Risk Reducers in Convex Order

Junnan He[a], Qihe Tang[b] and Huan Zhang[b] *

[a] Department of Economics
Washington University in St. Louis
Campus Box 1208, St. Louis, MO 63130-4899

[b] Department of Statistics and Actuarial Science
University of Iowa
241 Schaeffer Hall, Iowa City, IA 52242-1409

July 6, 2016

Abstract

Given a risk position X, a random addition Z is called a risk reducer for X if the new position $X + Z$ is less risky than $X + E[Z]$ in convex order. We utilize the concept of convex hull to give a structural description of risk reducers in the case of an atomless probability space. Then we study risk reducers that are fully dependent on X. Applications to multivariate stochastic ordering, index-linked hedging strategies, and optimal reinsurance are proposed.

Keywords: convex hull; co/counter-monotonicity; multivariate stochastic ordering; index-linked hedging strategies; optimal reinsurance

JEL: C44, G22, C73

1 Introduction

In this paper we study when a random addition can reduce the risk of a given risk position, where we mean to reduce the risk in convex order. Interests in stochastic ordering of risks have long been existing. Terminologies closely related to convex order such as second-order stochastic dominance, the Rothschild–Stiglitz increase in risk, majorization, mean preserving spread, and stop-loss order are popular in mathematics, statistics, economics, and decision theory. Pioneering works on these and related concepts include Lehmann (1955), Rothschild and Stiglitz (1970), and Day (1972). See the monographs Ross (1983), Stoyan (1983), Arnold (1987), Mas-Colell et al. (1995), Müller and Stoyan (2002), and Marshall et al. (2011) for more extensive discussions. See also Denuit et al. (2005) and Dhaene et al. (2005), among others, for discussions on convex order in the insurance context.

*Corresponding author: Huan Zhang; Email: huan-zhang-1@uiowa.edu; Cell: 319-594-2188

Suppose that all random variables under consideration are defined on an appropriate probability space $(\Omega, \mathcal{F}, P)$ and are integrable. For two random variables X and Y, which we interpret as risks or loss-profit variables throughout the paper, say Y is less risky than X in convex order, denoted by $Y \leq_{\mathrm{cx}} X$, if

$$E[v(Y)] \leq E[v(X)] \tag{1.1}$$

for every convex function v for which the two expectations exist.

This immediately reminds one of the Rothschild–Stiglitz increase in risk, with which the reader from economic theory or decision theory probably feels more familiar. Actually, $Y \leq_{\mathrm{cx}} X$ if and only if $-Y \leq_{\mathrm{cx}} -X$ if and only if $E[v_1(-Y)] \leq E[v_1(-X)]$ for every convex function v_1 for which the two expectations exist. Thus, $Y \leq_{\mathrm{cx}} X$ if and only if the inequality

$$E\left[u(\tilde{Y})\right] \geq E\left[u(\tilde{X})\right] \tag{1.2}$$

holds for every concave function $u = -v_1$ for which the two expectations exist, where $\tilde{X} = -X$ and $\tilde{Y} = -Y$ are correspondingly interpreted as gains or profit-loss variables. Relation (1.2) means that a risk averter, i.e., a person with a concave utility function, prefers $\tilde{Y}$ to $\tilde{X}$. Another closely related concept is the second-order stochastic dominance, which is defined by relation (1.2) for every non-decreasing concave utility function u for which the two expectations exist. For later use, let us also recall a well-known result that $Y \leq_{\mathrm{cx}} X$ if and only if $E[Y] = E[X]$ and

$$E\left[(Y-d)_+\right] \leq E\left[(X-d)_+\right] \qquad \text{for all } d \in \mathbb{R},$$

where $x_+ = \max\{x, 0\}$ denotes the positive part of a real number x. The inequality above defines that Y is less risky than X in stop-loss order; see, for example, Kaas and van Heerwaarden (1992).

Recently, Cheung et al. (2014) introduced the concept of risk reducer in convex order as follows:

Definition 1.1 *For a given integrable random variable X, a random variable Z is said to be its risk reducer, denoted by $Z \in R(X)$, if*

$$X + Z \leq_{\mathrm{cx}} X + E[Z]. \tag{1.3}$$

In this paper we aim at a structural description of the set $R(X)$. In other words, we follow the work of Cheung et al. (2014) to investigate under what conditions adding a negatively dependent risk will decrease the overall level of risk. Dually, under the expected utility framework, Finkelshtain et al. (1999), Denuit et al. (2011), and Li et al. (2016), among others, have investigated under what conditions adding a positively dependent risk will increase the overall level of risk. We remark that their problem reveals a similar flavor, but our problem is essentially a different one.

In Theorem 1 of Cheung et al. (2014), a sufficient condition for Z to be a counter-monotonic risk reducer for X was put forward. Recall that two random variables X_1 and X_2 are said to be almost surely comonotonic if the inequality

$$(X_1(\omega) - X_1(\omega'))\,(X_2(\omega) - X_2(\omega')) \geq 0$$

holds for every ω and ω' belong to Ω but a null set N. Hereafter, whenever we quote almost sure comonotonicity we shall drop the words "almost sure" for brevity. Two random variables X_1 and X_2 are said to be counter-monotonic if X_1 and $-X_2$ are comonotonic. See Schmeidler (1986, 1989), Landsberger and Meilijson (1994), and Dhaene et al. (2002a, 2002b, 2006) for the theory of comonotonicity and its applications to various areas of economics.

A counter-monotonic risk reducer has immediate relevance to insurance. Consider an individual who faces a potential monetary loss X. He purchases an insurance contract with indemnity payoff $I(X)$, where I is a non-decreasing function over the range of X and satisfies $0 \leq I(x) \leq x$, and he pays pure premium $E[I(X)]$. With $Z = -I(X)$, which is counter-monotonic with X, relation (1.3) implies that the insured's residual loss, $X - I(X) + E[I(X)]$, becomes less risky than the original loss X. See Section 5 of Cheung et al. (2014) for a related discussion on universal marketability.

Apparently, in order for (1.3) to hold, it is not necessary to require Z to be counter-monotonic with X. For example, restricted to the family of normal distributions, it is easy to see that (1.3) holds if X and Z jointly follow a bivariate normal distribution with correlation coefficient ρ satisfying

$$-1 \leq \rho \leq -\frac{1}{2}\sqrt{\frac{\mathrm{Var}[Z]}{\mathrm{Var}[X]}},$$

yielding a much broader range than with $\rho = -1$.

In this paper we aim at a structural description of the set $R(X)$ of all risk reducers instead of counter-monotonic ones. The study of general risk reducers is meaningful particularly in view of the increasing complexity of products in nowadays insurance and financial markets.

Consider, for example, a bundle of home and auto insurance, which is a prevalent practice in insurance. Denote by X_1 and X_2 the potential monetary losses on the home and the auto, respectively. To decide the feasibility of the corresponding indemnity payoffs $I_1(X_1)$ and $I_2(X_2)$, we hope to establish inequality (1.3) with $X = X_1 + X_2$ and $Z = -(I_1(X_1) + I_2(X_2))$. Then Z is not counter-monotonic with X anymore except for some artificial cases. In such a situation with multiple risks, the study of counter-monotonic risk reducers is of limited use.

Let us take variable annuities with guaranteed minimum accumulation benefits (GMAB) as another example. GMAB is a living benefit that guarantees a minimum contract value at maturity to protect the policyholder against decreases in value of the reference portfolio. For the insurer, the guarantee loss involves both financial risk and mortality risk. If, as usual, one uses a put option contract purely on the reference portfolio to hedge, it cannot be counter-monotonic with the guarantee loss because of the involved mortality risk. See Coleman et al. (2006, 2007) and Bauer et al. (2008) for discussions on hedging guarantee losses in various variable annuities.

We conclude this introduction with the following lemma, which collects some elementary properties of the set $R(X)$ of risk reducers:

Lemma 1.1 *For a given integrable random variable X, we have*

(a) $R(X+a) = R(X)$ for $a \in \mathbb{R}$;

(b) $R(bX) = bR(X) := \{bY : Y \in R(X)\}$ for $b \neq 0$;

(c) $R(X) \subset R(X+Y)$ for an integrable random variable Y independent of X;

(d) $R(X) \subset R(X+Y)$ for an integrable random variable Y comonotonic with X.

Items (a)–(c) follow immediately from the definition of $R(X)$, while item (d) can easily be verified by Corollary 1 of Dhaene et al. (2002a).

The rest of the paper is organized as follows: In Section 2, we utilize the concept of convex hull to give a structural description of risk reducers in the case of an atomless probability space and we propose an application of this result to bivariate convex ordering. In Section 3, we study risk reducers that are fully dependent on X and we propose applications to index-linked hedging strategies as well as optimal stop-loss reinsurance. Finally, all long proofs of those theorems, corollaries, and lemmas presented in the main body of the paper are collected in Section 4 as an Appendix.

2 Description of risk reducers using convex hull

2.1 Convex hull

The main purpose of this section is to give a structural description of $R(X)$ for a given integrable random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$. For this purpose, we need the concept of convex hull and some notation. The sets introduced below contain random variables all defined on the same probability space $(\Omega, \mathcal{F}, P)$ as X.

Denote by

$$C(X) = \{Y : Y \leq_{\text{cx}} X\}$$

the set of all positions that are less risky than X in convex order. Rewriting (1.3) as $X + Z - E[Z] \leq_{\text{cx}} X$, we see that

$$R(X) = C(X) - X + \mathbb{R}, \tag{2.1}$$

and thus we need to focus on the description of $C(X)$ only. By the way, in (2.1) we have used the following notation: for two sets A, B and an element x, we write $A+B=\{a+b: a\in A, b\in B\}$ and $A-x=\{a-x: a\in A\}$. We shall use such notation throughout the paper without additional explanation.

The classical Strassen's (1965) theorem states that, as restated in Theorem 1.5.20 of Müller and Stoyan (2002), for random variables X and Y defined on the probability space $(\Omega, \mathcal{F}, P)$, the inequality $Y\leq_{\text{cx}} X$ holds if and only if there are another probability space $(\Omega', \mathcal{F}', P')$ and random variables X' and Y' defined on $(\Omega', \mathcal{F}', P')$ identically distributed as X and Y, respectively, such that $Y'=E[X'|Y']$. Thus, applying Strassen's theorem we may obtain a martingale characterization of the set $C(X)$. However, the description of $C(X)$ that we pursue in this paper will be essentially different from and more structural than such a martingale characterization.

The convex hull of a set A, denoted by $\text{Conv}(A)$, is the set of all finite convex combinations of elements in A; that is

$$\text{Conv}(A)=\left\{\sum_{i=1}^{k} a_i X_i : k\in\mathbb{N},\ a_i\geq 0,\ \sum_{i=1}^{k} a_i = 1 \text{ and } X_i\in A\right\}.$$

Denote by $\overline{\text{Conv}}(A)$ the closure of $\text{Conv}(A)$ in L^1 space.

Moreover, denote by $D(X)$ the set of all random variables that are identically distributed as X; that is,

$$D(X)=\{X' : X'=_{\text{d}} X\}.$$

2.2 A general discussion

In this subsection, we aim at a result about $R(X)$ in which no restriction on the underlying probability space is imposed. For this purpose, we need the following:

Lemma 2.1 *Let X be an integrable random variable defined on a general probability space $(\Omega, \mathcal{F}, P)$. Then the set $C(X)$ is convex and closed in L^1 norm.*

Proof. To establish the convexity of $C(X)$, choose $Y_1, Y_2\in C(X)$ and $a\in[0,1]$. Since for all $d\in\mathbb{R}$,

$$E[((aY_1+(1-a)Y_2)-d)_+]\leq aE[(Y_1-d)_+]+(1-a)E[(Y_2-d)_+]\leq E[(X-d)_+],$$

it follows that $aY_1+(1-a)Y_2\leq_{\text{cx}} X$. Hence, $aY_1+(1-a)Y_2\in C(X)$.

To prove that $C(X)$ is closed, choose a sequence of random variables $\{Y_n, n\in\mathbb{N}\}$ from $C(X)$ such that $E[|Y_n-Y|]\to 0$ as $n\to\infty$ for a random variable Y. Then it holds for every $d \in \mathbb{R}$ that

$$\begin{aligned} & \left|E\left[(Y_n-d)_+\right]-E\left[(Y-d)_+\right]\right| \\ \leq\ & E\left[\left|(Y_n-d)_+-(Y-d)_+\right|\right] \\ =\ & E\left[(Y-d)\,1_{\{Y_n\leq d<Y\}}+(Y_n-d)\,1_{\{Y\leq d<Y_n\}}+|Y_n-Y|\,1_{\{Y_n>d\}\cap\{Y>d\}}\right] \\ \leq\ & E\left[|Y_n-Y|\right]\to 0, \qquad n\to\infty. \end{aligned}$$

Hence, $E\left[(Y-d)_+\right] \leq E\left[(X-d)_+\right]$ and $Y \in C(X)$. ■

Lemma 2.1 leads to the desired result:

Corollary 2.1 *Let X be an integrable random variable defined on a general probability space $(\Omega, \mathcal{F}, P)$. Then $\overline{\mathrm{Conv}}(D(X) \cup \{E[X]\}) \subset C(X)$.*

It is noteworthy that $E[X]$ is in general not a trivial addition in Corollary 2.1, though $E[X] \in \overline{\mathrm{Conv}}(D(X))$ is often true. We construct a simple example to illustrate this. Consider a probability space consisting of only two atoms and define random variables X and Y as follows:

Ω	ω_1	ω_2
P	$1/3$	$2/3$
X	-2	1
Y	1	$-1/2$

Then $D(X)$ is only a singleton $\{X\}$, and so is $\overline{\mathrm{Conv}}(D(X))$. Thus, $E[X] = 0 \notin \overline{\mathrm{Conv}}(D(X))$. This example also serves as an illustration for the possibility that $Y \leq_{\mathrm{cx}} X$ but $Y \notin \overline{\mathrm{Conv}}(D(X) \cup \{E[X]\})$, meaning that the reverse inclusion in Corollary 2.1 does not hold in general.

In the next subsection, we restrict our consideration to an atomless probability space and we show that the reverse inclusion in Corollary 2.1 holds true then.

2.3 The case of an atomless probability space

For two random variables X and Y uniformly distributed on two real sets $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$, respectively, according to 1.A.3 of Marshall et al. (2011), $Y \leq_{\mathrm{cx}} X$ if and only if the vector $\mathbf{y} = (y_1, \ldots, y_n)$ is in the convex hull of $n!$ permutations of the vector $\mathbf{x} = (x_1, \ldots, x_n)$. A similar result about random variables defined on the interval $[0, 1]$ was proved by Ryff (1965). We shall extend Theorem 5 of Ryff (1965) to any atomless standard Borel space.

Let us first recall several concepts; see pages 13 and 74 of Kechris (1995) for more details. A topological space is called Polish if it is separable and completely metrizable. A measurable space $(\Omega, \mathcal{B}_\Omega)$ is called a standard Borel space if Ω is endowed with a Polish topology and $\mathcal{B}_\Omega$ is the generated Borel sigma field. A probability measure P is called atomless if for any set $A \in \mathcal{B}_\Omega$ with $P(A) > 0$ there is a measurable subset B of A such that $0 < P(B) < P(A)$. These regularity assumptions describe a nice probability space that is enough for most applications. For example, $\left([0,1]^d, \mathcal{B}_{[0,1]^d}\right)$ equipped with the Lebesgue measure is a d-dimensional atomless standard Borel space.

We postpone the proof of the following result to the Appendix:

Lemma 2.2 *Let $(\Omega, \mathcal{B}_\Omega, P)$ be an atomless and standard Borel space and let X be an integrable random variable defined on it. Then $C(X) = \overline{\mathrm{Conv}}(D(X))$.*

The reason that $E[X]$ does not need to appear in Lemma 2.2 is as follows. The assumptions on the probability space imply the existence of a sequence of independent random variables identically distributed as (though not necessarily independent of) X such that the sample means converge to $E[X]$ almost surely and, hence, in L_1 since the sample means are uniformly integrable. Lemma 2.2 justifies a common view in portfolio theory that diversification reduces risk in convex order.

Lemma 2.2 and relation (2.1) lead to our first main result:

Theorem 2.1 *Let $(\Omega, \mathcal{B}_\Omega, P)$ be an atomless and standard Borel space and let X be an integrable random variable defined on it. Then*

$$R(X) = \overline{\mathrm{Conv}}(D(X)) - X + \mathbb{R}.$$

2.4 Application to multivariate stochastic ordering

In this subsection, we propose an application of Lemma 2.2 to bivariate convex ordering, and we remark that the extension to the multivariate case is straightforward.

A bivariate function $v : \mathbb{R}^2 \to \mathbb{R}$ is said to be 2-increasing (also called supermodular or superadditive in the literature of stochastic ordering) if it holds for every rectangle $A = (a_1, a_2] \times (b_1, b_2]$ that

$$v(a_2, b_2) - v(a_1, b_2) - v(a_2, b_1) + v(a_1, b_1) \geq 0.$$

If further v is continuous from above, it defines a sigma-finite measure over $\mathbb{R}^2$, which we still denote by v for simplicity.

Lemma 2.3 *Let $X^{(1)}$, $X^{(2)}$, $Y^{(1)}$, and $Y^{(2)}$ be integrable random variables defined on the same probability space such that $X^{(1)} =_{\mathrm{d}} Y^{(1)}$, $X^{(2)} =_{\mathrm{d}} Y^{(2)}$, and $X^{(1)}$ and $X^{(2)}$ are comonotonic, and let $v : \mathbb{R}^2 \to \mathbb{R}$ be 2-increasing and continuous from above. Then we have*

$$E\left[v(Y^{(1)}, Y^{(2)})\right] \leq E\left[v(X^{(1)}, X^{(2)})\right]$$

as long as the two expectations exist.

Our Lemma 2.3 partially generalizes Theorem 2.8 of Whitt (1976) and Corollary 2.2 of Tchen (1980) regarding a classical result of Lorentz (1953). By this lemma we can establish the following result, which serves as an application of Lemma 2.2:

Corollary 2.2 *Let $(\Omega, \mathcal{B}_\Omega, P)$ be an atomless and standard Borel space, and let $X^{(1)}$, $X^{(2)}$, $Y^{(1)}$, and $Y^{(2)}$ be defined on this space such that $Y^{(1)} \leq_{\mathrm{cx}} X^{(1)}$, $Y^{(2)} \leq_{\mathrm{cx}} X^{(2)}$, and $X^{(1)}$ and $X^{(2)}$ are comonotonic. Furthermore, let $v : \mathbb{R}^2 \to \mathbb{R}$ be 2-increasing and componentwise convex (hence, continuous). Then under either of the following conditions the inequality*

$$E\left[v(Y^{(1)}, Y^{(2)})\right] \leq E\left[v(X^{(1)}, X^{(2)})\right]$$

holds provided that the two expectations exist:

(a) the first-order partial derivatives of v exist and are bounded over the range of $(X^{(1)}, X^{(2)})$;

(b) the random variables $X^{(1)}$ and $X^{(2)}$ are bounded.

It is well known that if $Y^{(1)} \leq_{\mathrm{cx}} X^{(1)}$ and $Y^{(2)} \leq_{\mathrm{cx}} X^{(2)}$ then $Y^{(1)} + Y^{(2)} \leq_{\mathrm{cx}} X^{(1)} + X^{(2)}$; see, for example, Corollary 1 of Dhaene et al. (2002a). Now we show that Corollary 2.2(a) implies this result. Actually, for any stop-loss function $v^d(x) = (x - d)_+$, $d \in \mathbb{R}$, by a general approximation procedure one can always construct a sequence of differentiable and convex functions v_n^d non-increasing to v^d as $n \to \infty$ such that, for each n, v_n^d is identical to v^d except for $x \in (d - 1/n, d + 1/n)$. Clearly, each $v_n^d(x_1 + x_2)$ fulfills all conditions required for $v(x_1, x_2)$ in (a). Thus, $E\left[v_n^d(Y^{(1)} + Y^{(2)})\right] \leq E\left[v_n^d(X^{(1)} + X^{(2)})\right]$. By the monotone convergence theorem, letting $n \to \infty$ on both sides yields that $E\left[v^d(Y^{(1)} + Y^{(2)})\right] \leq E\left[v^d(X^{(1)} + X^{(2)})\right]$. Since d is arbitrary, this gives $Y^{(1)} + Y^{(2)} \leq_{\mathrm{cx}} X^{(1)} + X^{(2)}$.

Both proofs of Lemma 2.3 and Corollary 2.2 are postponed to the Appendix. Recent related discussions on multivariate stochastic ordering can be found in Denuit and Mesfioui (2010).

3 Fully dependent risk reducers

Deterministic transformation of risks with applications to comparative statics of risk has been studied by Meyer and Ormiston (1989), Quiggin (1991), and Levy and Wiener (1998), among others. Motivated by these works, we now restrict the discussion to risk reducers that are fully dependent on the given risk position.

3.1 Description of fully dependent risk reducers

Let $(\Omega, \mathcal{B}_\Omega, P)$ be an atomless and standard Borel space, and let the risk X defined on this space follow a continuous distribution F. We are dedicated to a structural description of the set of all risk reducers for X of the form $Z = h(X)$ for some measurable but not necessarily monotone function h; that is, we study the set

$$\tilde{R}(X) = \{Z : Z = h(X) \text{ for some measurable function } h \text{ such that } X + Z \leq_{\text{cx}} X + E[Z]\}.$$

Analogously to Subsection 2.1, introduce

$$\tilde{C}(X) = \left\{\tilde{Y} : \tilde{Y} = g(X) \text{ for some measurable function } g \text{ such that } \tilde{Y} \leq_{\text{cx}} X\right\}$$

to be the set of all positions that are less risky than X and fully dependent on X. Then, the same as before,

$$\tilde{R}(X) = \tilde{C}(X) - X + \mathbb{R}.$$

Thus, we only focus on the description of $\tilde{C}(X)$.

A key idea is to restrict the consideration to the smaller sigma field generated by X, that is, $X^{-1}(\mathcal{B}_{\mathbb{R}})$. Recall a well-known result that a random variable $\tilde{Y}$ can be expressed as a measurable function of X, i.e., $\tilde{Y} = g(X)$ for some measurable function g, if and only if $\tilde{Y}^{-1}(\mathcal{B}_{\mathbb{R}}) \subset X^{-1}(\mathcal{B}_{\mathbb{R}})$; see, for example, Theorem 20.1 of Billingsley (1995). Thus,

$$\tilde{C}(X) = \left\{\tilde{Y} \text{ defined on } (\Omega, X^{-1}(\mathcal{B}_{\mathbb{R}}), P) : \tilde{Y} \leq_{\text{cx}} X\right\}.$$

Note that $(\Omega, X^{-1}(\mathcal{B}_{\mathbb{R}}), P)$ is still an atomless and standard Borel space since X is continuous. By Lemma 2.2, we have

$$\tilde{C}(X) = \overline{\text{Conv}}(\tilde{D}(X)),$$

where

$$\tilde{D}(X) = \{\tilde{X} \text{ defined on } (\Omega, X^{-1}(\mathcal{B}_{\mathbb{R}}), P) : \tilde{X} =_{\text{d}} X\}. \tag{3.1}$$

The set $\tilde{D}(X)$ can also be understood in the following way. Introduce $U = F(X)$, which is a uniform random variable on $(\Omega, X^{-1}(\mathcal{B}_{\mathbb{R}}), P)$. Furthermore, if a measurable function $m : [0,1] \to [0,1]$ is Lebesgue measure preserving, then $m(U)$ is still a uniform random variable $(\Omega, X^{-1}(\mathcal{B}_{\mathbb{R}}), P)$. Denoting by $\mathcal{M}$ the set of all such measure-preserving functions m, it is easy to see that

$$\tilde{D}(X) = \{F^{\leftarrow}(m(U)) : m \in \mathcal{M}\}, \tag{3.2}$$

where $F^{\leftarrow}$, defined by $F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}$ for $q \in (0,1)$, is the well-known quantile function of F.

We conclude this subsection with the following result:

Theorem 3.1 *Let $(\Omega, \mathcal{B}_{\Omega}, P)$ be an atomless and standard Borel space and let the risk X follow a continuous distribution F. The set $\tilde{R}(X)$ of all fully dependent risk reducers for X can be described as*

$$\tilde{R}(X) = \overline{\text{Conv}}(\tilde{D}(X)) - X + \mathbb{R}, \tag{3.3}$$

where $\tilde{D}(X)$ is defined by (3.1) and can alternatively be understood as (3.2).

Obviously, Theorem 3.1 gives a much larger set of risk reducers than the set of countermonotonic risk reducers studied by Cheung et al. (2014).

3.2 Application to index-linked hedging strategies

As pointed out by Kellner and Gatzert (2013), index-linked hedging strategies are of high relevance to the financial and insurance industry. Such strategies link the payoff of a contract to the development of an index. Despite benefits such as high transparency, low transaction costs, and reduction of moral hazard, the use of an index leads to basis risk as the insurer's exposure is usually not fully dependent on the index; see, for example, Gatzert and Kellner (2013). Therefore, it is important to control the basis risk to an acceptable level. Cummins et al. (2004) conducted an empirical study to analyze the effectiveness of catastrophic-loss index options in hedging hurricane losses for Florida insurers using a windstorm simulation model developed by Applied Insurance Research (AIR). Motivated by their study, we propose the following application of Theorem 3.1 to index-linked hedging strategies.

Consider n insurers in a certain state having unhedged losses $L_1, \dots, L_n$ within a certain year. The sum

$$L = \sum_{j=1}^{n} L_j$$

represents a statewide industry loss index. A hedging strategy for insurer $j = 1$, say, under the loss index L defines a function $h(L)$ such that its hedged loss becomes

$$L_1^S = L_1 - h(L).$$

From the point of view of risk management, it is natural to require that the hedged loss L_1^S is less risky, or, more precisely, that $Z = -h(L)$ is a risk reducer for L_1 in convex order.

Suppose that the vector $\mathbf{L} = (L_1, \dots, L_n)$ follows a multivariate normal distribution $N_n(\mu, \mathbf{\Sigma})$ with mean vector $\mu = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ and covariance matrix $\mathbf{\Sigma} = [\sigma_{ij}] \in \mathbb{R}^{n \times n}$. To avoid triviality, assume that $\mathbf{\Sigma}$ is positive definite. Notice that

$$(L_1, L) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \end{bmatrix} \mathbf{L} = B\mathbf{L} \sim N_2\left(B\mu, B\mathbf{\Sigma}B'\right).$$

It is easy to see that there are some independent standard normal random variables ε_1 and ε_2 such that

$$\begin{cases} L = \varepsilon_1 \sqrt{\sum_{i,j=1}^{n} \sigma_{ij}} + \sum_{j=1}^{n} \mu_j, \\ L_1 = \varepsilon_1 \frac{\sum_{j=1}^{n} \sigma_{1j}}{\sqrt{\sum_{i,j=1}^{n} \sigma_{ij}}} + \varepsilon_2 \sqrt{\sigma_{11} - \frac{\left(\sum_{j=1}^{n} \sigma_{1j}\right)^2}{\sum_{i,j=1}^{n} \sigma_{ij}}} + \mu_1. \end{cases} \tag{3.4}$$

By the independence between ε_1 and ε_2 in the second equality in (3.4) and Lemma 1.1(a, b, c),

$$\frac{\sum_{j=1}^{n} \sigma_{1j}}{\sqrt{\sum_{i,j=1}^{n} \sigma_{ij}}} \tilde{R}\left(\varepsilon_1\right) \subset R(L_1).$$

Then by the first equality in (3.4) and Lemma 1.1(a, b),

$$\frac{\sum_{j=1}^{n}\sigma_{1j}}{\sqrt{\sum_{i,j=1}^{n}\sigma_{ij}}}\tilde{R}\left(\varepsilon_1\right)=\frac{\sum_{j=1}^{n}\sigma_{1j}}{\sum_{i,j=1}^{n}\sigma_{ij}}\tilde{R}\left(L\right),$$

meaning that any element Z of the set on the left-hand side is fully dependent on L. By Theorem 3.1, such risk reducers form the set

$$\frac{\sum_{j=1}^{n}\sigma_{1j}}{\sum_{i,j=1}^{n}\sigma_{ij}}\tilde{R}\left(L\right)=\frac{\sum_{j=1}^{n}\sigma_{1j}}{\sum_{i,j=1}^{n}\sigma_{ij}}\left(\overline{\mathrm{Conv}}(\tilde{D}\left(L\right))-L\right)+\mathbb{R}.$$

3.3 When the original and reduced positions are comonotonic

Next we restrict the discussion to the case in which $(X, X+Z)$ is comonotonic. The comonotonicity here is not just for technical convenience but of practical relevance. In the insurance context, if X is interpreted as a loss to the insured and $-Z$ as the amount paid by the insurer, then $X+Z$ is the amount retained to the insured. In this situation, the comonotonicity of $(X, X+Z)$ is naturally interpreted as preclusion of over insurance.

Recall that a random variable Z is negatively expectation dependent on X if

$$E[Z|X\leq d]\geq E[Z]\qquad\text{for every } d\in\mathbb{R},$$

or, equivalently,

$$E[Z|X>d]\leq E[Z]\qquad\text{for every } d\in\mathbb{R}.$$

See Wright (1987) for the original definition based on the first inequality above, see Denuit et al. (2015) for the equivalence of both inequalities, and see Li et al. (2016) for a generalization of so-called high-order expectation dependence.

The proof of the following theorem is postponed to the Appendix:

Theorem 3.2 *Let X and Z be two integrable random variables defined on a general probability space $(\Omega, \mathcal{F}, P)$ such that $Z =_{\text{a.s.}} h(X)$ for some measurable (but not necessarily monotone) function h and $(X, X+Z)$ is comonotonic. Then the following are equivalent:*

(a) $Z\in R(X)$;

(b) Z is negatively expectation dependent on X.

Cheung et al. (2014) studied the case of $Z=-h(X)$, where h is non-decreasing and 1-Lipschitz over the range of X. Note that their condition implies the comonotonicity of $(X, X+Z)$, as discussed there, but not vice versa. Moreover, a fully dependent risk reducer need not be a counter-monotonic one. We use two examples to demonstrate that our Theorem 3.2 considerably generalizes the corresponding results of Cheung et al. (2014).

Our first example is motivated by a discussion of Mas-Colell et al. (1995, page 199). Let X be uniformly distributed on $[0,1]$ and let

$$Z = \alpha \sin(2\pi X),$$

which is obviously not counter-monotonic with X. It is easy to check that, for $\alpha > 0$ small enough, $(X, X+Z)$ is comonotonic and Z is negatively expectation dependent on X. Thus, $Z \in R(X)$ by Theorem 3.2.

The following second example considers a hedging strategy using call options. Let X be the price of a stock at expiration. Let Z be a bear call ladder, which is constructed by buying a low strike call, selling an intermediate strike call and selling another high strike call; that is,

$$Z = (X - K_1)_+ - (X - K_2)_+ - (X - K_3)_+,$$

where $0 < K_1 < K_2 < K_3$. Obviously, Z is not counter-monotonic with X, but one can check that if the strike prices K_1, K_2 and K_3 are chosen such that $E[Z] = 0$, then Z is always a risk reducer for X. For more information about bear call ladders, see Section 3.5 of Cohen (2005).

It can be seen that Theorem 1 of Meyer and Ormiston (1989) corresponds to one direction of Theorem 3.2 above for a special case with the random position X distributed on $[0,1]$ and the transformation $h(x) + x$ non-decreasing, continuous, and piecewise differentiable. In essence, our result makes no regularity assumptions except for "preclusion of over insurance."

The condition that $Z =_{\text{a.s.}} h(X)$ for some measurable function h is redundant if X is continuous. Actually, since $(X, X+Z)$ is comonotonic, by Corollary 6.11 of Kallenberg (2002) there is some random variable U uniformly distributed on $[0,1]$ such that $X = F_X^{\leftarrow}(U)$ and $X + Z = F_{X+Z}^{\leftarrow}(U)$ almost surely. Then by the continuity of F, it holds almost surely that

$$Z = \left(F_{X+Z}^{\leftarrow} - F_X^{\leftarrow}\right)(U) = \left(F_{X+Z}^{\leftarrow} - F_X^{\leftarrow}\right)(F(X)),$$

which expresses Z into a measurable function of X almost surely.

3.4 Application to optimal stop-loss reinsurance

Consider an insurer who buys a stop-loss reinsurance with retention r for its insurance portfolio with aggregate loss of amount L. The reinsurer premium is equal to

$$p(r) = (1+\theta) E\left[(L-r)_+\right],$$

where $\theta \geq 0$ is the safety loading coefficient. Suppose that the insurer has a budget constraint c, $0 < c \leq p(0)$, on the reinsurance. For a given retention r, the insurer's retained loss is

$$X(r) = L - \frac{c}{p(r)}(L-r)_+.$$

Our goal is to identify an optimal retention level.

To avoid moral risk, the insurer is not allowed to buy more than one share of the stop-loss reinsurance; that is, r should be chosen such that $p(r) \geq c$. Consider the equation

$$p(r) = c \qquad \text{for } 0 < c \leq p(0). \tag{3.5}$$

As r varies from 0 to the upper endpoint of L, the premium function $p(r)$ is continuous and strictly decreasing from $p(0)$ to 0. Thus, the solution to this equation, denoted by r^*, exists and is unique. Equivalently, to avoid moral risk it is required that $r \leq r^*$.

The following result, which is a corollary of Theorem 3.2, shows that the optimal retention level is r^* in the sense of convex order:

Corollary 3.1 *Let r^* solve equation (3.5). Then $X(r^*) \leq_{\mathrm{cx}} X(r)$ for any $r < r^*$.*

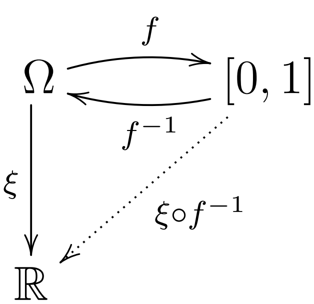
The proof of this corollary is postponed to the Appendix.

4 Appendix

4.1 Proof of Lemma 2.2

To begin with, recall that a mapping f from one topological space to another topological space is called a Borel isomorphism if it is a bijection and both f and f^{-1} are Borel measurable. See page 71 of Kechris (1995) for details of this concept.

By Theorem 17.41 of Kechris (1995), there is a Borel isomorphism $f : \Omega \to [0, 1]$ such that $P(B) = \lambda(f(B))$ for any $B \in \mathcal{B}_\Omega$, where λ is the Lebesgue measure. Hence, f is a measure-preserving mapping from Ω to $[0, 1]$ between the two spaces. Let ξ be a random variable defined on $(\Omega, \mathcal{B}_\Omega, P)$. As illustrated in the diagram below, through this mapping f we construct a random variable $\xi \circ f^{-1}$ defined on $\left([0, 1], \mathcal{B}_{[0,1]}, \lambda\right)$.



For any $B \in \mathcal{B}_\mathbb{R}$,

$$P(\xi \in B) = P\left(\xi^{-1}(B)\right) = \lambda\left(f\left(\xi^{-1}(B)\right)\right) = \lambda\left(\xi \circ f^{-1} \in B\right). \tag{4.1}$$

This implies that $\xi \circ f^{-1} =_{\mathrm{d}} \xi$, although the random variables $\xi \circ f^{-1}$ and ξ on both sides are defined on two different probability spaces. Similarly, we can prove that, for a random variable η defined on $\left([0, 1], \mathcal{B}_{[0,1]}, \lambda\right)$, we have $\eta \circ f =_{\mathrm{d}} \eta$.

Now we turn to the proof of Lemma 2.2. In view of Corollary 2.1, it suffices to show the inclusion $C(X) \subset \overline{\mathrm{Conv}}(D(X))$. Let $Y \in C(X)$. By the preliminary discussions above, the fact $Y \leq_{\mathrm{cx}} X$ implies that $Y \circ f^{-1} \leq_{\mathrm{cx}} X \circ f^{-1}$. By Theorem 5 of Ryff (1965), $Y \circ f^{-1} \in \overline{\mathrm{Conv}}\,(D(X \circ f^{-1}))$. Thus, there is a sequence of random variables $\{Y_n^*, n \in \mathbb{N}\}$ from $\mathrm{Conv}\,(D(X \circ f^{-1}))$ defined on $[0,1]$ such that

$$\lim_{n\to\infty} \int_{[0,1]} \left| Y_n^* - Y \circ f^{-1} \right| \mathrm{d}\lambda = 0.$$

For each Y_n^*, there are some positive integer k_n, nonnegative numbers $\{a_{n1}, \ldots, a_{nk_n}\}$ with $\sum_{i=1}^{k_n} a_{ni} = 1$, and random variables $\{X_{n1}^*, \ldots, X_{nk_n}^*\}$ defined on $[0,1]$ identically distributed as $X \circ f^{-1}$ such that

$$Y_n^* = \sum_{i=1}^{k_n} a_{ni} X_{ni}^*.$$

Then $Y_n^* \circ f = \sum_{i=1}^{k_n} a_{ni} \left(X_{ni}^* \circ f\right)$. Note that

$$X_{ni}^* \circ f =_{\mathrm{d}} X_{ni}^* =_{\mathrm{d}} X \circ f^{-1} =_{\mathrm{d}} X,$$

where the first and last equalities in distribution are due to the preliminary discussion at the beginning of this proof. Thus, $Y_n^* \circ f \in \mathrm{Conv}\,(D(X))$. Moreover, due to the same reasoning,

$$Y_n^* \circ f - Y = \left(Y_n^* - Y \circ f^{-1}\right) \circ f =_{\mathrm{d}} Y_n^* - Y \circ f^{-1}.$$

It follows that

$$\lim_{n\to\infty} \int_{\Omega} \left| Y_n^* \circ f - Y \right| \mathrm{d}P = \lim_{n\to\infty} \int_{[0,1]} \left| Y_n^* - Y \circ f^{-1} \right| \mathrm{d}\lambda = 0.$$

Hence, $Y \in \overline{\mathrm{Conv}}(D(X))$.

4.2 Proof of Lemma 2.3

Denote the four quadrants of the space $\mathbb{R}^2$ by $\Delta_1 = (0,\infty) \times (0,\infty)$, $\Delta_2 = (-\infty, 0] \times (0,\infty)$, $\Delta_3 = (-\infty, 0] \times (-\infty, 0]$, and $\Delta_4 = (0,\infty) \times (-\infty, 0]$. Introducing an auxiliary bivariate function

$$\tilde{v}(z_1, z_2) = v(z_1, z_2) - v(z_1, 0) - v(0, z_2) + v(0,0) \qquad \text{for } (z_1, z_2) \in \mathbb{R}^2,$$

it is easy to verify the following:

$$\begin{aligned}
I_1 &= \int_{\Delta_1} 1_{\{x_1 \le z_1, x_2 \le z_2\}} \mathrm{d}v(x_1, x_2) = \tilde{v}(z_1, z_2) 1_{\{(z_1, z_2) \in \Delta_1\}}, \\
I_2 &= -\int_{\Delta_2} 1_{\{x_1 > z_1, x_2 \le z_2\}} \mathrm{d}v(x_1, x_2) = \tilde{v}(z_1, z_2) 1_{\{(z_1, z_2) \in \Delta_2\}}, \\
I_3 &= \int_{\Delta_3} 1_{\{x_1 > z_1, x_2 > z_2\}} \mathrm{d}v(x_1, x_2) = \tilde{v}(z_1, z_2) 1_{\{(z_1, z_2) \in \Delta_3\}}, \\
I_4 &= -\int_{\Delta_4} 1_{\{x_1 \le z_1, x_2 > z_2\}} \mathrm{d}v(x_1, x_2) = \tilde{v}(z_1, z_2) 1_{\{(z_1, z_2) \in \Delta_4\}}.
\end{aligned}$$

Summing these up yields that

$$\tilde{v}(z_1, z_2) = \sum_{j=1}^{4} I_j, \qquad (z_1, z_2) \in \mathbb{R}^2. \tag{4.2}$$

Applying this decomposition in (4.2) to $\tilde{v}(Y^{(1)}, Y^{(2)})$ and taking expectation, we have

$$\begin{aligned}
& E\left[v(Y^{(1)}, Y^{(2)})\right] - E\left[v(Y^{(1)}, 0)\right] - E\left[v(0, Y^{(2)})\right] + v(0,0) \\
= & \int_{\Delta_1} P\left(x_1 \le Y^{(1)}, x_2 \le Y^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
& - \int_{\Delta_2} P\left(x_1 > Y^{(1)}, x_2 \le Y^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
& + \int_{\Delta_3} P\left(x_1 > Y^{(1)}, x_2 > Y^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
& - \int_{\Delta_4} P\left(x_1 \le Y^{(1)}, x_2 > Y^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
\le & \int_{\Delta_1} P\left(x_1 \le X^{(1)}, x_2 \le X^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
& - \int_{\Delta_2} P\left(x_1 > X^{(1)}, x_2 \le X^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
& + \int_{\Delta_3} P\left(x_1 > X^{(1)}, x_2 > X^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
& - \int_{\Delta_4} P\left(x_1 \le X^{(1)}, x_2 > X^{(2)}\right) \mathrm{d}v(x_1, x_2) \\
= & E\left[v(X^{(1)}, X^{(2)})\right] - E\left[v(X^{(1)}, 0)\right] - E\left[v(0, X^{(2)})\right] + v(0,0),
\end{aligned}$$

where in the first step we applied Fubini's theorem to interchange the order of integral and expectation, in the second step we applied the fact that for two vectors with the same marginal distributions the comonotonic one gives a larger joint survival probability, and in the last step we applied the decomposition in (4.2) again to $\tilde{v}(X^{(1)}, X^{(2)})$. Since $E\left[v(Y^{(1)}, 0)\right] = E\left[v(X^{(1)}, 0)\right]$ and $E\left[v(0, Y^{(2)})\right] = E\left[v(0, X^{(2)})\right]$, it follows that $E\left[v(Y^{(1)}, Y^{(2)})\right] \le E\left[v(X^{(1)}, X^{(2)})\right]$, as desired.

4.3 Proof of Corollary 2.2

According to Lemma 2.2, $Y^{(1)} \le_{\mathrm{cx}} X^{(1)}$ implies $Y^{(1)} \in \overline{\mathrm{Conv}}(D(X^{(1)}))$. Thus, there is a sequence of random variables $\{Y_n^{(1)}, n \in \mathbb{N}\}$ from $\mathrm{Conv}(D(X^{(1)}))$ such that

$$\lim_{n\to\infty} \int_{\Omega} \left|Y_n^{(1)} - Y^{(1)}\right| \mathrm{d}P = 0. \tag{4.3}$$

Similarly, there is another sequence of random variables $\{Y_n^{(2)}, n \in \mathbb{N}\}$ from $\mathrm{Conv}(D(X^{(2)}))$ such that

$$\lim_{n\to\infty} \int_{\Omega} \left|Y_n^{(2)} - Y^{(2)}\right| \mathrm{d}P = 0. \tag{4.4}$$

By the definition of convex hull, each $Y_n^{(1)} \in \mathrm{Conv}(D(X^{(1)}))$ can be expressed as a certain convex combination of finitely many random variables, say, $\{\xi_{ni}^{(1)}, i = 1, \dots, k_n^{(1)}\}$, each identically distributed as $X^{(1)}$, and each $Y_n^{(2)} \in \mathrm{Conv}(D(X^{(2)}))$ can also be expressed as a certain convex combination of finitely many random variables, say, $\{\xi_{nj}^{(2)}, j = 1, \dots, k_n^{(2)}\}$, each identically distributed as $X^{(2)}$. Applying the componentwise convexity of v, we arrive at an upper bound for $v\left(Y_n^{(1)}, Y_n^{(2)}\right)$ correspondingly expressed as a convex combination of $k_n^{(1)} \times k_n^{(2)}$ terms each of the form $v\left(\xi_{ni}^{(1)}, \xi_{nj}^{(2)}\right)$. By Lemma 2.3, $E\left[v\left(\xi_{ni}^{(1)}, \xi_{nj}^{(2)}\right)\right] \leq E\left[v(X^{(1)}, X^{(2)})\right]$. It follows that

$$E\left[v\left(Y_n^{(1)}, Y_n^{(2)}\right)\right] \leq E\left[v(X^{(1)}, X^{(2)})\right]. \tag{4.5}$$

We complete the rest of the proof according to the two cases:

(a) By the mean-value theorem, there are a random variable $\theta_n^{(1)}$ between $Y_n^{(1)}$ and $Y^{(1)}$ and another random variable $\theta_n^{(2)}$ between $Y_n^{(2)}$ and $Y^{(2)}$ such that

$$\begin{aligned} & v\left(Y_n^{(1)}, Y_n^{(2)}\right) - v(Y^{(1)}, Y^{(2)}) \\ = \; & v\left(Y_n^{(1)}, Y_n^{(2)}\right) - v\left(Y^{(1)}, Y_n^{(2)}\right) + v\left(Y^{(1)}, Y_n^{(2)}\right) - v(Y^{(1)}, Y^{(2)}) \\ = \; & v_1'\left(\theta_n^{(1)}, Y_n^{(2)}\right)\left(Y_n^{(1)} - Y^{(1)}\right) + v_2'\left(Y^{(1)}, \theta_n^{(2)}\right)\left(Y_n^{(2)} - Y^{(2)}\right). \end{aligned}$$

By the condition on the two partial derivatives, both $v_1'\left(\theta_n^{(1)}, Y_n^{(2)}\right)$ and $v_2'\left(Y^{(1)}, \theta_n^{(2)}\right)$ are uniformly bounded. Then by (4.3) and (4.4), it follows that

$$\lim_{n\to\infty} \int_\Omega \left| v\left(Y_n^{(1)}, Y_n^{(2)}\right) - v(Y^{(1)}, Y^{(2)}) \right| \mathrm{d}P = 0.$$

Thus, letting $n \to \infty$ on the left-hand side of (4.5) yields $E\left[v(Y^{(1)}, Y^{(2)})\right] \leq E\left[v(X^{(1)}, X^{(2)})\right]$, as desired.

(b) Since $X^{(1)}$ and $X^{(2)}$ are bounded, the random variables $\{Y_n^{(1)}, n \in \mathbb{N}\}$ and $\{Y_n^{(2)}, n \in \mathbb{N}\}$ are uniformly bounded. Then an application of the dominated convergence theorem on the left-hand side of (4.5) immediately yields the desired inequality.

4.4 Proof of Theorem 3.2

We need to prepare a lemma which plays a crucial role in proving Theorem 3.2. A similar result is Theorem 1 of Dhaene et al. (2002a). For a non-decreasing function $g : \mathbb{R} \to \mathbb{R}$, define its two generalized inverses as, for $y \in \mathbb{R}$,

$$g^{\leftarrow}(y) = \inf\left\{x \in \mathbb{R} : g(x) \geq y\right\}, \qquad g^{\rightarrow}(y) = \sup\left\{x \in \mathbb{R} : g(x) \leq y\right\},$$

where we have followed the usual conventions $\inf \emptyset = \infty$ and $\sup \emptyset = -\infty$. In particular, for a distribution function F, its inverse $F^{\leftarrow}(q)$ for $0 < q < 1$ is the quantile function, as introduced in Subsection 3.1. As usual, we use $f(x+0)$ and $f(x-0)$ to mean the right and left limits of a function f at x, respectively.

Lemma 4.1 *Let X be a real-valued random variable and let $Y = g(X)$ for some measurable function g.*

(a) If g is non-decreasing over the range of X, then $F_Y^{\leftarrow}(p) = g \circ F_X^{\leftarrow}(p)$ for almost all $p \in (0,1)$;

(b) If g is non-increasing over the range of X, then $F_Y^{\leftarrow}(1-p) = g \circ F_X^{\leftarrow}(p)$ for almost all $p \in (0,1)$.

Proof. (a) Since g is non-decreasing, it is easy to see that, for every $y \in g(\mathbb{R})$,

$$x < g^{\leftarrow}(y) \implies g(x) \le y \implies x \le g^{\rightarrow}(y). \tag{4.6}$$

The first implication in (4.6) gives that $P(g(X) \le y) \ge P(X < g^{\leftarrow}(y))$. Hence,

$$F_Y^{\leftarrow}(p) = \inf\{y : P(g(X) \le y) \ge p\} \le \inf\{y : P(X < g^{\leftarrow}(y)) \ge p\}.$$

Letting $z = g^{\leftarrow}(y)$ and noticing that this implies $y \le g(z+0)$, we have

$$F_Y^{\leftarrow}(p) \le \inf\{g(z+0) : P(X < z) \ge p\} \le g\left(F_X^{\leftarrow}(p) + 0\right), \tag{4.7}$$

where the last step is due to the fact that $P(X \le F_X^{\leftarrow}(p)) \ge p$. Symmetrically, the second implication in (4.6) gives that

$$F_Y^{\leftarrow}(p) = \inf\{y : P(g(X) \le y) \ge p\} \ge \inf\{y : P(X \le g^{\rightarrow}(y)) \ge p\}.$$

Letting $z = g^{\rightarrow}(y)$ and noticing that this implies $y \ge g(z-0)$, we have

$$F_Y^{\leftarrow}(p) \ge \inf\{g(z-0) : P(X \le z) \ge p\} \ge g\left(F_X^{\leftarrow}(p) - 0\right). \tag{4.8}$$

Combining (4.7)–(4.8) yields that

$$g\left(F_X^{\leftarrow}(p) - 0\right) \le F_Y^{\leftarrow}(p) \le g\left(F_X^{\leftarrow}(p) + 0\right).$$

Since g has at most countably many jumps, $F_Y^{\leftarrow}(p) = g \circ F_X^{\leftarrow}(p)$ holds almost everywhere.

(b) First notice a fact. For a random variable Z and $0 < p < 1$, it holds that

$$\begin{aligned} F_Z^{\leftarrow}(1-p) &= \inf\{z : P(Z \le z) \ge 1-p\} \\ &= \inf\{z : P(Z > z) \le p\} \\ &= \inf\{z : P(-Z < -z) \le p\} \\ &= -\sup\{z : P(-Z < z) \le p\} \\ &= -F_{-Z}^{\rightarrow}(p). \end{aligned}$$

Thus, $F_Z^{\leftarrow}(1-p) = -F_{-Z}^{\leftarrow}(p)$ holds almost everywhere over $0 < p < 1$. By this and (a) we have, almost everywhere of $p \in (0,1)$,

$$F_Y^{\leftarrow}(1-p) = -F_{-g(X)}^{\leftarrow}(p) = -(-g) \circ F_X^{\leftarrow}(p) = g \circ F_X^{\leftarrow}(p).$$

This completes the proof of Lemma 4.1. ■

Proof of Theorem 3.2. In this proof, we introduce an intermediate condition below:

(c) It holds for every $p \in (0,1)$ that

$$\frac{1}{1-p}\int_p^1 h \circ F_X^{\leftarrow}(t)dt \le E[Z]. \tag{4.9}$$

Then we formulate the proof of Theorem 3.2 into the following three steps:

(a)$\Longleftrightarrow$(c): Recalling Definition 1.1, by Theorem 2.5 of Bäuerle and Müller (2006), (a) holds if and only if

$$\frac{1}{1-p}\int_p^1 F_{X+Z}^{\leftarrow}(t)\mathrm{d}t \le \frac{1}{1-p}\int_p^1 F_{X+E[Z]}^{\leftarrow}(t)\mathrm{d}t, \qquad p \in (0,1). \tag{4.10}$$

Introduce $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x + h(x)$. Due to the comonotonicity of $(X, X + Z)$, the function $g(\cdot)$ is non-decreasing over the range of X. By Lemma 4.1, $F_{X+Z}^{\leftarrow}(t) = F_{g(X)}^{\leftarrow}(t) = g \circ F_X^{\leftarrow}(t)$ almost everywhere for $t \in (0,1)$. Also note the identity $F_{X+E[Z]}^{\leftarrow}(t) = F_X^{\leftarrow}(t) + E[Z]$. Thus, inequality (4.10) is rewritten as

$$\frac{1}{1-p}\int_p^1 g \circ F_X^{\leftarrow}(t)\mathrm{d}t \le \frac{1}{1-p}\int_p^1 F_X^{\leftarrow}(t)\mathrm{d}t + E[Z], \qquad p \in (0,1),$$

which is equivalent to (c).

(c)$\Longrightarrow$(b): For arbitrarily fixed $d \in \mathbb{R}$, let $p = F_X(d)$. Then $F_X^{\leftarrow}(p) \le d$ and $F_X(F_X^{\leftarrow}(p)) = F_X(d) = p$. Hence,

$$\begin{aligned} E\left[Z|X > d\right] &= E\left[h(X)|X > d\right] \\ &= \frac{1}{1 - F_X(d)}\int_d^{\infty} h(x)\mathrm{d}F_X(x) \\ &= \frac{1}{1-p}\int_{F_X^{\leftarrow}(p)}^{\infty} h(x)\mathrm{d}F_X(x) \\ &= \frac{1}{1-p}\int_p^1 h \circ F_X^{\leftarrow}(t)\mathrm{d}t, \end{aligned} \tag{4.11}$$

where the third step is due to $P\left(F_X^{\leftarrow}(p) < X \le d\right) = 0$ and the last step due to the change of variables $x = F_X^{\leftarrow}(t)$. This proves the implication.

(b)$\Longrightarrow$(c): If $p \in F_X(\mathbb{R})$, which means that $p = F_X(d)$ for some $d \in \mathbb{R}$, then the derivation of (4.11) is still valid and, hence, by (b), inequality (4.9) holds. Now let $p \in (0,1) - F_X(\mathbb{R})$. Define $d = F_X^{\leftarrow}(p)$, which is a discontinuity point of F, and define $p_1 = P(X < d)$, $p_2 = F_X(d)$. Clearly, $p_1 \le p < p_2$. In order to prove inequality (4.9) for this p, first look at p_1 and p_2. Inequality (4.9) already holds for p_2 since $p_2 \in F_X(\mathbb{R})$. Choose a sequence $d_n \in F_X(\mathbb{R})$, $n \in \mathbb{N}$, approaching $d = F_X^{\leftarrow}(p)$ from below and write $q_n = P(X \le d_n)$, $n \in \mathbb{N}$, which approaches $p_1 = P(X < d)$. By (4.11) and (b),

$$\frac{1}{1-p_1}\int_{p_1}^1 h \circ F_X^{\leftarrow}(t)\mathrm{d}t = \lim_{n\to\infty}\frac{1}{1-q_n}\int_{q_n}^1 h \circ F_X^{\leftarrow}(t)\mathrm{d}t = \lim_{n\to\infty} E\left[Z|X > d_n\right] \le E[Z].$$

Clearly, $F_X^{\leftarrow}(p) \equiv d$ for $p_1 \leq p < p_2$, implying that $\frac{1}{1-p}\int_p^1 h(F_X^{\leftarrow}(t))\mathrm{d}t$ is a monotonic function in $p \in [p_1, p_2]$. Hence, for $p \in [p_1, p_2)$,

$$\frac{1}{1-p}\int_p^1 h(F_X^{\leftarrow}(t))\mathrm{d}t \leq \max\left\{\frac{1}{1-p_1}\int_{p_1}^1 h(F_X^{\leftarrow}(t))\mathrm{d}t,\ \frac{1}{1-p_2}\int_{p_2}^1 h(F_X^{\leftarrow}(t))\mathrm{d}t\right\} \leq E[Z].$$

This completes the proof of Theorem 3.2. ∎

4.5 Proof of Corollary 3.1

To apply Theorem 3.2, introduce

$$Z(r) = \frac{c}{p(r)}(L-r)_+ - \frac{c}{p(r^*)}(L-r^*)_+, \tag{4.12}$$

so that $X(r^*) = X(r) + Z(r)$ and $E[Z(r)] = 0$. Thus, to prove $X(r^*) \leq_{\mathrm{cx}} X(r)$ we need to verify $Z(r) \in R(X(r))$. Define

$$g_r(l) = l - \frac{c}{p(r)}(l-r)_+,$$

which is continuous and strictly increasing in $l \in (0, \infty)$ when $p(r) > c$ (which is implied by $r < r^*$), and is continuous and non-decreasing in $l \in (0, \infty)$ when $r = r^*$. Noticing that $X(r) = g_r(L)$ and $X(r^*) = g_{r^*}(L)$, it follows that $X(r)$ is comonotonic with $X(r^*) = X(r) + Z(r)$. By plugging $L = g_r^{-1}(X(r))$ into the expression for $Z(r)$ in (4.12), we see that $Z(r)$ is a measurable function of $X(r)$. Finally, expand the expression for $Z(r)$ to

$$Z(r) = \begin{cases} 0, & L \leq r, \\ \frac{c}{p(r)}(L-r)_+, & r < L \leq r^*, \\ \left(\frac{c}{p(r)} - \frac{c}{p(r^*)}\right)L - \frac{c}{p(r)}r + \frac{c}{p(r^*)}r^*, & r^* < L. \end{cases}$$

Notice that $Z(r)$ as a function of L is nonnegative over the range $0 \leq L \leq r^*$ and is non-increasing over the range $L > r^*$. Since $E[Z(r)] = 0$, it holds for every $d \in \mathbb{R}$ that

$$E[Z(r)|X(r) \leq d] = E\left[Z(r)|L \leq g_r^{-1}(d)\right] \geq 0.$$

By Theorem 3.2, $Z(r) \in R(X(r))$.

Acknowledgments. The authors would like to thank Dr. Yichun Chi at the Central University of Finance and Economics of China, Dr. Ambrose Lo at the University of Iowa, Dr. Jose A. Rodrigues Neto at Australian National University, and Dr. Jonathan Weinstein at Washington University in St. Louis for their helpful discussions. This research was supported by an NSF grant (CMMI-1435864) and a Centers of Actuarial Excellence (CAE) Research Grant (2013–2016) from the Society of Actuaries (SOA). Junnan He acknowledges the support of travel funds from the Economics Department and Graduate School of Arts and Sciences at Washington University in St. Louis, and Huan Zhang acknowledges the support of the SOA Hickman Scholars Program.

References

[1] Arnold, B. C. Majorization and the Lorenz Order: a Brief Introduction. Springer-Verlag, Berlin, 1987.

[2] Bauer, D.; Kling, A.; Russ, J. A universal pricing framework for guaranteed minimum benefits in variable annuities. ASTIN Bulletin 38 (2008), 621–651.

[3] Bäuerle, N.; Müller, A. Stochastic orders and risk measures: consistency and bounds. Insurance: Mathematics and Economics 38 (2006), 132–148.

[4] Billingsley, P. Probability and Measure (third edition). John Wiley & Sons, New York, 1995.

[5] Cheung, K. C.; Dhaene, J.; Lo, A.; Tang, Q. Reducing risk by merging counter-monotonic risks. Insurance: Mathematics and Economics 54 (2014), 58–65.

[6] Cohen, G. The Bible of Options Strategies: the Definitive Guide for Practical Trading Strategies. Pearson Education, 2005.

[7] Coleman, T. F.; Kim, Y.; Li, Y.; Patron, M. Robustly hedging variable annuities with guarantees under jump and volatility risks. Journal of Risk and Insurance 74 (2007), 347–376.

[8] Coleman, T. F.; Li, Y.; Patron, M. Hedging guarantees in variable annuities under both equity and interest rate risks. Insurance: Mathematics and Economics 38 (2006), 215–228.

[9] Cummins, J. D.; Lalonde, D.; Phillips, R. D. The basis risk of catastrophic-loss index securities. Journal of Financial Economics 71 (2004), 77–111.

[10] Day, P. W. Rearrangement inequalities. Canadian Journal of Mathematics 24 (1972), 930–943.

[11] Denuit, M.; Dhaene, J.; Goovaerts, M.; Kaas, R. Actuarial Theory for Dependent Risks. John Wiley & Sons, 2005.

[12] Denuit, M.; Eeckhoudt, L.; Menegatti, M. Correlated risks, bivariate utility and optimal choices. Economic Theory 46 (2011), 39–54.

[13] Denuit, M.; Huang, R.; Tzeng, L. Almost expectation and excess dependence notions. Theory and Decision 79 (2015), 375–401.

[14] Denuit, M.; Mesfioui, M. Generalized increasing convex and directionally convex orders. Journal of Applied Probability 47 (2010), 264–276.

[15] Dhaene, J.; Denuit, M.; Goovaerts, M. J.; Kaas, R.; Vyncke, D. The concept of comonotonicity in actuarial science and finance: theory. Insurance: Mathematics and Economics 31 (2002a), 3–33.

[16] Dhaene, J.; Denuit, M.; Goovaerts, M. J.; Kaas, R.; Vyncke, D. The concept of comonotonicity in actuarial science and finance: applications. Insurance: Mathematics and Economics 31 (2002b), 133–161.

[17] Dhaene, J.; Vanduffel, S.; Goovaerts, M. J.; Kaas, R.; Tang, Q.; Vyncke, D. Risk measures and comonotonicity: a review. Stochastic Models 22 (2006), 573–606.

[18] Dhaene, J.; Vanduffel, S.; Goovaerts, M. J.; Kaas, R.; Vyncke, D. Comonotonic approximations for optimal portfolio selection problems. Journal of Risk and Insurance 72 (2005), 253–300.

[19] Finkelshtain, I.; Kella, O.; Scarsini, M. On risk aversion with two risks. Journal of Mathematical Economics 31 (1999), 239–250.

[20] Gatzert, N.; Kellner, R. The effectiveness of gap insurance with respect to basis risk in a shareholder value maximization setting. Journal of Risk and Insurance 81 (2013), 831–860.

[21] Kaas, R.; van Heerwaarden, A. E. Stop–loss order, unequal means, and more dangerous distributions. Insurance: Mathematics and Economics 11 (1992), 71–77.

[22] Kallenberg, O. Foundations of Modern Probability (second edition). Springer-Verlag, New York, 2002.

[23] Kechris, A. S. Classical Descriptive Set Theory. Springer-Verlag, New York, 1995.

[24] Kellner, R.; Gatzert, N. Estimating the basis risk of index-linked hedging strategies using multivariate extreme value theory. Journal of Banking & Finance 37 (2013), 4353–4367.

[25] Landsberger, M.; Meilijson, I. Co-monotone allocations, Bickel–Lehmann dispersion and the Arrow–Pratt measure of risk aversion. Annals of Operations Research 52 (1994), 97–106.

[26] Lehmann, E. L. Ordered families of distributions. Annals of Mathematical Statistics 26 (1955), 399–419.

[27] Levy, H.; Wiener, Z. Stochastic dominance and prospect dominance with subjective weighting functions. Journal of Risk and Uncertainty 16 (1998), 147–163.

[28] Li, J.; Liu, D.; Wang, J. Risk aversion with two risks: a theoretical extension. Journal of Mathematical Economics 63 (2016), 100–105.

[29] Lorentz, G. G. An inequality for rearrangements. American Mathematical Monthly 60 (1953), 176–179.

[30] Marshall, A. W.; Olkin, I.; Arnold, B. C. Inequalities: Theory of Majorization and Its Applications (second edition). Springer, New York, 2011.

[31] Mas-Colell, A.; Whinston, M. D.; Green, J. R. Microeconomic Theory. Oxford University Press, Oxford, 1995.

[32] Meyer, J.; Ormiston, M. B. Deterministic transformations of random variables and the comparative statics of risk. Journal of Risk and Uncertainty 2 (1989), 179–188.

[33] Müller, A.; Stoyan, D. Comparison Methods for Stochastic Models and Risks. John Wiley & Sons, Ltd., Chichester, 2002.

[34] Quiggin, J. Comparative statics for rank-dependent expected utility theory. Journal of Risk and Uncertainty 4 (1991), 339–350.

[35] Ross, S. M. Stochastic Processes. John Wiley & Sons, Inc., New York, 1983.

[36] Rothschild, M.; Stiglitz, J. E. Increasing risk. I. A definition. Journal of Economic Theory 2 (1970), 225–243.

[37] Ryff, J. V. Orbits of L1-functions under doubly stochastic transformations. Transactions of the American Mathematical Society 117 (1965), 92–100.

[38] Schmeidler, D. Integral representation without additivity. Proceedings of the American Mathematical Society 97 (1986), 255–261.

[39] Schmeidler, D. Subjective probability and expected utility without additivity. Econometrica 57 (1989), 571–587.

[40] Stoyan, D. Comparison Methods for Queues and Other Stochastic Models. Translation from the German edited by Daryl J. Daley. John Wiley & Sons, Ltd., Chichester, 1983.

[41] Strassen, V. The existence of probability measures with given marginals. Annals of Mathematical Statistics 36 (1965), 423–439.

[42] Tchen, A. H. Inequalities for Distributions with Given Marginals. Annals of Probability 8 (1980), 814–827.

[43] Whitt, W. Bivariate distributions with given marginals. Annals of Statistics 4 (1976), 1280–1289.

[44] Wright, R. Expectation dependence of random variables, with an application in portfolio theory. Theory and Decision 22 (1987), 111–124.