conditional joint mixability, the above bounds are indeed the exact values of (1.2). For strict inverse-S-shaped distortion functions, conditional joint mixability is shown to be necessary and sufficient for the attainability of the above bounds. These mathematical results will serve as a building block for future research on behavioral investment problems and decision analysis under model uncertainty. Furthermore, we discuss a particular application of our main results, which we call a *pessimism effect*. We shall see that, with the aversion to dependence uncertainty, a decision maker makes decisions according to an auxiliary risk-averse dual utility that is more pessimistic than her original utility functional, but with the dependence uncertainty removed. This shows that the uncertainty aversion can be "transferred" and built into the dual utility itself.

The rest of the paper is organized as follows. In Section 2 we gather some preliminaries on dual utilities. In Section 3 we introduce conditional joint mixability and present some of its basic properties. In Section 4 bounds on dual utilities are established, and their sharpness is shown under conditional joint mixability; moreover we discuss particular results in the case of inverse-S-shaped distortion functions and in the case $n = 2$. Through two concrete examples, in Section 5 we illustrate the pessimism effect derived from our main results. Finally we conclude in Section 6.

2 Preliminaries

2.1 Dual utilities

We first fix some notation. Let L^0 be the set of all random variables in an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In the following, for any $X \in L^0$, we use F_X for the cdf of X. For any distribution (cdf) F, its generalized inverse (or quantile) F^{-1} is given by

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

Let U_X be a uniform random variable such that $X = F_X^{-1}(U_X)$ almost surely; such U_X is called a *uniform transform* of X and always exists. Let $\mathcal{X}$ be the set of essentially bounded random variables in this probability space, $\mathcal{D}$ the set of distributions of essentially bounded random variables, and $\mathcal{D}_0$ the set of all distributions on $\mathbb{R}$. For $F \in \mathcal{D}_0$, write $\mathcal{B}_F = \{Y \in L^0 : Y \sim F\}$. Throughout this paper n represents a fixed positive integer.

Denote by $\mathcal{H}$ the set of non-decreasing functions $h : [0, 1] \mapsto [0, 1]$ with $h(0) = h(0^+) = 0$ and $h(1^-) = h(1) = 1$. In this paper, a *dual utility* $\mathcal{U}_h : \mathcal{X} \to \mathbb{R}$ with $h \in \mathcal{H}$ is defined as[2]

$$\mathcal{U}_h(X) = \int_0^\infty h(\mathbb{P}(X > x))\,\mathrm{d}x + \int_{-\infty}^0 (h(\mathbb{P}(X > x)) - 1)\,\mathrm{d}x, \quad X \in \mathcal{X}. \tag{2.1}$$

The function h is called a *probability perception function*, a *weighting function* or a *distortion function* in the economic literature.

If at least one of h and F_X^{-1} is continuous, then through an integration by parts and a change of variable, $\mathcal{U}_h$ can be written as[3]

$$\mathcal{U}_h(X) = \int_0^1 F_X^{-1}(1-t)\,\mathrm{d}h(t), \quad X \in \mathcal{X}. \tag{2.2}$$

Recall that two random variables X and Y are *comonotonic* if there exists $\Omega_0 \subseteq \Omega$ with $\mathbb{P}(\Omega_0) = 1$ such that

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0 \ \text{ for all } \omega, \omega' \in \Omega_0.$$

[2]Due to the work of Yaari (1987), the functional $\mathcal{U}_h$ is also often referred to as a *Yaari's dual utility (functional)*.

[3]In the risk management literature, a distortion risk measure is defined, by convention, as $\rho_g(X) = \int_0^1 F_X^{-1}(t)\,\mathrm{d}g(t)$, $X \in \mathcal{X}$; see e.g. McNeil et al. (2015). It is easy to see that the two definitions reconcile if g is taken as the so-called dual of h, namely, $g(t) = 1 - h(1-t)$, $t \in [0, 1]$.

One of the key properties of dual utilities is *comonotonic additivity*; that is, a dual utility is always additive for comonotonic random variables. It is also well known that the following four statements are equivalent: (i) h is convex; (ii) $\mathcal{U}_h$ is concave; (iii) $\mathcal{U}_h$ is superadditive[4]; (iv) $\mathcal{U}_h$ is monotone with respect to the second-order stochastic dominance (SSD). See Lemma 2.1 below for the above assertions.

The main objective of this paper is to study the infimum and the supremum of $\mathcal{U}_h\left(\sum_{i=1}^n Y_i\right)$ over all random variables $Y_1,\dots,Y_n$ with given respective distributions $F_1,\dots,F_n \in D_0$; that is, to find the values of

$$\inf_{Y_i\in\mathcal{B}_{F_i},i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) \tag{2.3}$$

and

$$\sup_{Y_i\in\mathcal{B}_{F_i},i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right). \tag{2.4}$$

Problems (2.3) and (2.4) represent the worst-case and the best-case dual utility of an aggregate risk under dependence uncertainty, respectively. We emphasize that in this paper we do not assume the probability distortion function h to be concave or convex. For instance, h may be inverse-S-shaped (see Section 4.2 below for a precise definition), a prominent feature in behavioral economics and finance (e.g. in CPT and RDUT).

2.2 Properties of dual utilities

For $h \in \mathcal{H}$ and $t \in [0,1]$, define its concave and convex hulls h^* and h_* respectively by

$$h^*(t) = \inf\left\{g(t):\ g:[0,1]\mapsto[0,1],\ g\geqslant h,\ g \text{ is non-decreasing and concave on } [0,1]\right\},$$

$$h_*(t) = \sup\left\{g(t):\ g:[0,1]\mapsto[0,1],\ g\leqslant h,\ g \text{ is non-decreasing and convex on } [0,1]\right\}.$$

Figure 2.1 gives an illustration of h, h^* and h_*, where h is the inverse-S-shaped function in Tversky and Kahneman (1992).

Below we collect a few lemmas on properties of dual utilities which will be useful later in the paper. Lemma 2.1 concerns comonotonic additivity, concavity and superadditivity of dual utilities; Lemma 2.2 gives a simple relationship between $\mathcal{U}_h$, $\mathcal{U}_{h^*}$ and

[4] A functional $\mathcal{U}$ on $\mathcal{X}$ is said to be *superadditive* (resp. *subadditive*) if $\mathcal{U}(X+Y)\geqslant$ (resp. $\leqslant$) $\mathcal{U}(X)+\mathcal{U}(Y)$ for all $X,Y\in\mathcal{X}$.

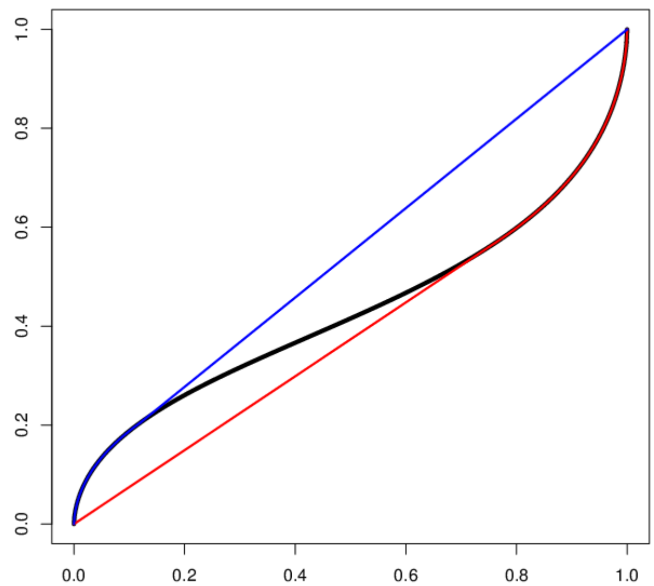

Figure 2.1: h (black), h^* (blue) and h_* (red)

$\mathcal{U}_{h_*}$, and Lemma 2.3 is a classic property of the convex and concave hulls h^* and h_* in relation to h (cf. Figure 2.1).

Lemma 2.1 (Theorems 1-2 of Yaari (1987) and Theorem 4.1 of Acerbi (2002))**.** *For $h \in \mathcal{H}$, $\mathcal{U}_h$ is comonotonic additive. Moreover, the following four statements are equivalent: (i) h is convex; (ii) $\mathcal{U}_h$ is concave; (iii) $\mathcal{U}_h$ is superadditive; (iv) $\mathcal{U}_h$ is monotone with respect to the second-order stochastic dominance (SSD).*

Lemma 2.2 (Lemma 3.1 of Wang et al. (2015))**.** *For any $h \in \mathcal{H}$, we have*

(i) $h^* \in \mathcal{H}$ and $\mathcal{U}_{h^*}$ is the smallest subadditive dual utility dominating $\mathcal{U}_h$;

(ii) $h_* \in \mathcal{H}$ and $\mathcal{U}_{h_*}$ is the largest superadditive dual utility dominated by $\mathcal{U}_h$.

Lemma 2.3 (Lemma 5.1 of Brighi and Chipot (1994))**.** *Suppose that $h \in \mathcal{H}$ is continuous, then the set $\{t \in [0,1] : h(t) \neq h_*(t)\}$ (resp. $\{t \in [0,1] : h(t) \neq h^*(t)\}$) is the union of some disjoint open intervals, and h_* (resp. h^*) is linear on each of the above intervals.*

Before ending this section, we present some well known results in the cases that h is convex or concave. In particular, the case that h is convex corresponds to *strong risk* *aversion* in decision theory (Lemma 2.1 (iv)). The following proposition is straightforward from the above lemmas.

Proposition 2.4. *Let $X_1, \ldots, X_n \in \mathcal{X}$ be random variables with respective distributions $F_1, \ldots, F_n$. If $h \in \mathcal{H}$ is convex, then*

$$\min_{Y_i \in \mathcal{B}_{F_i}, i=1,\ldots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathcal{U}_h(X_i). \tag{2.5}$$

If $h \in \mathcal{H}$ is concave, then

$$\max_{Y_i \in \mathcal{B}_{F_i}, i=1,\ldots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathcal{U}_h(X_i). \tag{2.6}$$

Furthermore, these bounds are attained if $X_1, \ldots, X_n$ are comonotonic.

Proof. By Lemma 2.2, we have the "$\leqslant$" sign in (2.5). Since $\mathcal{U}_h$ is comonotonic additive (Lemma 2.1), we have "$\geqslant$" sign in (2.5). Equation (2.6) is similar. □

For the upper bound in the case of a convex h and the lower bound in the case of a concave h, explicit results are available for $n = 2$, and these bounds are attained by counter-monotonic random variables. Two random variables X_1 and X_2 are said to be *counter-monotonic* if X_1 and $-X_2$ are comonotonic.

Proposition 2.5. *Let X_1 and X_2 be counter-monotonic random variables with respective distributions F_1 and F_2, and $U \sim \mathrm{U}(0,1)$. If $h \in \mathcal{H}$ is convex, then*

$$\max_{Y_i \in \mathcal{B}_{F_i}, i=1,2} \mathcal{U}_h(Y_1 + Y_2) = \mathcal{U}_h(X_1 + X_2) = \mathcal{U}_h(F_1^{-1}(U) + F_2^{-1}(1-U)).$$

If $h \in \mathcal{H}$ is concave, then

$$\min_{Y_i \in \mathcal{B}_{F_i}, i=1,2} \mathcal{U}_h(Y_1 + Y_2) = \mathcal{U}_h(X_1 + X_2) = \mathcal{U}_h(F_1^{-1}(U) + F_2^{-1}(1-U)).$$

Proof. From Lemma 2.1, $\mathcal{U}_h$ is SSD-monotone if h is convex, and among all dependence structures, counter-monotonicity maximizes the sum of two random variables in this order (see e.g. Corollary 3.28 of Rüschendorf (2013)). □

Bounds of the type Proposition 2.5 cannot be generalized to $n \geqslant 3$ since the concept of counter-monotonicity cannot be trivially extended to $n \geqslant 3$; see Puccetti and Wang (2015) for extremal dependence concepts. As mentioned above, result in the case that

h is neither convex or concave is very limited in the literature except for the case that $\mathcal{U}_h$ is a quantile (see Embrechts et al. (2015)). The next few sections are dedicated to the study of this problem for general $h \in \mathcal{H}$ and $n \in \mathbb{N}$.

3 Conditional joint mixability

In this section, we first recall the definition of joint mixability in Wang et al. (2013), and then introduce a concept most relevant to the study of dual utilities under dependence uncertainty.

Definition 3.1. An n-tuple $(F_1, \dots, F_n)$ of probability distributions on $\mathbb{R}$ is *jointly mixable* if there exist n random variables $X_1 \sim F_1, \dots, X_n \sim F_n$ such that $X_1 + \cdots + X_n$ is almost surely a constant.

The determination of whether a given tuple of distributions is jointly mixable is a highly non-trivial task. Some analytically characterized classes of distributions that are jointly mixable are given in Wang and Wang (2011, 2016). In the trivial case of $n = 2$, the joint mixability of (F, G) is equivalent to that F and G are symmetric to each other[5].

Joint mixability has shown to be instrumental in calculating worst-case values for quantile (VaR) aggregation under dependence uncertainty; see e.g. Embrechts et al. (2013). For general dual utility functionals, we generalize this idea to a concept of *conditional joint mixability*. Here and henceforth, for any distribution F and $0 \leqslant a < b \leqslant 1$, we denote by $F^{(a,b)}$ the distribution of $F^{-1}(U)$ where $U \sim \mathrm{U}(a, b)$. Clearly, $F^{(0,1)} = F$.

Definition 3.2 (Conditional joint mixability)**.** Let $\mathcal{I}$ be a set of open subintervals of $(0, 1)$. The n-tuple of distributions $(F_1, \dots, F_n)$ is *$\mathcal{I}$-jointly mixable* if $(F_1^I, \dots, F_n^I)$ is jointly mixable for each interval $I \in \mathcal{I}$.

By convention any n-tuple of distributions is $\varnothing$-jointly mixable (here and hereafter $\varnothing$ denotes the empty set). Obviously, $\{(0, 1)\}$-joint mixability is equivalent to joint mixability defined in Definition 3.1.

[5]The distributions F and G are *symmetric to each other* if there exists $c \in \mathbb{R}$ such that $F(c + x) = 1 - G(c - x)$ for almost everywhere $x \in \mathbb{R}$.

It is not hard to see that conditional joint mixability shares many common properties as joint mixability, such as those in Proposition 2.3 of Wang and Wang (2016). However, since a general analytical determination of joint mixability is still unavailable, so is conditional joint mixability. Below we collect some elementary results on conditional joint mixability.

Proposition 3.1. *Let $\mathcal{I}$ be a set of disjoint open subintervals of $(0,1)$. For any $F \in \mathcal{D}_0$ (resp. $F \in \mathcal{D}$), there exists $G \in \mathcal{D}_0$ (resp. $G \in \mathcal{D}$) such that (F, G) is $\mathcal{I}$-jointly mixable.*

Proof. We only show the case of $F \in \mathcal{D}_0$ since the case of $F \in \mathcal{D}$ is implied from the construction of G below. As $\mathcal{I}$ is a set of disjoint open intervals, it is at most countable. Write $\mathcal{I} = \{I_k : k \in K\}$, $K \subset \mathbb{N}$, $I_k = (a_k, b_k)$, and denote by $I = \bigcup_{k \in K} I_k$. Define for $u \in [0,1]$,

$$T(u) = \sum_{k \in K} (F^{-1}(b_k) + F^{-1}(a_k) - F^{-1}(a_k + b_k - u)) \mathrm{I}_{\{u \in I_k\}} + F^{-1}(u) \mathrm{I}_{\{u \in I^c\}}.$$

Let G be the distribution of $T(U)$ where $U \sim \mathrm{U}(0,1)$. Note that for $k \in K$, T is increasing on I_k and for $u \in I_k$,

$$T(a_k) = F^{-1}(a_k) \leqslant T(u) \leqslant F^{-1}(b_k) = T(b_k).$$

Therefore, T is an increasing function on $[0,1]$, and thus $T = G^{-1}$ almost everywhere on $(0,1)$. One can easily check that for $k \in K$, if $U_k \sim \mathrm{U}(a_k, b_k)$, then $F^{-1}(b_k) + F^{-1}(a_k) - F^{-1}(a_k + b_k - U_k) \in \mathcal{B}_{G^{I_k}}$ and $F^{-1}(a_k + b_k - U_k) \in \mathcal{B}_{F^{I_k}}$. This shows that (F^{I_k}, G^{I_k}) is jointly mixable for each $k \in K$. □

Joint mixability and conditional joint mixability can be explicitly characterized for uniform distributions.

Proposition 3.2. *Suppose that for $i = 1, \ldots, n$, F_i is a uniform distribution on an interval of length $a_i > 0$. The following statements are equivalent.*

(i) $\sum_{i=1}^{n} a_i \geqslant 2 \max\{a_i : i = 1, \ldots, n\}$.

(ii) $(F_1, \ldots, F_n)$ *is jointly mixable.*

(iii) $(F_1, \ldots, F_n)$ *is $\mathcal{I}$-jointly mixable for all sets $\mathcal{I}$ of open subintervals of* $(0,1)$.

(iv) $(F_1, \dots, F_n)$ *is* $\mathcal{I}$*-jointly mixable for some non-empty set* $\mathcal{I}$ *of open subintervals of* $(0,1)$.

Proof. By Theorem 3.1 of Wang and Wang (2016), an n-tuple of uniform distributions with lengths $b_1, \dots, b_n > 0$ is jointly mixable if and only if $\sum_{i=1}^n b_i \geqslant 2\max\{b_i : i = 1, \dots, n\}$. Note that for each $I \in \mathcal{I}$ with length $|I|$, $(F_1^I, \dots, F_n^I)$ is an n-tuple of uniform distributions with lengths $a_1|I|, \dots, a_n|I|$. From there, the implications (ii)⇔(i)⇒(iii)⇒(iv)⇒(i) are all straightforward. □

4 Dual utilities with dependence uncertainty

In this section we partially solve the problems (2.3) and (2.4) with the help of conditional joint mixability. We first present some general results, followed by a special case when the distortion functions are inverse-S-shaped.

4.1 General results

Recall that the set $\mathcal{X}$ consists of essentially bounded random variables; so $\mathcal{U}_h(X)$ is properly defined for all $h \in \mathcal{H}$ and $X \in \mathcal{X}$. We first show that if $h \in \mathcal{H}$ is continuous, then the infimum in (2.3) can always be attained and thus it is a minimum. Similar for the supremum in (2.4).

Lemma 4.1. *Suppose that* $h \in \mathcal{H}$ *is continuous and* $F_1, \dots, F_n \in \mathcal{D}$*. Then there exist* $X_i \in \mathcal{B}_{F_i}$*,* $i = 1, \dots, n$*, such that*

$$\mathcal{U}_h\left(\sum_{i=1}^n X_i\right) = \min_{Y_i \in \mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right). \tag{4.1}$$

Proof. Denote by $\mathcal{S} = \{X_1 + \dots + X_n : X_i \in \mathcal{B}_{F_i},\ i = 1, \dots, n\}$. We use two facts about the set $\mathcal{S}$. First, $\mathcal{U}_h$ is continuous on $\mathcal{S}$ with respect to weak convergence, due to Theorem 2.4 of Embrechts et al. (2015). Second, for any sequence $T_1, T_2, \dots \in \mathcal{S}$, there exists a subsequence $T_{k_1}, T_{k_2}, \dots$ and $T \in \mathcal{S}$ such that $T_{k_j} \to T$ weakly as $j \to \infty$; see Theorem 2.1 (vii) of Bernard et al. (2014) and its proof. Combining the above two facts establishes the desired result. □

Note that the attainability in Lemma 4.1 is generally not true for a non-continuous distortion function h, as illustrated by the following example.

Example 4.1 (Non-attainability)**.** For $p \in (0,1)$, let $h(t) = \mathrm{I}_{\{t \geqslant 1-p\}}$, $t \in [0,1]$. By definition in (2.1), the functional $\mathcal{U}_h : \mathcal{X} \to \mathbb{R}$ is the right-p-quantile, namely,

$$\begin{aligned}\mathcal{U}_h(X) &= \int_0^\infty \mathrm{I}_{\{\mathbb{P}(X>x)\geqslant 1-p\}}\,\mathrm{d}x - \int_{-\infty}^0 \mathrm{I}_{\{\mathbb{P}(X>x)<1-p\}}\,\mathrm{d}x \\ &= \inf\{x \in \mathbb{R} : \mathbb{P}(X > x) < 1-p\}.\end{aligned}$$

Let F be a uniform distribution on $[0,1]$. Then by Lemma 4.5 of Bernard et al. (2014) and Proposition 8.31 of McNeil et al. (2015),

$$\inf_{Y_1,Y_2\in\mathcal{B}_F} \mathcal{U}_h\left(Y_1+Y_2\right) = \inf_{x\in[0,p]}\{F^{-1}(x) + F^{-1}(p-x)\} = p.$$

Now we verify that for any $X_1, X_2 \sim F$ one has $\mathcal{U}_h(X_1+X_2) > p$. Suppose otherwise, that is, $\inf\{x \in \mathbb{R} : F_{X_1+X_2}(x) > p\} = p$. Because the cdf $F_{X_1+X_2}$ is right-continuous, we have $F_{X_1+X_2}(p) \geqslant p$. Note that $\{X_1 + X_2 \leqslant p\} \subset \{X_i \leqslant p\}$ and $\mathbb{P}(X_i \leqslant p) = p \leqslant \mathbb{P}(X_1+X_2 \leqslant p)$, $i = 1,2$. It follows that $\{X_1+X_2 \leqslant p\} = \{X_1 \leqslant p\} = \{X_2 \leqslant p\}$ almost surely, and as a consequence, $\{X_1 + X_2 > p\} = \{X_1 > p\} = \{X_2 > p\}$ almost surely. This implies $\mathbb{P}(X_1 + X_2 > 2p) = \mathbb{P}(X_1 > p) = 1-p$, and hence $\mathbb{P}(X_1 + X_2 \leqslant 2p) = p$. Finally,

$$\mathcal{U}_h(X_1+X_2) = \inf\left\{x \in \mathbb{R} : \mathbb{P}(X_1+X_2 \leqslant x) > p\right\} \geqslant 2p.$$

Thus, the minimum in (4.1) is not attainable.

For any continuous, non-concave distortion function $h \in \mathcal{H}$, from Lemma 2.3, there exist (countably many) disjoint open intervals on which $h \neq h_*$. For $h \in \mathcal{H}$, write $g(t) = 1 - h(1-t)$, $g_1(t) = 1 - h_*(1-t)$ and $g_2(t) = 1 - h^*(1-t)$, $t \in [0,1]$. Denote by $\mathcal{I}^h$ the set of disjoint open intervals on which $g \neq g_1$. In other words, $\mathcal{I}^h$ contains intervals $(1-b, 1-a)$ where (a,b) is an open interval on which $h \neq h_*$. Similarly, denote by $\mathcal{J}^h$ the set of disjoint open intervals on which $g \neq g_2$. The following theorem contains our main technical result.

Theorem 4.2. *Suppose that $h \in \mathcal{H}$ and $X_1, \dots, X_n \in \mathcal{X}$ with respective distributions $F_1, \dots, F_n$.*

(i) We have

$$\sum_{i=1}^n \mathcal{U}_{h_*}(X_i) \leqslant \inf_{Y_i\in\mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right)$$
$$\leqslant \sup_{Y_i\in\mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) \leqslant \sum_{i=1}^n \mathcal{U}_{h^*}(X_i). \tag{4.2}$$

(ii) If h is continuous on $[0,1]$ and $(F_1,\dots,F_n)$ is $\mathcal{I}^h$-jointly mixable, then

$$\min_{Y_i\in\mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathcal{U}_{h_*}(X_i). \tag{4.3}$$

(iii) If h is continuous on $[0,1]$ and $(F_1,\dots,F_n)$ is $\mathcal{J}^h$-jointly mixable, then

$$\max_{Y_i\in\mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathcal{U}_{h^*}(X_i). \tag{4.4}$$

Proof. We only show (i) and (ii), as (iii) is symmetric to (ii).

(i) From Lemma 2.2, we have

$$\sum_{i=1}^n \mathcal{U}_{h_*}(X_i) \leqslant \mathcal{U}_{h_*}\left(\sum_{i=1}^n X_i\right) \leqslant \mathcal{U}_h\left(\sum_{i=1}^n X_i\right) \leqslant \mathcal{U}_{h^*}\left(\sum_{i=1}^n X_i\right) \leqslant \sum_{i=1}^n \mathcal{U}_{h^*}(X_i).$$

By taking infimum and supremum in the above inequality over random variables with distributions respective distributions $F_1,\dots,F_n$, we obtain (4.2).

(ii) Since h is continuous, we directly work with (2.2). Write $\mathcal{I}^h = \{I_k,\ k\in K\}$ where $K\subset\mathbb{N}$ and $I_k=(a_k,b_k)$, $k\in K$. Since $(F_1,\dots,F_n)$ is $\mathcal{I}^h$-jointly mixable, for each $k\in K$ there exist random variables $Y_{1k},\dots,Y_{nk}$, such that Y_{ik} is distributed as $F_i^{I_k}$, $i=1,\dots,n$, and $Y_{1k}+\dots+Y_{nk}$ is a constant μ_k where

$$\mu_k = \sum_{i=1}^n \mathbb{E}[Y_{ik}] = \sum_{i=1}^n \frac{\int_{(a_k,b_k)} F_i^{-1}(t)\mathrm{d}t}{b_k-a_k},\ \ k\in K. \tag{4.5}$$

Take $U\sim \mathrm{U}(0,1)$ independent of Y_{ik}, $i=1,\dots,n$, $k\in K$. It is harmless to assume the existence of U, as we have only imposed some requirement on the distributions of Y_{ik}, $i=1,\dots,n$, $k\in K$. Write

$$S_n = F_1^{-1}(U)+\dots+F_n^{-1}(U). \tag{4.6}$$

and

$$Y_i^* = F_i^{-1}(U)\mathrm{I}_{\{U\notin\cup_{k\in K}I_k\}} + \sum_{k\in K} Y_{ik}\mathrm{I}_{\{U\in I_k\}}, \quad i=1,\dots,n.$$

It is easy to check that $Y_i^* \in \mathcal{B}_{F_i}$, $i=1,\dots,n$, and due to the comonotonic additivity of $\mathcal{U}_{h_*}$ in Lemma 2.1,

$$\mathcal{U}_{h_*}(S_n) = \sum_{i=1}^n \mathcal{U}_{h_*}(F_i^{-1}(U)) = \sum_{i=1}^n \mathcal{U}_{h_*}(X_i).$$

Denote by $R_n = Y_1^* + \cdots + Y_n^*$ and write

$$G(u) = \begin{cases} F_1^{-1}(u) + \cdots + F_n^{-1}(u), & u \in (0,1) \setminus (\cup_{k\in K} I_k), \\ \mu_k, & u \in I_k,\ k \in K. \end{cases} \tag{4.7}$$

Clearly $R_n = G(U)$. Noting that G is an increasing function on $(0,1)$, we have $G(t) = F_{R_n}^{-1}(t)$ for almost every $t \in (0,1)$. Write $g(t) = 1 - h(1-t)$ and $g_1(t) = 1 - h_*(1-t)$, $t \in [0,1]$. From the fact that g_1 is linear on each I_k, $k \in K$, and

$$F_{S_n}^{-1}(t) = \sum_{i=1}^n F_i^{-1}(t) \quad \text{for } t \in (0,1),$$

we obtain

$$\begin{aligned} &\int_{(a_k,b_k)} F_{S_n}^{-1}(t)\mathrm{d}g_1(t) - \int_{(a_k,b_k)} F_{R_n}^{-1}(t)\mathrm{d}g_1(t) \\ &= \frac{g_1(b_k)-g_1(a_k)}{b_k-a_k} \sum_{i=1}^n \int_{(a_k,b_k)} F_i^{-1}(t)\mathrm{d}t - \sum_{i=1}^n \int_{(a_k,b_k)} \mu_k \mathrm{d}g_1(t) = 0. \end{aligned}$$

It follows that

$$\begin{aligned} \mathcal{U}_{h_*}(S_n) - \mathcal{U}_{h_*}(R_n) &= \int_0^1 F_{S_n}^{-1}(1-t)\mathrm{d}h_*(t) - \int_0^1 F_{R_n}^{-1}(1-t)\mathrm{d}h_*(t) \\ &= \int_0^1 F_{S_n}^{-1}(t)\mathrm{d}g_1(t) - \int_0^1 F_{R_n}^{-1}(t)\mathrm{d}g_1(t) \\ &= \sum_{k\in K}\left[\int_{a_k}^{b_k} F_{S_n}^{-1}(t)\mathrm{d}g_1(t) - \int_{a_k}^{b_k} F_{R_n}^{-1}(t)\mathrm{d}g_1(t)\right] = 0, \end{aligned} \tag{4.8}$$

that is, $\mathcal{U}_{h_*}(S_n) = \mathcal{U}_{h_*}(R_n)$. Integration by parts yields

$$\begin{aligned} \mathcal{U}_{h_*}(R_n) - \mathcal{U}_h(R_n) &= \int_0^1 F_{R_n}^{-1}(t)\mathrm{d}g_1(t) - \int_0^1 F_{R_n}^{-1}(t)\mathrm{d}g(t) \\ &= \int_0^1 (g(t) - g_1(t))\,\mathrm{d}F_{R_n}^{-1}(t) \\ &= \sum_{k\in K} \int_{(a_k,b_k)} (g(t) - g_1(t))\,\mathrm{d}F_{R_n}^{-1}(t) = 0, \end{aligned} \tag{4.9}$$

where the last equality follows since $F^{-1}_{R_n}(t)$ is a constant on (a_k, b_k).

Therefore,

$$\mathcal{U}_h(R_n) = \mathcal{U}_{h_*}(S_n) = \sum_{i=1}^{n} \mathcal{U}_{h_*}(X_i), \tag{4.10}$$

and from part (i), we obtain (4.3). □

Remark 4.1. Theorem 4.2 implies the classic results for convex or concave h in Proposition 2.4. If h is convex, then it is continuous and equal to h_*. Moreover, $\mathcal{I}^h$ is an empty set, hence $(F_1, \ldots, F_n)$ is always $\mathcal{I}^h$-jointly mixable. Therefore, (4.3) is equivalent to (2.5). The case of a concave h is similar.

Remark 4.2. Assuming (4.3) holds, the dependence structure that attains the bound is generally not unique. Indeed, for $n \geqslant 3$ and a tuple of marginal distributions, there can be many dependence structures that leads to a constant sum. Another question is whether the conditional joint mixability is necessary for (4.3) to hold. In Theorem 4.4 below, we show that, for strict inverse-S-shaped distortion functions, this condition is indeed necessary and sufficient for (4.3) to hold. In general, such an equivalence may fail to hold; see Remark 4.5 for further discussions.

For general distributions, the $\mathcal{I}^h$- and $\mathcal{J}^h$-joint mixability in Theorem 4.2 is not easy to verify. One particularly useful result for this verification is Theorem 3.2 of Wang and Wang (2016) on the joint mixability of an n-tuple of distributions with increasing (or decreasing) densities: If for each $i = 1, \ldots, n$, F_i has mean μ_i and an increasing density on its finite support $[a_i - l_i, a_i]$, then $(F_1, \ldots, F_n)$ is jointly mixable if and only if

$$\sum_{i=1}^{n} a_i - \sum_{i=1}^{n} \mu_i \geqslant \max_{i=1,\ldots,n} l_i. \tag{4.11}$$

Condition (4.11) is straightforward to check and will be used in Section 5 later.

Remark 4.3. Throughout this paper we work with the set $\mathcal{X}$ of essentially bounded random variables so that all dual utilities $\mathcal{U}_h$, $\mathcal{U}_{h_*}$ and $\mathcal{U}_{h^*}$ are properly defined on the same domain $\mathcal{X}$. If one likes to study $\mathcal{U}_h$ defined on a set $\mathcal{Y}$ larger than $\mathcal{X}$, then it has to be assumed that both $\mathcal{U}_{h_*}(Y)$ and $\mathcal{U}_{h^*}(Y)$ are finite for all $Y \in \mathcal{Y}$. In that case, the results in Theorem 4.2 are still valid based on the same proof.

Remark 4.4. The asymptotics as $n \to \infty$ of the worst-case aggregate values of dual utilities (under the term *distortion risk measures*) has been studied in Wang et al. (2015) and Cai et al. (2018) under different settings. In particular, Cai et al. (2018) showed that the ratio

$$\frac{\sup\left\{\mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) : Y_i \in \mathcal{B}_{F_i},\ i = 1, \dots, n\right\}}{\sum_{i=1}^n \mathcal{U}_{h^*}(X_i)}$$

is 1 asymptotically as $n \to \infty$, under some regularity conditions on the distributions $F_1, F_2, \dots$ and the distortion function h. In other words, for large n, the bounds in Theorem 4.2 provide reasonable approximations for the values of (2.3) and (2.4), even when the assumption of conditional joint mixability is violated.

Example 4.2 (Uniform risks)**.** Suppose that $h \in \mathcal{H}$ is continuous and for $i = 1, \dots, n$, F_i is a uniform distribution on an interval $[b_i, b_i + a_i]$, $a_i > 0$, and $\sum_{i=1}^n a_i \geqslant 2\max\{a_i : i = 1, \dots, n\}$. The $\mathcal{I}^h$-joint mixability of $(F_1, \dots, F_n)$ is given by Proposition 3.2. For a $\mathcal{U}_h$-decision maker, the worst-case utility for an aggregation with marginal distributions $F_1, \dots, F_n$ is given by

$$\min_{Y_i \in \mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \left(a_i \int_0^1 (1-t)\, \mathrm{d}h_*(t) + b_i\right).$$

4.2 Inverse-S-shaped distortion functions

An inverse-S-shaped function is arguably the most prominent example of probability distortion function in a dual utility used in behavioral economics and finance (such as in CPT or RDUT); see Tversky and Kahneman (1992) and Prelec (1998).

Definition 4.1. A distortion function $h \in \mathcal{H}$ is *inverse-S-shaped* if h is twice differentiable on $(0, 1)$, and there exists $t_h \in [0, 1]$ such that h is concave on $[0, t_h]$ and h is convex on $[t_h, 1]$. Moreover, h is *strict* if it is strictly convex on $[0, t_h]$ and strictly convex on $[t_h, 1]$ for some $t_h \in [0, 1]$.

A typical inverse-S-shaped distortion function is presented in Figure 2.1. The following proposition can be easily verified from Lemma 2.3.

Proposition 4.3. *For an inverse-S-shaped distortion function h, there exist $p_h, q_h \in [0, 1]$ such that* *(i)* $h = h_*$ *on* $[p_h, 1]$*,* $h > h_*$ *on* $(0, p_h)$ *and* h_* *is linear on* $[0, p_h]$*;*

(ii) $h = h^*$ *on* $[0, q_h]$*,* $h < h^*$ *on* $(q_h, 1)$ *and* h^* *is linear on* $[q_h, 1]$*.*

It is immediate that $p_h \leqslant t_h \leqslant q_h$, h is concave on $[0, q_h]$ and h is convex on $[p_h, 1]$. Note that t_h is in general not unique, unless we assume h is strict. On the other hand, p_h and q_h are unique for a given inverse-S-shaped h.

For inverse-S-shaped $h \in \mathcal{H}$, we have $\mathcal{I}^h = \{(1-p_h, 1)\}$ and $\mathcal{J}^h = \{(0, 1-q_h)\}$. In the following, if $p_h = 0$ (resp. $q_h = 1$), we interpret $\{(1-p_h, 1)\} = \varnothing$ (resp. $\{(0, 1-q_h)\} = \varnothing$). Using Theorem 4.2, for $F_1, \dots, F_n \in \mathcal{D}$, and $X_i \in \mathcal{B}_{F_i}$, $i = 1, \dots, n$,

$$\min_{Y_i \in \mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) \geqslant \sum_{i=1}^n \mathcal{U}_{h_*}(X_i), \tag{4.12}$$

and

$$\max_{Y_i \in \mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) \leqslant \sum_{i=1}^n \mathcal{U}_{h^*}(X_i). \tag{4.13}$$

Moreover, the inequalities in (4.12) and (4.13) and hold as equalities under conditional joint mixability specified in Theorem 4.2. In the following theorem, we establish a stronger claim: for a strict inverse-S-shaped distortion function h, the above bounds are attainable only if conditional joint mixability holds. This result suggests that conditional joint mixability is essential to the problem of dual utilities with dependence uncertainty, and under some conditions it gives the only possible dependence structure that attains the bounds.

Theorem 4.4. *Suppose that* $h \in \mathcal{H}$ *is inverse-S-shaped and strict,* $F_1, \dots, F_n \in \mathcal{D}$*, and* $X_i \in \mathcal{B}_{F_i}$*,* $i = 1, \dots, n$*. The inequality in* (4.12) *holds as an equality if and only if* $(F_1, \dots, F_n)$ *is* $\{(1-p_h, 1)\}$*-jointly mixable, and the inequality in* (4.13) *holds as an equality if and only if* $(F_1, \dots, F_n)$ *is* $\{(0, 1-q_h)\}$*-jointly mixable.*

Proof. Theorem 4.2 gives the "if" statements. We only show the first "only-if" statement: the inequality in (4.12) is an equality only if $(F_1, \dots, F_n)$ is $\{(1-p_h, 1)\}$-jointly mixable. The second "only-if" statement is analogous.

There is nothing to show if $p_h = 0$, and below we assume $p_h > 0$. Denote by I the interval $(1-p_h, 1)$. Suppose that there exist $Y_i \in \mathcal{B}_{F_i}, i = 1, \dots, n$ such that they attain the minimum

$$\mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathcal{U}_{h_*}(Y_i). \tag{4.14}$$

Denote by $T = \sum_{i=1}^n Y_i$. Note that (4.14) implies

$$\sum_{i=1}^n \mathcal{U}_{h_*}(Y_i) = \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) \geqslant \mathcal{U}_{h_*}\left(\sum_{i=1}^n Y_i\right) \geqslant \sum_{i=1}^n \mathcal{U}_{h_*}(Y_i).$$

and hence $\mathcal{U}_h(T) = \mathcal{U}_{h_*}(T)$. By definition of $\mathcal{U}_h$ in (2.1), we have

$$\int_{-\infty}^{\infty} (h(\mathbb{P}(T > x)) - h_*(\mathbb{P}(T > x)))\,\mathrm{d}x = 0.$$

As $h \geqslant h_*$, we have $h(\mathbb{P}(T > x)) = h_*(\mathbb{P}(T > x))$ a.e. $x \in \mathbb{R}$. Further, by Proposition 4.3, $h(t) > h_*(t)$ for $t \in (0, p_h)$. Therefore, $\mathbb{P}(T > x)$ does not take values in $(0, p_h)$. As a consequence, there exists $M \in \mathbb{R}$ such that $\mathbb{P}(T \leqslant M) = 1$ and $\mathbb{P}(T = M) \geqslant p_h$.

For $i = 1, \ldots, n$, let $A_i \in \mathcal{F}$ be an event of probability p_h on which Y_i takes its largest possible values (that is, $A_i = \{U_{Y_i} > 1 - p_h\}$ a.s. for a uniform transform U_{Y_i} of Y_i), and let $A \in \mathcal{F}$ be an event of probability p_h on which T takes its largest possible values. Note that $Y_i \sim F_i^I$ conditional on A_i, $i = 1, \ldots, n$. Denote by

$$M' = \sum_{i=1}^n \mathbb{E}[Y_i|A_i] = \sum_{i=1}^n \int_I F_i^{-1}(t)\,\mathrm{d}t.$$

Note that since $\mathbb{P}(T = M) \geqslant p_h$, we have

$$M = \mathbb{E}[T|A] = \sum_{i=1}^n \mathbb{E}[Y_i|A] \leqslant \sum_{i=1}^n \mathbb{E}[Y_i|A_i] = M'.$$

Assume for the moment that $M = M'$. We have, for each i, $\mathbb{E}[Y_i|A] = \mathbb{E}[Y_i|A_i]$, and as a consequence, Y_i also takes its largest possible values on A. Hence, $Y_i \sim F_i^I$ conditional on A. Since $Y_1 + \cdots + Y_n = M$ conditional on A, we conclude that $(F_1, \ldots, F_n)$ is I-jointly mixable. Thus, our desired statement holds if $M = M'$. It remains to show $M = M'$, which we illustrate below.

Let the function G be given by, as in (4.7),

$$G(u) = \begin{cases} F_1^{-1}(u) + \cdots + F_n^{-1}(u), & u \in (0, 1 - p_h], \\ M', & u \in (1 - p_h, 1] \end{cases} \tag{4.15}$$

and $R = G(U)$, where $U \sim \mathrm{U}(0,1)$. By (4.10), $\mathcal{U}_{h_*}(R) = \mathcal{U}_{h_*}(T)$. Let $\leqslant_{\mathrm{cx}}$ and $\leqslant_{\mathrm{mps}}$ denote *convex order* and *mean-preserving spread order* in Definitions 1.5.1 and 1.5.25

of Müller and Stoyan (2002), respectively. Write $S=\sum_{i=1}^n F_i^{-1}(U)$. As comonotonicity maximizes convex order of the sum, we have $T\leqslant_{\mathrm{cx}} S$, which implies $\mathbb{E}[(T-x)_-]\geqslant \mathbb{E}[(S-x)_-]$ for all $x\in\mathbb{R}$. In particular, since $S=R$ for $U\leqslant 1-p_h$, and $\mathbb{E}[T]=\mathbb{E}[S]=\mathbb{E}[R]$, for all $x\leqslant F_1^{-1}(1-p_h)+\cdots+F_n^{-1}(1-p_h)$, we have $\mathbb{E}[(T-x)_-]\geqslant \mathbb{E}[(R-x)_-]$, and equivalently, $\mathbb{E}[(T-x)_+]\leqslant \mathbb{E}[(R-x)_+]$. For $x>F_1^{-1}(1-p_h)+\cdots+F_n^{-1}(1-p_h)$, it is clear that $\mathbb{E}[(R-x)_+]=\mathbb{E}[(M'-x)_+]\geqslant \mathbb{E}[(M-x)_+]\geqslant \mathbb{E}[(T-x)_+]$. Therefore, $\mathbb{E}[(T-x)_+]\leqslant \mathbb{E}[(R-x)_+]$ for all $x\in\mathbb{R}$. By Theorem 1.5.7 of Müller and Stoyan (2002), together with the fact that $\mathbb{E}[T]=\mathbb{E}[R]$, we have $T\leqslant_{\mathrm{cx}} R$.

By Theorem 1.5.27 of Müller and Stoyan (2002), there exists a sequence $\{T_1,T_2,\dots\}$ with $T_k\leqslant_{\mathrm{mps}} T_{k+1}$ for all k, such that $T=T_1$ and $T_k\to R$ in distribution. Since $\mathcal{U}_{h_*}$ is SSD-monotone (Lemma 2.1), and $\leqslant_{\mathrm{mps}}$ is stronger than $\leqslant_{\mathrm{cx}}$ (which implies the oppose direction of SSD), we have

$$\mathcal{U}_{h_*}(T)=\mathcal{U}_{h_*}(T_1)\geqslant \mathcal{U}_{h_*}(T_2)\geqslant\ldots\geqslant \mathcal{U}_{h_*}(R)=\mathcal{U}_{h_*}(T).$$

Therefore, $\mathcal{U}_{h_*}(T_k)=\mathcal{U}_{h_*}(T_{k+1})$.

By definition of $\leqslant_{\mathrm{mps}}$, there exists $x_0\in\mathbb{R}$ such that $F_{T_k}(x)\leqslant F_{T_{k+1}}(x)$ for $x<x_0$ and $F_{T_k}(x)\geqslant F_{T_{k+1}}(x)$ for $x>x_0$. Inverting the distribution functions, there exists $t_0\in(0,1)$ such that $F_{T_k}^{-1}(t)\geqslant F_{T_{k+1}}^{-1}(t)$ for $t<t_0$ and $F_{T_k}^{-1}(t)\leqslant F_{T_{k+1}}^{-1}(t)$ for $t>t_0$. Let $\Phi(t)=F_{T_{k+1}}^{-1}(t)-F_{T_k}^{-1}(t)$, $t\in(0,1)$. Then $\Phi(t)\leqslant 0$ for $t\in(0,t_0)$ and $\Phi(t)\geqslant 0$ for $t\in(t_0,1)$.

Since T_k and T_{k+1} have the same mean because $T_k\leqslant_{\mathrm{mps}} T_{k+1}$, we have $\int_0^1\Phi(t)\,\mathrm{d}t=\mathbb{E}[T_{k+1}]-\mathbb{E}[T_k]=0$, and hence,

$$-\int_0^{t_0}\Phi(t)\,\mathrm{d}t=\int_{t_0}^1\Phi(t)\,\mathrm{d}t. \tag{4.16}$$

On the other hand, $\mathcal{U}_{h_*}(T_k)=\mathcal{U}_{h_*}(T_{k+1})$ implies

$$-\int_0^{t_0}\Phi(t)h_*'(1-t)\,\mathrm{d}t=\int_{t_0}^1\Phi(t)h_*'(1-t)\,\mathrm{d}t, \tag{4.17}$$

where h_*' is the left-derivative of h_*. Let $\xi=\inf_{t\in(0,t_0)}h_*'(1-t)$. Putting (4.16) and (4.17) together, we have

$$-\int_0^{t_0}\Phi(t)(h_*'(1-t)-\xi)\,\mathrm{d}t=\int_{t_0}^1\Phi(t)(h_*'(1-t)-\xi)\,\mathrm{d}t. \tag{4.18}$$

Note that the left-hand side of (4.18) is non-negative since $h_*'(1-t) \geqslant \xi$ and $\Phi(t) \leqslant 0$ for $t \in (0, t_0)$, and the right-hand side of (4.18) is non-positive since $h_*'(1-t) \leqslant \xi$ and $\Phi(t) \geqslant 0$ for $t \in (t_0, 1)$. Therefore, we have

$$\int_0^{t_0} \Phi(t)(h_*'(1-t) - \xi)\,\mathrm{d}t = 0 = \int_{t_0}^1 \Phi(t)(h_*'(1-t) - \xi)\,\mathrm{d}t. \tag{4.19}$$

As h (and h_*) is strictly convex on $(0, p_h)$, we know that $h_*'(t) > h_*'(s)$ a.e. for $t \in (p_h, 1)$ and $s \in (0, p_h)$. There are two possibilities:

(a) $t_0 \leqslant 1 - p_h$. In this case, $h_*'(1-t) < \xi$ for $t \in (1-p_h, 1)$, and (4.19) implies $\Phi(t) = 0$ for $t \in (1-p_h, 1)$.

(b) $t_0 > 1 - p_h$. In this case, $h_*'(1-t) > \xi$ for $t \in (0, 1-p_h)$, and (4.19) implies $\Phi(t) = 0$ for $t \in (0, 1-p_h)$.

In both cases, by (4.16), we have $\int_{1-p_h}^1 \Phi(t) = 0$. Therefore, for all $k \in \mathbb{N}$,

$$\int_{1-p_h}^1 F_{T_k}^{-1}(t)\,\mathrm{d}t = \int_{1-p_h}^1 F_{T_k+1}^{-1}(t)\,\mathrm{d}t. \tag{4.20}$$

Note that $T_k \leqslant_{\mathrm{cx}} R$ implies $T_k \leqslant M'$ a.s. Since $F_{T_k}^{-1} \to F_R^{-1} = M$ a.e. and $F_{T_k}^{-1}$ is bounded above by M', the Bounded Convergence Theorem gives, as $k \to \infty$,

$$p_h M = \int_{1-p_h}^1 F_T^{-1}(t) = \int_{1-p_h}^1 F_{T_k}^{-1}(t)\,\mathrm{d}t \to \int_{1-p_h}^1 F_R^{-1}(t)\,\mathrm{d}t = p_h M'.$$

That is, $M = M'$ and the proof is complete. □

As a special case of Theorem 4.4, if h is strictly convex, then

$$\max_{Y_i \in \mathcal{B}_{F_i}, i=1,\dots,n} \mathcal{U}_h\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}[Y_i]. \tag{4.21}$$

if and only if $(F_1, \dots, F_n)$ is jointly mixable. This fact can be verified directly. Intuitively, for a risk-averse decision maker, the best-possible risk given its mean is a constant.

Remark 4.5. As we can see from the proof of Theorem 4.4, the strictness assumption of h is used to assure $h_*'(t) > h_*'(s)$ a.e. for $t \in (p_h, 1)$ and $s \in (0, p_h)$. Hence, this assumption can be weakened to that h_* is not linear in a neighbourhood of p_h. If h_* is indeed linear in a neighbourhood (a, b) of p_h, using the same proof of Theorem 4.2, we can see that the bound (4.12) is attainable if $(F_1, \ldots, F_n)$ is $\{(1-b,1)\}$-jointly mixable but not necessarily $\{(1-p_h,1)\}$-jointly mixable. Note that $\{(1-b,1)\}$- and $\{(1-p_h,1)\}$-joint mixability do not imply each other in general. Hence, such an assumption is not dispensable.

Remark 4.6. Commonly used distortion functions in decision theory are strict. For instance, the famous probability perception functions (distortion functions) introduced by Tversky and Kahneman (1992) and Prelec (1998) are both strict.

5 The pessimism effect

In this section we present an application of our main results to decision theory in terms of what we call the *pessimism effect*. Let us first look at an inverse-S-shaped distortion function as in Figure 2.1. To compare h and h_*, one can see that h puts high weights on both extreme gains and extreme losses, whereas h_* puts the same high weight on the losses but much less weight for gains. Chateauneuf et al. (2005) define the index of pessimism for $\mathcal{U}_h$ as

$$P_h = \inf_{v \in (0,1)} \left\{ \frac{(1-h(v))v}{h(v)(1-v)} \right\}.$$

It is clear from the definition that $P_{h_*} \geqslant P_h$ as $h_* \leqslant h$. Therefore, a $\mathcal{U}_{h_*}$-decision maker is more pessimistic (i.e. with a greater index of pessimism) than a $\mathcal{U}_h$-decision maker, and for this statement to hold we do not need to restrict ourselves to inverse-S-shaped distortion functions.

From Theorem 4.2, we have seen that the two dual utilities $\mathcal{U}_h$ and $\mathcal{U}_{h_*}$ are already very well connected in the presence of dependence uncertainty. Below we make this connection more precise from a decision analysis perspective, via two concrete examples. We will see that a $\mathcal{U}_h$-decision maker with dependence uncertainty may act like the (artificial) concave $\mathcal{U}_{h_*}$-decision maker without dependence uncertainty; thus, conceptually, the fear for uncertainty translates into a more pessimistic preference.

Example 5.1. The first example concerns insurance risks with a truncated Pareto distribution. For $\theta > 1$, let the distribution of a random variable X_θ be given by

$$\mathbb{P}(X_\theta \leqslant x) = \begin{cases} 0, & x < 1, \\ 1 - x^{-2}, & 1 \leqslant x < \theta, \\ 1, & x \geqslant \theta. \end{cases}$$

That is, X_θ has a truncated Pareto(2) distribution with truncation point θ. Let F_θ be the distribution of $-X_\theta$. The distribution F_θ is often used to model insurance losses with a limit θ in the payment (e.g. in auto insurance). It is straightforward that

$$F_\theta^{-1}(t) = \begin{cases} -\theta, & 0 < t \leqslant \theta^{-2}, \\ -t^{-1/2}, & \theta^{-2} < t \leqslant 1. \end{cases}$$

Suppose that $h \in \mathcal{H}$ is inverse-S-shaped with $p_h = 0.84$ (see e.g. Figure 2.1)[6], and an insurer A with utility functional $\mathcal{U}_h$ insures n risks $Y_1, \ldots, Y_n$ identically distributed as F_θ with uncertain dependence. Insurer A collects a premium $p(\theta)$ for each risk she insures, in which p is an increasing function determined by the market. Her problem is to design an insurance policy, i.e. to determine θ over $\theta \in [2.5, \infty)$ to maximize her utility.

For given $\theta \in [2.5, \infty)$, the utility for A under dependence uncertainty is

$$\min_{Y_i \in \mathcal{B}_{F_\theta}, i=1,\ldots,n} \mathcal{U}_h \left(np(\theta) + \sum_{i=1}^n Y_i \right).$$

Note that $p_h \geqslant \theta^{-2}$, and $F_\theta^{(1-p_h,1)}$ is the conditional distribution of F_θ on $[F_\theta^{-1}(1 - p_h), -1] = [-2.5, -1]$, which has a mean of

$$\mu = \frac{1}{p_h} \int_{1-p_h}^{1} -t^{-1/2} \, \mathrm{d}t = -2 \frac{1 - (1 - p_h)^{1/2}}{p_h} = -\frac{10}{7}.$$

Because $F_\theta^{(1-p_h,1)}$ has an increasing density on its support, the joint mixability of the n-tuple $(F_\theta^{(1-p_h,1)}, \ldots, F_\theta^{(1-p_h,1)})$ is equivalent to condition (4.11), which in the present case is

$$n(F_\theta^{-1}(1) - \mu) \geqslant F_\theta^{-1}(1) - F_\theta^{-1}(1 - p_h) = 1.5,$$

[6]The value $p_h = 0.84$ is chosen for the ease of calculation.

or $n \geqslant 3.5$. Therefore, for $n \geqslant 4$, we know that the n-tuple $(F_\theta, \dots, F_\theta)$ is $(1 - p_h, 1)$-jointly mixable. Now, by Proposition 4.3, we have, for $n \geqslant 4$,

$$\min_{Y_i \in \mathcal{B}_{F_\theta}, i=1,\dots,n} \mathcal{U}_h \left(np(\theta) + \sum_{i=1}^{n} Y_i \right) = n(\mathcal{U}_{h_*}(-X_\theta) + p(\theta)).$$

Thus, insurer A will make decision over $\theta \in [2.5, \infty)$ in the same way as she would with the utility $\mathcal{U}_{h_*}(-X_\theta) + p(\theta)$, which is independent of n as long as $n \geqslant 4$. In other words, due to the fear for dependence uncertainty, the insurer A's decision is dictated by $\mathcal{U}_{h_*}$ instead of her original utility $\mathcal{U}_h$. □

Before presenting the next example, we have the following result which is interesting in its own right.

Theorem 5.1. *For any continuous $h \in \mathcal{H}$ and $X \in \mathcal{X}$, we have*

$$\min_{Y \in \mathcal{X}} \{\mathcal{U}_h(X+Y) - \mathcal{U}_{h_*}(Y)\} = \mathcal{U}_{h_*}(X), \tag{5.1}$$

and

$$\max_{Y \in \mathcal{X}} \{\mathcal{U}_h(X+Y) - \mathcal{U}_{h^*}(Y)\} = \mathcal{U}_{h^*}(X). \tag{5.2}$$

Proof. We only show (5.1) since (5.2) is symmetric to (5.1). First, from Theorem 4.2 (i),

$$\mathcal{U}_h(X+Y) \geqslant \mathcal{U}_{h_*}(X) + \mathcal{U}_{h_*}(Y)$$

for all $Y \in \mathcal{X}$. This shows

$$\min_{Y \in \mathcal{X}} \{\mathcal{U}_h(X+Y) - \mathcal{U}_{h_*}(Y)\} \geqslant \mathcal{U}_{h_*}(X). \tag{5.3}$$

Let F be the distribution of X. From Proposition 3.1, there exist $G \in \mathcal{D}$ such that (F, G) is $\mathcal{I}^h$-jointly mixable. By Theorem 4.2 (ii), there exist $Y_1 \in B_F$ and $Y_2 \in B_G$ such that $\mathcal{U}_h(Y_1 + Y_2) = \mathcal{U}_{h_*}(Y_1) + \mathcal{U}_{h_*}(Y_2)$. Take $Y \in \mathcal{X}$ such that (X, Y) is identically distributed as (Y_1, Y_2). Then we have $\mathcal{U}_h(X+Y) - \mathcal{U}_{h_*}(Y) = \mathcal{U}_{h_*}(X)$. This shows

$$\min_{Y \in \mathcal{X}} \{\mathcal{U}_h(X+Y) - \mathcal{U}_{h_*}(Y)\} \leqslant \mathcal{U}_{h_*}(X). \tag{5.4}$$

Combining (5.3)-(5.4) we obtain (5.1). □

Remark 5.1. For any $X, Y \in \mathcal{X}$, it follows from Theorem 4.2 that

$$\mathcal{U}_{h_*}(X) + \mathcal{U}_{h_*}(Y) \leqslant \mathcal{U}_h(X+Y) \leqslant \mathcal{U}_{h^*}(X) + \mathcal{U}_{h^*}(Y).$$

One may wonder for a given $X \in \mathcal{X}$, whether there exist $Y \in \mathcal{X}$ such that one of the above two inequalities holds as an equality. Theorem 5.1 confirms that the answer is positive. Indeed, Theorem 5.1 can be formulated equivalently as follows. For any continuous $h \in \mathcal{H}$ and $X \in \mathcal{X}$, there exist $Y \in \mathcal{X}$ and $Z \in \mathcal{X}$ such that $\mathcal{U}_h(X+Y) = \mathcal{U}_{h^*}(X) + \mathcal{U}_{h^*}(Y)$ and $\mathcal{U}_h(X+Z) = \mathcal{U}_{h_*}(X) + \mathcal{U}_{h_*}(Z)$.

Example 5.2. This example is based on Theorem 5.1, in which we do not need to assume an inverse-S-shaped distortion function h. Suppose that a $\mathcal{U}_h$-decision maker B has an intractable claim Y, for which the only information that she knows is $\mathcal{U}_{h_*}(Y) \geqslant k$ where k is some constant. This may be interpreted as: an imaginary $\mathcal{U}_{h_*}$-decision maker C, who is more pessimistic than B, would agree that this risk Y has a utility at least k. Hence, the utility of Y is also at least k for B, as $\mathcal{U}_h \geqslant \mathcal{U}_{h_*}$.

Now, the decision maker B is concerned about the worst-case utility of her investment X over an admissible set $\mathcal{A}$. By Theorem 5.1, the resulting utility of $X \in \mathcal{A}$ for her is

$$\min_{Y \in \mathcal{X},\, \mathcal{U}_{h_*}(Y) \geqslant k} \{\mathcal{U}_h(X+Y)\} = \mathcal{U}_{h_*}(X) + k.$$

That is, due to the fear for uncertainty in both the distribution of Y and the dependence between X and Y, the investor B becomes pessimistic so that she would make decision on $X \in \mathcal{A}$ according to the artificial concave utility functional $\mathcal{U}_{h_*}$. Note that the imaginary investor, C, would have the same worst-case utility because $(h_*)_* = h_*$.

A technical by-product is that, although the original optimization problem $\max\limits_{X \in \mathcal{A}} \mathcal{U}_h(X)$ is not a convex problem, the optimization problem under uncertainty, $\max\limits_{X \in \mathcal{A}} \mathcal{U}_{h_*}(X)$, becomes convex. This creates great convenience for further research on optimal investment. $\square$

The above two examples suggest that dual utilities are capable of absorbing people's attitude towards uncertainty into the functional forms and, in particular, into the corresponding distortion functions.

We can obtain the same conclusion by symmetry for an optimistic investor. For instance, in the setup of Example 5.2, if a $\mathcal{U}_h$-decision maker D is concerned about the best-case utility of $X+Y$ for an investment $X \in \mathcal{A}$,

$$\mathcal{U}_h(X+Y) \text{ subject to } Y \in \mathcal{X},\ \mathcal{U}_{h^*}(Y) \leqslant k,$$

then she will end up with

$$\max\left\{\mathcal{U}_h(X+Y) : Y \in \mathcal{X},\ \mathcal{U}_{h^*}(Y) \leqslant k\right\} = \mathcal{U}_{h^*}(X) + k.$$

That is, she is indeed as optimistic as a decision maker whose preference is modeled by the convex utility functional $\mathcal{U}_{h^*}$, which only puts high weights on extreme gains.

6 Concluding remarks

In this paper we introduce the notion of conditional joint mixability, and establish (under the assumption of conditional joint mixability) the exact best- and worst-case values of dual utility functions for an aggregate risk in which the dependence is uncertain. For strict inverse-S-shaped distortion functions, we show that the above bounds are attainable if and only if the assumption of conditional joint mixability holds. An economic implication of our results is what we call the pessimism effect: in the presence of dependence uncertainty, a decision maker equipped with a dual utility tends to be more pessimistic and she makes decision according to an auxiliary concave dual utility.

We recognize that the exact best and worst values in Theorem 4.2 require a rather strong assumption of conditional joint mixability, which is not easy to check for general distributions. Nevertheless, by imposing this condition we observe interesting consequences such as those presented in Section 5, leading to a new interpretation of dependence uncertainty in decision making. Moreover, the bounds in Theorem 4.2 may be used as approximations if the assumption of conditional joint mixability is dropped (see Remark 4.4).

Moving forward, various decision problems such as optimal consumption or portfolio selection with dual utilities and dependence uncertainty may be studied. Another possible direction of further research is to evaluate risk aggregation under uncertainty for RDUT and CPT functionals. Although mathematical formulas for RDUT and CPT may not be as elegant as in Theorem 4.2 of this paper, they may nonetheless serve well for other broad classes of decision problems and applications.

References

Acerbi, C. (2002). Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**(7), 1505–1518.

Bernard, C., Jiang, X. and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.

Brighi, B. and Chipot, M. (1994). Approximated convex envelope of a function. *SIAM Journal on Numerical Analysis*, **31**, 128–148.

Cai, J., Liu, H. and Wang, R. (2018). Asymptotic equivalence of risk measures under dependence uncertainty. *Mathematical Finance*, **28**(1), 29–49.

Chateauneuf, A., Cohen, M. and Meilijson, I. (2005). More pessimism than greediness: a characterization of monotone risk aversion in the rank-dependent expected utility model. *Economic Theory*, **25**, 649–667.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic response to Basel 3.5. *Risks*, **2**(1), 25-48.

Embrechts, P., Wang, B. and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, **19**(4), 763–790.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**, 141–153.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review.* **91**(2), 60–66.

Hou, D. and Xu, Z. Q. (2016). A robust Markowitz mean-variance portfolio selection model with an intractable claim. *SIAM Journal on Financial Mathematics*, **7**(1), 124–151.

He, X. and Zhou, X. (2016). Hope, fear, and aspirations. *Mathematical Finance*, **26**(1), 3–50.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica*, **46**, 171–185.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. *Advances in Experimental Social Psychology*, **20**, 255–295.

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

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Statistical Models and Risks*. Wiley, England.

Prelec, D. (1998). The probability weighting function. *Econometrica*, **66**(3), 497–527.

Puccetti, G. and Wang R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.

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

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

Schmeidler, D. (1989). Subject probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**(4), 297–323.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, **102**(10), 1344–1360.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, **41**(3), 808–826.

Wang, R., Bignozzi, V. and Tsakanas, A. (2015). How superadditive can a risk measure be? *SIAM Jounral on Financial Mathematics*, **6**, 776–803.

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst value-at-risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

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