

On convex order and supermodular order without finite mean

Benjamin Côté* Ruodu Wang†

February 26, 2026

Abstract

Many results on the convex order in the literature were stated for random variables with finite mean. For instance, a fundamental result in dependence modeling is that the sum of a pair of random random variables is upper bounded in convex order by that of its comonotonic version and lower bounded by that of its counter-monotonic version, and all existing proofs of this result require the random variables' expectations to be finite. We show that the above result remains true even when discarding the finite-mean assumption, and obtain several other results on the comparison of infinite-mean random variables via the convex order. To our surprise, we find two deceptively similar definitions of the convex order, both of which exist widely in the literature, and they are not equivalent for random variables with infinite mean. This subtle discrepancy in definitions also applies to the supermodular order, and it gives rise to some incorrect statements, often found in the literature.

Keywords: Comonotonicity, counter-monotonicity, heavy-tailed distributions, optimal transport, concordance order. JEL: C69.

1 Introduction

The convex order is designed to compare the variability of distributions, and it is one of the most popular notions of stochastic comparison. The convex order and its variants, such as increasing convex order or concave order, are prominent tools in many fields; for instance, they can model attitudes of risk aversion in economics ([Rothschild and Stiglitz \(1970\)](#)), risk aggregation and risk sharing in actuarial science (e.g., [Dhaene et al. \(2002\)](#) and [Denuit and Dhaene \(2012\)](#)), assessment of diversification in quantitative finance (e.g., [Föllmer and Schied](#)

*Department of Statistics and Actuarial Science, University of Waterloo, Canada. b3cote@uwaterloo.ca

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

(2016) and Mao and Wang (2020)), and comparison between statistical experiments (Torgersen (1991)). Through the classic representation theorem of Strassen (1965), the convex order is fundamental to the construction of martingales (Hirsch et al. (2011)) and the theory of martingale optimal transport (Beiglböck et al. (2013)). Comprehensive textbook treatments of the convex order can be found in Müller and Stoyan (2002, Chapter 1) and Shaked and Shanthikumar (2007, Chapter 3).

In this paper we explore some hidden aspects of the celebrated notion of convex order, and its related notion of supermodular order. Results in the literature on the convex order are often stated for the case of finite-mean random variables. Some authors, including the standard textbook Müller and Stoyan (2002), exclude infinite-mean or undefined-mean random variables in their definitions. Heavy-tailed distributions are prominent in many scientific fields. In some contexts, the assumption of a finite mean is harmless, but probabilistic models of infinite mean are encountered in many settings in finance and insurance; see Chen and Wang (2025, Section 3) for a survey. One naturally wonders whether the assumption of finite expectation only excludes trivial cases or it makes a real difference to main properties of the convex order. This is the primary purpose of the paper.

We start the discussion by recalling a well-known result on the convex order, which is fundamental in the research area of risk aggregation. Let $d \geq 2$, (X_1, \dots, X_d) be a random vector with finite-mean components, and $(X_1^{\text{co}}, \dots, X_d^{\text{co}})$ be a comonotonic random vector such that $X_i^{\text{co}} \stackrel{d}{=} X_i$ for each $i \in \{1, \dots, d\}$.¹ We have (e.g., Dhaene et al. (2002, Theorem 7))

$$\mathbb{E}[f(X_1 + \dots + X_d)] \leq \mathbb{E}[f(X_1^{\text{co}} + \dots + X_d^{\text{co}})] \quad (1)$$

for all convex functions f such that the two expectations are finite.

The relation (1) is often seen as a special case of a more general relation²

$$\mathbb{E}[\varphi(X_1, \dots, X_d)] \leq \mathbb{E}[\varphi(X_1^{\text{co}}, \dots, X_d^{\text{co}})] \quad (2)$$

for all supermodular functions φ such that the two expectations are finite.

The relation (2) is widely believed to hold true; see e.g., Müller and Stoyan (2002, Theorem 3.9.8, Property 5) and we list several other references later. The connection between (1) and (2) can be seen from the fact that $(x_1, \dots, x_d) \mapsto f(x_1 + \dots + x_d)$ is supermodular whenever f is convex.

¹Definitions of comonotonicity and counter-monotonicity are formally presented in Section 4.

²A function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ is *supermodular* if $\varphi(\mathbf{x}) + \varphi(\mathbf{y}) \leq \varphi(\mathbf{x} \wedge \mathbf{y}) + \varphi(\mathbf{x} \vee \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, where \wedge represents taking the componentwise minimum and \vee , the componentwise maximum; see Section 5.

Our first, perhaps surprising, example shows that (2) is actually not true, at least when $d \geq 3$. As a consequence, (1) does not directly follow from (2).

Example 1. Consider the random vector (X, Y, Z) where all three random variables follow a standard uniform distribution. Assume that (X, Y) is counter-monotonic, and $(1/X + 1/Y, Z)$ is counter-monotonic. Write $V = 2/Z$ and $W = 1/X + 1/Y$. Consider the function $\varphi : \mathbb{R}^3 \mapsto \mathbb{R}$ given by

$$\varphi(x, y, z) = \frac{1}{x} \mathbb{1}_{\{0 < x < 1\}} + \frac{1}{y} \mathbb{1}_{\{0 < y < 1\}} - \frac{2}{z} \mathbb{1}_{\{0 < z < 1\}}, \quad \text{for } x, y, z \in \mathbb{R}. \quad (3)$$

This function is separable in 3 variables, and therefore it is supermodular. We have

$$\mathbb{E}[\varphi(X, Y, Z)] = \mathbb{E}[W - V] = \int_0^1 (F_W^{-1}(u) - F_V^{-1}(u)) du,$$

where F_W^{-1} and F_V^{-1} denote the left quantile functions of W and V , respectively,³ and the second equality follows from the comonotonicity of (W, V) . One easily finds $F_W^{-1}(u) = 4/(1 - u^2)$ and $F_V^{-1}(u) = 2/(1 - u)$ for $u \in (0, 1)$. Hence,

$$\mathbb{E}[\varphi(X, Y, Z)] = \int_0^1 \frac{4}{1 - u^2} - \frac{2}{1 - u} du = \int_0^1 \frac{2}{1 + u} du = 2 \log 2.$$

For the comonotonic random vector $(X^{\text{co}}, Y^{\text{co}}, Z^{\text{co}})$ with standard uniform marginal distributions, $\varphi(X^{\text{co}}, Y^{\text{co}}, Z^{\text{co}}) = 0$, which yields $\mathbb{E}[\varphi(X^{\text{co}}, Y^{\text{co}}, Z^{\text{co}})] = 0 < \mathbb{E}[\varphi(X, Y, Z)]$. Therefore, (2) does not hold for the bounded random vectors (X, Y, Z) and $(X^{\text{co}}, Y^{\text{co}}, Z^{\text{co}})$.

In Example 1, the unboundedness of φ allowed for the uniform random variables at play to be transformed into infinite-mean Pareto random variables, whose sum has an unusual behavior under negative dependence, as recently studied by [Chen et al. \(2025b\)](#). Their Theorem 1 states that W stochastically dominates V ; thus, $\varphi(X, Y, Z) > 0$ almost surely, and with no surprise it has a positive expectation. A similar example to Example 1, involving Cauchy distributions, can be built from [Simons \(1977\)](#). Other counter-examples can be constructed from the fact that the sum of standard Cauchy random variables can be a non-zero constant; see [Puccetti et al. \(2019, Theorem 4.2\)](#). Although providing useful examples, none of the above papers pointed out that (2) fails to hold. Note that the relation (2) also fails for strictly supermodular functions, by adding a very small strictly supermodular term to φ in the example.

Example 1 seems to refute many statements of (2) in well-cited papers, including [Puccetti and Wang \(2015, Theorem 2.1\(d\)\)](#) (from one of the authors of the current paper), [Shaked and Shanthikumar \(2007, Theorem 9.A.21\)](#), [Müller and Stoyan \(2002, Theorem 3.9.8, Property 5\)](#),

³Formally, $F_X^{-1}(t) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq t\}$ for $t \in (0, 1)$ and any random variable X .

Müller and Scarsini (2000, Theorem 3.1), Demuit et al. (2005, Proposition 6.3.7) and Rüschendorf (2013, Theorem 6.14). Most of these results relied on the upper bound for expectations of supermodular functions provided in Tchen (1980, Theorem 5). However, requirements on the integrability of the functions used by Tchen (1980) were stricter than (2). Proper restrictions on φ for (2) to hold can be found in optimal transport (OT) theory as in Villani (2009, Theorem 5.10) and Rachev and Rüschendorf (2006, Theorem 3.1.2); these conditions involve some form of regularity and integrability. For instance, the inequality in (2) holds true for φ that is bounded and lower semicontinuous. This issue will be formally discussed in Section 5.

A key feature of Example 1 is that $\mathbb{E}[\varphi(X, Y, Z)]$ involves terms that have infinite expectations, although these terms may be canceled for some particular dependence structures of (X, Y, Z) as in Example 1. If the standard uniform random variables X, Y, Z are independent, then $\mathbb{E}[\varphi(X, Y, Z)]$ is actually undefined, as it involves $\infty - \infty$. Therefore, the real trouble to cause (2) to fail is the presence of infinite mean.

With (2) failing when infinite-mean terms arise, one naturally wonders whether (1) also fails when the random variables do not have a finite mean. It turns out that (1) still holds true, even when we allow infinite expectations of $\mathbb{E}[f(X_1, \dots, X_n)]$ and $\mathbb{E}[f(X_1^{\text{co}}, \dots, X_n^{\text{co}})]$; this result is presented as Theorem 5. Notably, when comparing two random variables X and Y in convex order, requiring $\mathbb{E}[f(X)]$ and $\mathbb{E}[f(Y)]$ to be finite or not leads to two distinct definitions of the convex order, co-existing in the literature and commonly believed to be equivalent. This issue is formally discussed in Section 2. We show that both definitions result in the same partial order on the space of finite-mean random variables, but the equivalence does not hold if the means of random variables are allowed to be infinite.

In Section 3, we give some equivalent conditions for the convex order to hold for infinite-mean random variables, which are slightly different from common conditions used for finite-mean random variables. The obtained conditions are useful to establish some later results.

We proceed to study convex order relations between sums of random variables in Section 4. In dimension $d = 2$, (1) becomes the upper bound in the following relation (e.g., Rüschendorf (2013, Corollary 3.28)): For any X, Y with finite mean,

$$\mathbb{E}[f(X^{\text{ct}} + Y^{\text{ct}})] \leq \mathbb{E}[f(X + Y)] \leq \mathbb{E}[f(X^{\text{co}} + Y^{\text{co}})], \quad (4)$$

for all convex functions f such that the three expectations are finite,

where $(X^{\text{co}}, Y^{\text{co}})$ is comonotonic, $(X^{\text{ct}}, Y^{\text{ct}})$ is counter-monotonic, $X^{\text{co}}, X^{\text{ct}}$, and X are identically distributed, and $Y^{\text{co}}, Y^{\text{ct}}$, and Y are identically distributed. A first version of this result, formulated quite differently, may be traced back at least to Lorentz (1953); for a further historical remark, see Puccetti and Wang (2015). Proofs of (4) in the literature rely heavily on the

assumption, although sometimes not emphasized, that the means of X and Y are finite. This is the case, for instance, for [Kaas et al. \(2000, Proposition 2\)](#), [Dhaene et al. \(2002, Theorem 7\)](#) and [Denuit et al. \(2005, Section 3.4\)](#), all focusing on the bound provided by the comonotonic version. By removing the assumption of finiteness of means, we establish (1), (4), and a corresponding lower bound in dimension $d \geq 3$, with the greatest generality.

As mentioned earlier, Section 5 addresses the supermodular order and its connection to OT theory, where we obtain several results that hold for random variables with infinite mean. As we can see from the discussions in Section 5, many questions remain open, including whether (2) holds when $d = 2$. Section 6 discusses implications of our main results for risk management and decision making. Whereas short proofs are kept within the corresponding sections, long proofs are provided in Section 7.

2 Two definitions of convex order

Let L^0 be the space of all random variables on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$, L^1 be the space of all random variables with finite mean, and L^∞ be the space of all essentially bounded random variables. We write $X \stackrel{d}{=} Y$ to represent that X and Y follow the same distribution. Throughout, by saying that an expectation $\mathbb{E}[X]$ is well-defined, we mean $\mathbb{E}[X_+] < \infty$ or $\mathbb{E}[X_-] < \infty$, where $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$ for $x \in \mathbb{R}$. Denote by \mathcal{U}_{cx} the set of all convex functions on \mathbb{R} .

Although it may not be obvious to many researchers, two definitions of the convex order co-exist in the literature, and they are deceptively similar.

Definition 1. For $X, Y \in L^0$, we say that Y dominates X

- (i) in *convex order*, denoted by $X \leq_{\text{cx}} Y$, if $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all $u \in \mathcal{U}_{\text{cx}}$ such that the two expectations are well-defined;
- (ii) in *†convex order*, denoted by $X \leq_{\text{cx}}^\dagger Y$, if $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all $u \in \mathcal{U}_{\text{cx}}$ such that the two expectations are finite.

The definition as per item (i) is found in [Müller and Stoyan \(2002\)](#), notably; the one as per item (ii) is found in [Rüschendorf \(2013\)](#), [Denuit et al. \(2005\)](#) and [Shaked and Shanthikumar \(2007\)](#). We added a dagger † to dispel possible ambiguity as to which definition we refer to henceforth. The difference is whether one must consider functions in \mathcal{U}_{cx} rendering infinite expectations to establish comparisons: one needs to for \leq_{cx} but not for \leq_{cx}^\dagger . Because in the latter case $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ is required for a smaller subset of functions $u \in \mathcal{U}_{\text{cx}}$, we immediately

have $X \leq_{\text{cx}} Y \implies X \leq_{\text{cx}}^\dagger Y$. When comparing random variables on L^∞ , the definitions are obviously equivalent since all $u \in \mathcal{U}_{\text{cx}}$ inevitably produce finite expectations.

The most plausible cause for this discrepancy to have arisen is that most versions of the convex order's definition state that the expectations must *exist*, which can refer to either expectations be well-defined or finite depending on the authors' wording choices. In Müller and Stoyan (2002) and Rachev and Rüschendorf (2006), existence of expectations means they are well-defined; for Denuit et al. (2005) and Shaked and Shanthikumar (2007), existence means finiteness.⁴ This semantic inconsistency is not the product of either set of textbooks and is in fact not circumscribed to the literature on stochastic orders. Even in the standard probability textbook Durrett (2019, Fifth edition), while it is declared that an expectation that exists may be infinite (p. 28), in some places the existence of an expectation is understood as finiteness (p. 29).

Another plausible cause for the discrepancy is that the nuance may have come off as inconsequential. In the following proposition, we show that the two definitions coincide when means are finite. It is not the case, however, when on the space of all random variables.

Proposition 1. *On L^1 , the two relations \leq_{cx} and \leq_{cx}^\dagger are equivalent. On L^0 , \leq_{cx} implies \leq_{cx}^\dagger , but the converse is not true.*

Proof. As discussed above, the implication $X \leq_{\text{cx}} Y \implies X \leq_{\text{cx}}^\dagger Y$ holds under both settings of L^1 and L^0 . We first show the converse holds on L^1 . By Theorems 1.5.3 and 1.5.7 of Müller and Stoyan (2002), we have, for $X, Y \in L^1$,

$$X \leq_{\text{cx}} Y \iff \mathbb{E}[X] = \mathbb{E}[Y] \text{ and } \mathbb{E}[(X - w)_+] \leq \mathbb{E}[(Y - w)_+] \text{ for all } w \in \mathbb{R}. \quad (5)$$

By Proposition 3.4.3 of Denuit et al. (2005) we have, for $X, Y \in L^1$,

$$X \leq_{\text{cx}}^\dagger Y \iff \mathbb{E}[X] = \mathbb{E}[Y] \text{ and } \mathbb{E}[(X - w)_+] \leq \mathbb{E}[(Y - w)_+] \text{ for all } w \in \mathbb{R}. \quad (6)$$

Putting (5) and (6) together, we get $X \leq_{\text{cx}} Y \iff X \leq_{\text{cx}}^\dagger Y$. To show that \leq_{cx}^\dagger does not imply \leq_{cx} on L^0 , a counter-example is provided in Example 2 below. \square

Because the \dagger convex order gets rid of infinite expectations, whether positive or negative, one can design a comparison scheme where all elements $u \in \mathcal{U}_{\text{cx}}$, except for constant functions, are removed because of either random variable.

⁴They do not explicitly indicate what the existence of expectations signifies, but it can be inferred from other results, e.g., Property 1.8.9 and its proof in Denuit et al. (2005) and Example 1.B.23 in Shaked and Shanthikumar (2007).

Example 2. Consider two random variables $X \geq 0$ and $Y \leq 0$ with $\mathbb{E}[X] = -\mathbb{E}[Y] = \infty$. Any convex function u with $\mathbb{E}[u(X)] < \infty$ must not have any strictly increasing part, and that with $\mathbb{E}[u(Y)] < \infty$ must not have any strictly decreasing part. With both requirements, we are left only with the constant functions, for which $\mathbb{E}[u(X)] = \mathbb{E}[u(Y)]$ clearly holds, and this implies both $X \leq_{\text{cx}}^{\dagger} Y$ and $Y \leq_{\text{cx}}^{\dagger} X$. Meanwhile, for convex functions $u_1 : x \mapsto x_+$ and $u_2 : x \mapsto x_-$, we note $\mathbb{E}[u_1(X)] = \infty > 0 = \mathbb{E}[u_1(Y)]$ and $\mathbb{E}[u_2(X)] = 0 < \infty = \mathbb{E}[u_2(Y)]$; therefore, neither $X \leq_{\text{cx}} Y$ nor $Y \leq_{\text{cx}} X$ holds.

Müller and Stoyan (2002) restricted their definition of the convex order to random variables in L^1 , based on limitations of the convex order beyond L^1 identified by Elton and Hill (1992). Given Proposition 1, Müller and Stoyan (2002), Denuit et al. (2005), Shaked and Shanthikumar (2007), and Rüschendorf (2013), as well as many others, all effectively work under the same definition on L^1 .

Example 2 illustrates that $X, Y \in L^0$ may be comparable with respect to the \dagger convex order even if $\mathbb{E}[X] \neq \mathbb{E}[Y]$, when these means are infinite. It moreover showcases that $\leq_{\text{cx}}^{\dagger}$ does not satisfy antisymmetry on L^0 , as $X \leq_{\text{cx}}^{\dagger} Y$ and $Y \leq_{\text{cx}}^{\dagger} X$ despite $X \not\stackrel{\text{d}}{=} Y$. Neither does the order \leq_{cx} satisfy antisymmetry on L^0 ; see Example 3 below.

Example 3. Let X follow a standard Cauchy distribution, and $Y = 2X$. Recall that the Cauchy distribution has no finite moment and is symmetric; in this vein, no convex function u yields finite expectations $\mathbb{E}[u(X)]$ and $\mathbb{E}[u(Y)]$ besides u constant. If expectations are well-defined, then $\mathbb{E}[u(X)] = \mathbb{E}[u(Y)] = \infty$. Hence, $X \leq_{\text{cx}} Y$ and $Y \leq_{\text{cx}} X$ hold. The two random variables, however, are not identically distributed.

We should note that for random variables X and Y that do not have a finite mean, $Y \leq_{\text{cx}} X$ loses the classic interpretation that the distribution of X is more spread out than that of Y . For instance, in the above example, $Y = 2X$ and X has a symmetric distribution with respect to 0. Thus, the distribution of Y is more spread out than that of X , despite $Y \leq_{\text{cx}} X$.

The following example was inspired from the discussion in Müller (1997a) and showcases that $\leq_{\text{cx}}^{\dagger}$ does not satisfy transitivity.

Example 4. Let $X \in L^0$ follow a standard Cauchy distribution, and $Y, Z \in L^1$ be such that $\mathbb{E}[Y] \neq \mathbb{E}[Z]$. Because $\mathbb{E}[u(X)]$ is finite for no $u \in \mathcal{U}_{\text{cx}}$ other than constant ones, we have the chain of relations $Y \leq_{\text{cx}}^{\dagger} X \leq_{\text{cx}}^{\dagger} Z \leq_{\text{cx}}^{\dagger} X \leq_{\text{cx}}^{\dagger} Y$. Yet, from (6), $Y \not\leq_{\text{cx}}^{\dagger} Z$ and $Z \not\leq_{\text{cx}}^{\dagger} Y$.

The flaw exhibited in Example 4 does not spread to \leq_{cx} , as we attest by the next proposition. Still, \leq_{cx} is not a partial order on L^0 because it fails to satisfy antisymmetry; in contrast, \leq_{cx} and $\leq_{\text{cx}}^{\dagger}$ are partial orders on L^1 .

Proposition 2. *The relation \leq_{cx} is a preorder on L^0 .*

Proof. The reflexivity of \leq_{cx} follows trivially from definition; remains to show transitivity. Consider $X, Y, Z \in L^0$ and suppose $X \leq_{\text{cx}} Y$ and $Y \leq_{\text{cx}} Z$. Let $u \in \mathcal{U}_{\text{cx}}$ be such that $\mathbb{E}[u(X)]$ and $\mathbb{E}[u(Z)]$ are well-defined. If $\mathbb{E}[u(Y)]$ is also well-defined, then $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)] \leq \mathbb{E}[u(Z)]$. If $\mathbb{E}[u(Y)]$ is not well-defined, then $\infty = \mathbb{E}[u(Y)_+] \leq \mathbb{E}[u(Z)_+]$ since the function $x \mapsto u(x)_+$ is convex. Hence, $\mathbb{E}[u(Z)] = \infty$ and thus $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Z)]$. This shows $X \leq_{\text{cx}} Z$. \square

The convex order falls within the more general class of integral stochastic orders (Whitt (1986); Müller (1997a)). An integral stochastic order $\leq_{\mathfrak{F}}$, indexed by a set \mathfrak{F} of real-valued functions, is defined on a set \mathcal{L} of random vectors taking values in \mathbb{R}^d via

$$X \leq_{\mathfrak{F}} Y \iff \mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)] \text{ for all } u \in \mathfrak{F}; \quad X, Y \in \mathcal{L}. \quad (7)$$

The set \mathcal{L} is called the domain of $\leq_{\mathfrak{F}}$. The convex order and the supermodular order (see Section 5) are instances of integral stochastic orders with $d = 1$. To avoid situations such as in Example 4 and ensure integral orders be preorders, the setting of Müller (1997a) additionally requires

$$\text{there exists } b_{\mathfrak{F}} : \mathbb{R}^d \rightarrow [1, \infty) \text{ such that } \sup_{x \in \mathbb{R}^d} \frac{|u(x)|}{b_{\mathfrak{F}}(x)} < \infty \text{ for all } u \in \mathfrak{F}, \quad (8)$$

and excludes from the set \mathcal{L} every X with $\mathbb{E}[b_{\mathfrak{F}}(X)] = \infty$. Expectations in (7) are then finite for all $X, Y \in \mathcal{L}$. The set \mathcal{L} is restricted by \mathfrak{F} via the function $b_{\mathfrak{F}}$. Note that our definitions of \leq_{cx} and \leq_{cx}^\dagger do not restrict the set \mathcal{L} . When defined on L^1 , \leq_{cx} and \leq_{cx}^\dagger are consistent with the above restriction: Taking $\mathfrak{F} = \{I_{\mathbb{R}}, -I_{\mathbb{R}}\} \cup \{x \mapsto (x - w)_+ : w \in \mathbb{R}\}$, where $I_{\mathbb{R}}$ is the identity on \mathbb{R} , one can take $b_{\mathfrak{F}} = I_{\mathbb{R}} \vee 1$ and $\mathcal{L} = L^1$ in (8). By (5) and (6), the relation $\leq_{\mathfrak{F}}$ is equivalent to \leq_{cx} and \leq_{cx}^\dagger on L^1 .

An advantage of the restriction (8) for integral stochastic orders is that one never has to worry about infinite expectations and the troubles arising along. The drawbacks are limitations on the applicability of these orders. For instance, one cannot consider the complete set of random variables in L^0 unless \mathfrak{F} only contains bounded functions; note that bounded convex functions are constant. As we will discuss in the upcoming sections, \leq_{cx} continues to satisfy main properties of the convex order on L^0 . Therefore, narrowing arbitrarily the domain for \leq_{cx} seems unnecessary. This is indeed one of the main arguments of the paper: one need not limit to L^1 . In Section 5, we will discuss how, on the contrary, a narrowing of the domain can be beneficial for the supermodular order.

3 Equivalent conditions for the convex order

There are many conditions to verify convex order. A common equivalent condition for $X \leq_{\text{cx}} Y$ when $X, Y \in L^1$ is given in (5); another one is (Shaked and Shanthikumar (2007, Theorem 3.A.5))

$$\mathbb{E}[X] = \mathbb{E}[Y] \quad \text{and} \quad \int_p^1 F_X^{-1}(t) dt \leq \int_p^1 F_Y^{-1}(t) dt \quad \text{for all } p \in (0, 1). \quad (9)$$

Condition (9), as well as the second condition in (5), is not sufficient for either $X \leq_{\text{cx}} Y$ or $X \leq_{\text{cx}}^\dagger Y$ with $X, Y \in L^0$, because (9) holds as soon as $\mathbb{E}[X_+] = \mathbb{E}[Y_+] = \infty$, $\mathbb{E}[X_-] < \infty$, and $\mathbb{E}[Y_-] < \infty$, but $X \leq_{\text{cx}}^\dagger Y$ and $X \leq_{\text{cx}} Y$ fail to hold when $\mathbb{E}[X_-] > \mathbb{E}[Y_-]$ because $x \mapsto x_-$ is convex. Moreover, even when the means are well-defined, (9) is not necessary for $X \leq_{\text{cx}}^\dagger Y$ as we see in Example 2. Nevertheless, similar conditions to (5) and (9) turn out to be equivalent to $X \leq_{\text{cx}} Y$ for all $X, Y \in L^0$.

Theorem 1. *For $X, Y \in L^0$, the following are equivalent:*

- (i) $X \leq_{\text{cx}} Y$;
- (ii) *it holds that*

$$\int_0^p F_X^{-1}(t) dt \geq \int_0^p F_Y^{-1}(t) dt \quad \text{and} \quad \int_p^1 F_X^{-1}(t) dt \leq \int_p^1 F_Y^{-1}(t) dt \quad \text{for all } p \in (0, 1); \quad (10)$$

- (iii) *it holds that*

$$\mathbb{E}[(X - w)_-] \leq \mathbb{E}[(Y - w)_-] \quad \text{and} \quad \mathbb{E}[(X - w)_+] \leq \mathbb{E}[(Y - w)_+] \quad \text{for all } w \in \mathbb{R}. \quad (11)$$

We note that, regardless of whether X and Y have finite mean, the integrals and expectations in (10) and (11) are always well-defined, as each of them integrates a function that is bounded from one side. The proof of Theorem 1 relies on Lemma 1 below, which gives a simplification of the convex order when one of the two random variables at comparison is outside L^1 . Let us present a few more stochastic orders first.

Definition 2. Let \mathcal{U}_{st} (resp. \mathcal{U}_{icx} , \mathcal{U}_{dcx}) be the set of all increasing (resp. increasing convex, decreasing convex) real functions on \mathbb{R} . For random variables $X, Y \in L^0$ and $* \in \{\text{st}, \text{icx}, \text{dcx}\}$, we write $X \leq_* Y$ if $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all $u \in \mathcal{U}_*$ such that the expectations are well-defined.

Lemma 1. *For $X, Y \in L^0$,*

- (i) *if $\mathbb{E}[Y_+] = \infty$, then $X \leq_{\text{cx}} Y$ is equivalent to $X \leq_{\text{dcx}} Y$;*

- (ii) if $\mathbb{E}[Y_-] = \infty$, then $X \leq_{\text{cx}} Y$ is equivalent to $X \leq_{\text{icx}} Y$;
- (iii) if $\mathbb{E}[Y_+] = \mathbb{E}[Y_-] = \infty$, then $Y \leq_{\text{cx}}^\dagger X \leq_{\text{cx}} Y$ always holds.

By Lemma 1, for random variables $X, Y \in L^0 \setminus L^1$, one may observe $X \geq_{\text{st}} Y$ meanwhile $X \leq_{\text{cx}} Y$. This situation cannot occur for X, Y with finite means (unless $X \stackrel{\text{d}}{=} Y$) since convex ordering then requires $\mathbb{E}[X] = \mathbb{E}[Y]$, from (9). The convex order has been mainly designed and hitherto employed to compare the variability of distributions. On L^1 , the variability may be vaguely interpreted as some form of deviation from the mean. When the mean is infinite, for instance when $\mathbb{E}[Y] = \infty$, this deviation is larger when the distribution is stochastically smaller, and hence \leq_{cx} becomes \leq_{dcx} . Therefore, the interpretation of convex order as a comparative tool for deviation from the mean remains valid even on L^0 , as we see in parts (i) and (ii) of Lemma 1.

Each of (10) and (11) is sufficient, but not necessary, for $X \leq_{\text{cx}}^\dagger Y$ on L^0 . Example 2 illustrates one instance in which (10) and (11) fail but $X \leq_{\text{cx}}^\dagger Y$ holds. The following theorem provides necessary conditions to establish \dagger convex order in the setting of L^0 .

Theorem 2. *For $X, Y \in L^0$, the following are equivalent:*

- (i) $X \leq_{\text{cx}}^\dagger Y$;
- (ii) the inequalities in (10) hold with the relaxation that $\infty \leq x$ and $x \leq -\infty$ are treated as true for all $x \in [-\infty, \infty]$;
- (iii) the inequalities in (11) hold with the relaxation that $\infty \leq x$ and $x \leq -\infty$ are treated as true for all $x \in [-\infty, \infty]$.

Comparing Theorems 1 and 2, we can get the conclusion in Proposition 1 that \leq_{cx} and \leq_{cx}^\dagger are equivalent on L^1 but not on L^0 .

The conditions' relaxation in items (ii) and (iii) may be deemed unnatural, and it gives rise to peculiar cases where a positive random variable may be comparable to a negative random variable in \dagger convex order, as in Example 2. To exclude these incongruent cases, one might prefer using \leq_{cx} over \leq_{cx}^\dagger in the setting of L^0 . In other words, when the means of X and Y are infinite, the \dagger convex order is not very well-suited to establish comparisons as its requirement on the finiteness of $\mathbb{E}[u(X)]$ and $\mathbb{E}[u(Y)]$ excludes too many $u \in \mathcal{U}_{\text{cx}}$ for the comparison $X \leq_{\text{cx}}^\dagger Y$ to be practically useful.

Theorems 1 and 2 also imply that both \leq_{cx} and \leq_{cx}^\dagger are closed under distributional mixtures on L^0 . Note that this does not directly follow from the definition (whereas the case on L^∞ follows directly by definition), because the functions $u \in \mathcal{U}_{\text{cx}}$ that needs to be checked change when distributions are mixed.

4 Convex order between sums of random variables

We first define a few classic concepts of dependence that are relevant for convex order relations. A pair of random variables (X, Y) is *comonotonic* if $X = f(Z)$ and $Y = g(Z)$ almost surely for some random variable Z and increasing functions f and g . A pair of random variables (X, Y) is *counter-monotonic* if $(X, -Y)$ is comonotonic. Comonotonicity and counter-monotonicity for random vectors are meant pairwise. For a random vector (X_1, \dots, X_d) , we say that (X'_1, \dots, X'_d) is its *comonotonic* (resp. *counter-monotonic*) *version* if (X'_1, \dots, X'_d) is comonotonic (resp. counter-monotonic) and $X_i \stackrel{d}{=} X'_i$ for each $i \in [d] = \{1, \dots, d\}$. For any given random vector, its comonotonic version always exists, whereas its counter-monotonic version may not exist for $d \geq 3$ (Dall'Aglio (1972)).

As mentioned in the Introduction, comonotonicity and counter-monotonicity are connected to convex order on L^1 through $X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} X + Y \leq_{\text{cx}} X^{\text{co}} + Y^{\text{co}}$. Our next result shows that this connection continues to hold on L^0 .

Theorem 3. *For any $X, Y \in L^0$, we have*

$$X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} X + Y \leq_{\text{cx}} X^{\text{co}} + Y^{\text{co}}, \quad (12)$$

where $(X^{\text{co}}, Y^{\text{co}})$ is a comonotonic version of (X, Y) and $(X^{\text{ct}}, Y^{\text{ct}})$ is a counter-monotonic version of (X, Y) .

Since $Z \leq_{\text{cx}} W \implies Z \leq_{\text{cx}}^\dagger W$, the result of Theorem 3 is also valid for \leq_{cx}^\dagger . Another implication of Theorem 3 is the following corollary, previously shown by Simons (1977).

Corollary 1 (Simons (1977)). *Consider $X, Y, Z, W \in L^0$ and suppose $X \stackrel{d}{=} Z$ and $Y \stackrel{d}{=} W$. If $\mathbb{E}[X + Y]$ and $\mathbb{E}[Z + W]$ are both well-defined, then $\mathbb{E}[X + Y] = \mathbb{E}[Z + W]$.*

Proof. Note that the convex order relation $X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} X + Y$ from Theorem 3 implies $\mathbb{E}[(X^{\text{ct}} + Y^{\text{ct}})_+] \leq \mathbb{E}[(X + Y)_+]$ and $\mathbb{E}[(X^{\text{ct}} + Y^{\text{ct}})_-] \leq \mathbb{E}[(X + Y)_-]$ by (11). Hence, $\mathbb{E}[X^{\text{ct}} + Y^{\text{ct}}]$ is well-defined because $\mathbb{E}[X + Y]$ is well-defined. With this and the relations $X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} X + Y$ and $X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} Z + W$, we get $\mathbb{E}[X^{\text{ct}} + Y^{\text{ct}}] = \mathbb{E}[X + Y] = \mathbb{E}[Z + W]$ since both $x \mapsto x$ and $x \mapsto -x$ are convex. \square

Note that the proof of Corollary 1 requires the ordering in (12) for \leq_{cx} specifically (rather than for \leq_{cx}^\dagger), as we compare possibly infinite expectations. While Corollary 1 appears evident on L^1 , it is not trivial on L^0 since the relation $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ may not hold because one random variable's mean is ∞ and the other's is $-\infty$ or because at least one has an undefined mean. The noteworthiness of the result hence lies in that two random variables' infinite means

cannot offset each other in many different ways so that their sum's mean varies upon the dependence scheme. This is curiously not the case when summing three or more random variables, as exhibited in Example 1 and discussed by Simons (1977). The proof techniques of Simons (1977) are similar to the ones of Theorem 3, and in fact they inspired our proof.

We next compare sums of comonotonic random variables. This relation on L^1 , stated for $d \geq 2$ random variables, was presented in Dhaene et al. (2002, Corollary 1).

Theorem 4. For $X, Y, Z, W \in L^0$, if $X \leq_{\text{cx}} Z$ and $Y \leq_{\text{cx}} W$, then

$$X + Y \leq_{\text{cx}} Z^{\text{co}} + W^{\text{co}},$$

where $(Z^{\text{co}}, W^{\text{co}})$ is a comonotonic version of (Z, W) .

Proof. Using Theorem 3, it suffices to prove the result when (X, Y) is comonotonic. This result follows directly by using the equivalent condition (10) in Theorem 1 and the relation $F_{X+Y}^{-1}(t) = F_X^{-1}(t) + F_Y^{-1}(t)$, $t \in (0, 1)$, for comonotonic random variables. \square

Remark 1. The same conclusion in Theorem 4 holds also for \leq_{cx}^\dagger following a similar proof, which we omit.

A potential counterpart to Theorem 4, namely $X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} Z + W$ under the same condition, fails to hold even for random variables in L^∞ ; see Cheung et al. (2014, Example 1). An intuitive explanation is that $Z + W$ may be a constant, but $X^{\text{ct}} + Y^{\text{ct}}$ is not necessarily a constant under the given conditions.

Using Theorem 4, we can immediately show that the upper bound in (12) can be generalized to an arbitrary dimension d on L^0 .

Theorem 5. Let $d \in \mathbb{N}$. For any $X_1, \dots, X_d \in L^0$, we have

$$X_1 + \dots + X_d \leq_{\text{cx}} X_1^{\text{co}} + \dots + X_d^{\text{co}}, \tag{13}$$

where $(X_1^{\text{co}}, \dots, X_d^{\text{co}})$ is a comonotonic version of (X_1, \dots, X_d) .

Proof. Because sums of any number a comonotonic random vector's components are comonotonic, the result is obtained by combining iterative applications of Theorems 3 and 4. \square

As mentioned in Section 3, for $X, Y \in L^0 \setminus L^1$, it is possible that $X \leq_{\text{cx}} Y$ meanwhile $Y \leq_{\text{st}} X$. The implication of this phenomenon in risk management and decision making is discussed in Section 6.

In the next result, we show that the lower bound in (12) can also be generalized to an arbitrary dimension d . This result requires a separate proof because there is no counter-monotonic

counterpart of Theorem 4 (which is used to prove Theorem 5). When $d \geq 3$ and X_1, \dots, X_d have at least 3 non-degenerate components, this result follows from a more general conclusion in Theorem 7, addressed in the next section, and in the setting of L^1 it is given in Cheung and Lo (2014, Theorem 5.1).

Theorem 6. *Let $d \in \mathbb{N}$. For $X_1, \dots, X_d \in L^0$, we have*

$$X_1^{\text{ct}} + \dots + X_d^{\text{ct}} \leq_{\text{cx}} X_1 + \dots + X_d,$$

if a counter-monotonic version $(X_1^{\text{ct}}, \dots, X_d^{\text{ct}})$ of (X_1, \dots, X_d) exists.

Let us repeat that \leq_{cx} implies \leq_{cx}^\dagger , and hence, the statements of Theorems 5 and 6 also hold true for the \dagger convex order. If the marginal distributions of (X_1, \dots, X_d) do not allow for a counter-monotonic version to exist, the set of possible distributions of $X_1 + \dots + X_d$ with given marginal distributions of X_1, \dots, X_d may possibly not admit a smallest element with respect to \leq_{cx} or \leq_{cx}^\dagger , as discussed in Section 3.2 of Bernard et al. (2014) with a discrete example.

5 Optimal transport and supermodular order

Theorem 3 can be interpreted as a result in OT with infinite mean. For given distributions of $X, Y \in L^0$, the classical Kantorovich OT problem is to find a joint distribution of (X, Y) , called a transport plan, that minimizes the expectation $\mathbb{E}[c(X, Y)]$ for a given real-valued cost function c on \mathbb{R}^2 . It is straightforward to check that functions of the form $(x, y) \mapsto u(x + y)$ are supermodular for $u \in \mathcal{U}_{\text{cx}}$; see Müller (1997b, Theorem 3.1). Hence, for any cost function of the form $c : (x, y) \mapsto u(x + y)$ with $u \in \mathcal{U}_{\text{cx}}$, the optimal transport plan is given by the distribution of $(X^{\text{ct}}, Y^{\text{ct}})$. In the same vein, the distribution of $(X^{\text{co}}, Y^{\text{co}})$ is the optimal transport plan for functions of the form $c : (x, y) \mapsto u(x - y)$ with $u \in \mathcal{U}_{\text{cx}}$. The quadratic optimal cost $(x - y)^2$ and the absolute value cost $|x - y|$ are examples of such functions. Theorem 3 implies that the above conclusions hold even if X and Y do not have a mass barycenter. The statement of Theorem 3 connects to a fundamental OT result for supermodular cost functions, as we discuss below.

Recall that a function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ is *supermodular* if

$$\varphi(\mathbf{x}) + \varphi(\mathbf{y}) \leq \varphi(\mathbf{x} \wedge \mathbf{y}) + \varphi(\mathbf{x} \vee \mathbf{y}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,$$

where \wedge represents taking the componentwise minimum and \vee , the componentwise maximum. We denote by Φ_{sm}^d the set of all supermodular functions on \mathbb{R}^d .

Definition 3. For $\mathbf{X}, \mathbf{Y} \in (L^0)^d$, $d \in \mathbb{N}$, we say that \mathbf{Y} dominates \mathbf{X}

- (i) in *supermodular order*, denoted by $\mathbf{X} \leq_{\text{sm}} \mathbf{Y}$, if $\mathbb{E}[\varphi(\mathbf{X})] \leq \mathbb{E}[\varphi(\mathbf{Y})]$ for all $\varphi \in \Phi_{\text{sm}}^d$ such that the two expectations are well-defined;
- (ii) in \dagger *supermodular order*, denoted by $\mathbf{X} \leq_{\text{sm}}^{\dagger} \mathbf{Y}$, if $\mathbb{E}[\varphi(\mathbf{X})] \leq \mathbb{E}[\varphi(\mathbf{Y})]$ for all $\varphi \in \Phi_{\text{sm}}^d$ such that the two expectations are finite;
- (iii) in \diamond *supermodular order*, denoted by $\mathbf{X} \leq_{\text{sm}}^{\diamond} \mathbf{Y}$, if $\mathbb{E}[\varphi(\mathbf{X})] \leq \mathbb{E}[\varphi(\mathbf{Y})]$ for all bounded and right-continuous $\varphi \in \Phi_{\text{sm}}^d$;
- (iv) in *concordance order*, denoted by $\mathbf{X} \leq_{\text{c}} \mathbf{Y}$, if $F_{\mathbf{X}}(\mathbf{x}) \leq F_{\mathbf{Y}}(\mathbf{x})$ and $\bar{F}_{\mathbf{X}}(\mathbf{x}) \leq \bar{F}_{\mathbf{Y}}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$, where $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$ are the joint distribution functions of \mathbf{X} and \mathbf{Y} , respectively, and $\bar{F}_{\mathbf{X}}$ and $\bar{F}_{\mathbf{Y}}$ are their joint survival functions.

Note that there was no value in defining a “ \diamond convex” order in the previous sections because bounded convex functions on \mathbb{R} are constant. In the \diamond supermodular order, we required the right continuity of φ , so that it is consistent with Tchen (1980) and Rachev and Rüschendorf (2006).

Remark 2. The restriction (8) for integral stochastic orders in Section 2 allows for a continuum of orders ranging from $\leq_{\text{sm}}^{\diamond}$ to \leq_{sm} , depending on the function $b_{\mathfrak{F}}$ circumscribing the set \mathfrak{F} considered in (7). The domain embedding the orders varies accordingly. While the domain of $\leq_{\text{sm}}^{\diamond}$ is indeed $(L^0)^d$, for \leq_{sm} the domain would reduce to random vectors following a discrete distribution with a finite numbers of atoms, which is inconvenient to work with. Since most authors usually define supermodular order or \dagger supermodular order on larger spaces, such as $(L^{\infty})^d$, $(L^1)^d$ or $(L^0)^d$, we will continue to do so in this section and discuss the problems that consequently arise.

For any choice of $w_1, \dots, w_n \in \mathbb{R}$, functions $\mathbf{x} \mapsto \prod_{i=1}^d \mathbb{1}_{\{x_i \leq w_i\}}$ and $\mathbf{x} \mapsto \prod_{i=1}^d \mathbb{1}_{\{x_i > w_i\}}$ are supermodular, bounded and right-continuous. The expected values of these functions respectively yield the joint distribution function and the joint survival function and are evidently always finite. We therefore have the following implications

$$\mathbf{X} \leq_{\text{sm}} \mathbf{Y} \implies \mathbf{X} \leq_{\text{sm}}^{\dagger} \mathbf{Y} \implies \mathbf{X} \leq_{\text{sm}}^{\diamond} \mathbf{Y} \implies \mathbf{X} \leq_{\text{c}} \mathbf{Y} \quad \text{for any } \mathbf{X}, \mathbf{Y} \in (L^0)^d, \quad (14)$$

since, as for the convex order, $\leq_{\text{sm}}^{\dagger}$ considers a smaller subset of functions than \leq_{sm} within Φ_{sm}^d ; idem for $\leq_{\text{sm}}^{\diamond}$. Note that none of the reverse implications in (14) are obvious to be true. Indeed, as we will see below, the reverse directions of the second and the third implications do not hold in general. We are not aware of whether the reverse direction of the first implication in (14) is true.

From the distribution functions and survival functions orderings in the definition of concordance, one immediately sees that $\mathbf{X} \leq_{\text{c}} \mathbf{Y}$ implies $F_{X_i}(x) = F_{Y_i}(x)$, for all $x \in \mathbb{R}$ and

$i \in [d]$. Hence, by (14), all the orders in Definition 3 require \mathbf{X} and \mathbf{Y} to have identical marginal distributions.

In the same vein as our discussions in Section 4 for the convex and \dagger convex orders, we examine whether, for any random vectors, a comonotonic version and a counter-monotonic version provide an upper and a lower bound with respect to \leq_{sm} , \leq_{sm}^\dagger , $\leq_{\text{sm}}^\diamond$ and \leq_c . First, it is straightforward to check for \leq_c , which follows directly from the Fréchet-Hoeffding bounds.

Proposition 3. *Let $d \geq 2$. For any random vector $\mathbf{X} \in (L^0)^d$, we have $\mathbf{X} \leq_c \mathbf{X}^{\text{co}}$, where \mathbf{X}^{co} is a comonotonic version of \mathbf{X} . Moreover, $\mathbf{X}^{\text{ct}} \leq_c \mathbf{X}$, where \mathbf{X}^{ct} is a counter-monotonic version of \mathbf{X} , if it exists.*

The bounds in Proposition 3 transpose directly to $\leq_{\text{sm}}^\diamond$ in dimension 2 thanks to the following equivalence.

Proposition 4. *On $(L^0)^2$, the two relations \leq_c and $\leq_{\text{sm}}^\diamond$ are equivalent. They are not equivalent when $d \geq 3$ on $(L^\infty)^d$.*

Proof. In dimension $d = 2$, by Tchen (1980, Corollary 2.1), for any $X, Y, Z, W \in L^0$, the ordering $F_{X,Y} \leq F_{Z,W}$ implies $\mathbb{E}[\varphi(X, Y)] \leq \mathbb{E}[\varphi(Z, W)]$ for all bounded, right-continuous $\varphi \in \Phi_{\text{sm}}^2$. For an example where $\leq_c \not\Rightarrow \leq_{\text{sm}}^\diamond$ in dimension $d = 3$, see Müller (1997b, Example 1), which also generalizes to $d \geq 3$. \square

Proposition 4 is well-known in the literature, although references, while relying on Tchen (1980)'s proof, often omit his requirements on φ ; notably, $\varphi(X^*, Y^*)$ must be uniformly integrable for all (X^*, Y^*) in the Fréchet class of (X, Y) , that is, the class of random vectors with the specified marginal distributions. A similar result can be found in Rachev and Rüschendorf (2006, Theorem 3.1.2 (a)) with the lighter requirement of $\varphi(X, Y)$, $\varphi(Z, W)$, and also $\varphi(X^\perp, Y^\perp)$ to produce finite expectations, where $X^\perp \stackrel{\text{d}}{=} X$, $Y^\perp \stackrel{\text{d}}{=} Y$ and X^\perp and Y^\perp are independent. Hence, one cannot quite establish the equivalence between \leq_c and \leq_{sm}^\dagger from their result.

Combining Propositions 3 and 4, we immediately arrive at the following conclusion.

Proposition 5. *For any $X, Y \in L^0$, we have*

$$(X^{\text{ct}}, Y^{\text{ct}}) \leq_{\text{sm}}^\diamond (X, Y) \leq_{\text{sm}}^\diamond (X^{\text{co}}, Y^{\text{co}}) \quad (15)$$

where $(X^{\text{co}}, Y^{\text{co}})$ is a comonotonic version of (X, Y) and $(X^{\text{ct}}, Y^{\text{ct}})$ is a counter-monotonic version of (X, Y) .

From an OT perspective, Proposition 5 amounts to the statement that a counter-monotonic transport map is a minimizer for any right-continuous and bounded supermodular cost function

$\varphi \in \Phi_{\text{sm}}^2$. Indeed, this can be extended to lower semicontinuous supermodular functions φ that are bounded from below by a separable form of integrable functions; see Villani (2009, Theorem 5.10). As this requirement is not satisfied by all functions generating \leq_{sm} and \leq_{sm}^\dagger , it is unclear to us whether (15) also holds for \leq_{sm} or \leq_{sm}^\dagger .

More generally, we do not know if \leq_{sm} , \leq_{sm}^\dagger and $\leq_{\text{sm}}^\diamond$ all render the same partial ordering on any one of $(L^\infty)^2$, $(L^1)^2$ or $(L^0)^2$. We suspect this could be the case. Corollary 1 gives evidence towards this conjecture because one thus cannot obtain a bivariate counterpart to Example 1 using separable supermodular functions. Moreover, the four orders satisfy antisymmetry. They are not plagued by the same problem as \leq_{cx} and \leq_{cx}^\dagger that L^0 contains random variables for which there are not enough elements in \mathcal{U}_{cx} yielding finite or well-defined expectations to fully characterize the distribution. Functions $\{\mathbf{x} \mapsto \prod_{i=1}^d \mathbb{1}_{\{x_i \leq w_i\}} : \mathbf{w} \in \mathbb{R}^d\}$ are supermodular, bounded, right-continuous, and the sequence of their expectations fully characterizes the distributions. As a result, one cannot design counter-examples similar to the one in Example 2 where the \dagger convex order relation holds simply because only constant functions were to be compared.

As indicated in Proposition 3, when $d \geq 3$, the concordance and \diamond supermodular orders are different. The \diamond supermodular order $\leq_{\text{sm}}^\diamond$ is also distinct from \leq_{sm} and \leq_{sm}^\dagger when $d \geq 3$. The difference can be inferred from our discussion in the Introduction and the following proposition.

Proposition 6. *For any $d \geq 2$ and $\mathbf{X} \in (L^0)^d$, we have*

$$\mathbf{X} \leq_{\text{sm}}^\diamond \mathbf{X}^{\text{co}}, \tag{16}$$

where \mathbf{X}^{co} is a comonotonic version of \mathbf{X} . For $d \geq 3$,

$$\mathbf{X} \leq_{\text{sm}}^\dagger \mathbf{X}^{\text{co}} \text{ and } \mathbf{X} \leq_{\text{sm}} \mathbf{X}^{\text{co}} \text{ fail to hold for some } \mathbf{X} \in (L^\infty)^d.$$

Proof. The first part of this result is due to Lorentz (1953). A slightly more general version can also be found in Tchen (1980, Theorem 5) along with a shorter proof. Example 1 shows that (16) cannot be extended to \leq_{sm} or \leq_{sm}^\dagger on $(L^\infty)^d$. Note that the function φ therein is unbounded, but it produces finite expectations for the two random variables $\varphi(\mathbf{X})$ and $\varphi(\mathbf{X}^{\text{co}})$ in the example. Therefore, it is an element of Φ_{sm}^d required to establish comparisons according to \leq_{sm} and \leq_{sm}^\dagger , but not $\leq_{\text{sm}}^\diamond$. \square

If one reduces the domain of the supermodular order via the restriction (8), then the issue raised by Example 1 does not occur, as the function φ in (3) does not satisfy (8) for any $b_{\mathfrak{F}}$ with $\mathbb{E}[b_{\mathfrak{F}}(X, Y, Z)] < \infty$, and hence such (X, Y, Z) is excluded.

Proposition 6 indicates that, in general, $\leq_{\text{sm}}^\diamond \not\Rightarrow \leq_{\text{sm}}$ and $\leq_{\text{sm}}^\diamond \not\Rightarrow \leq_{\text{sm}}^\dagger$ when $d \geq 3$, even

on $(L^\infty)^d$. In other words, \leq_{sm} and \leq_{sm}^\dagger cannot be generated solely by right-continuous bounded functions. By Theorem 2.4.2(g) of Müller and Stoyan (2002), this means that the supermodular order and the \dagger supermodular order are not closed with respect to weak convergence.

In contrast to Proposition 6, counter-monotonic random vectors in dimensions $d \geq 3$, when they exist, are indeed the smallest elements within their Fréchet class with respect to \leq_{sm} (hence also \leq_{sm}^\dagger and $\leq_{\text{sm}}^\diamond$).

Theorem 7. *Let $d \geq 3$. For any $\mathbf{X} \in (L^0)^d$, with at least 3 non-degenerate components, we have*

$$\mathbf{X}^{\text{ct}} \leq_{\text{sm}} \mathbf{X},$$

if a counter-monotonic version \mathbf{X}^{ct} of \mathbf{X} exists.

It may appear surprising that counter-monotonicity is a lower bound with respect to \leq_{sm} , while comonotonicity fails to produce an upper bound. The asymmetry comes from the specific requirements on a random vector's marginal distributions for a counter-monotonic version of it to exist. For instance, a vector comprising 3 standard uniform random variables, as in Example 1, do not admit a counter-monotonic version; see Dall'Aglio (1972). Note that $\mathbf{X}^{\text{ct}} \leq_{\text{sm}}^\diamond \mathbf{X}$ holds true without requiring \mathbf{X} to have at least 3 non-degenerate components, and this follows by combining Theorem 7 and Proposition 5. Whether $\mathbf{X}^{\text{ct}} \leq_{\text{sm}} \mathbf{X}$ or $\mathbf{X}^{\text{ct}} \leq_{\text{sm}}^\dagger \mathbf{X}$ holds without this condition is unclear to us, and it boils down to whether the relations in Proposition 5 hold for \leq_{sm} and \leq_{sm}^\dagger .

6 Implications for risk management and decision making

As we stated in the Introduction, the convex order is a prominent tool for risk management and decision making. We hereby discuss possible insights and some limitations of our main results in these contexts.

6.1 Choosing between two risks with possibly infinite mean

Suppose that a decision-maker is faced with having to choose between two infinite-mean risks. Such a situation is not unrealistic: despite the financial world containing a finite amount of money, infinite-mean distributions are relevant to multiple risk-modeling situations, catastrophic risks and operational risks for example; see Chen and Wang (2025) for a review. If the decision-maker is risk averse, its preferences for one of the two risks should be consistent with any convex ordering of the two. This is how strong risk aversion is defined for finite-mean risks, following the classical paper by Rothschild and Stiglitz (1970). Convex order thus discerns risk in some

sorts; more precisely, all risk-averse agents agree that, of two risks where one is dominated by the other, the latter is the riskier. A risk measure that is increasing with respect to \leq_{cx} is called \leq_{cx} -consistent. Some examples are the Expected Shortfall (ES) at level $p \in (0, 1)$,

$$\text{ES}_p(X) = \frac{1}{1-p} \int_p^1 F_X^{-1}(u) \, du, \quad X \in L^0,$$

whose \leq_{cx} -consistency can be checked by Theorem 1, and the entropic risk measure (ER) with parameter $\beta > 0$,

$$\text{ER}_\beta(X) = \frac{1}{\beta} \log \mathbb{E}[e^{\beta X}], \quad X \in L^0,$$

whose \leq_{cx} -consistency can be checked by definition. Both ES and ER are well-defined on all of L^0 , although they can take infinite values.

A characterization of all monetary risk measures on L^1 satisfying \leq_{cx} -consistency is obtained in Mao and Wang (2020). Many \leq_{cx} -consistent risk measures are also convex, and all convex risk measures on L^p for $p \geq 1$ satisfying the Fatou property are \leq_{cx} -consistent. Indeed, Filipović and Svindland (2012) showed that the canonical space for convex risk measures is L^1 . General convex risk measures on L^0 are delicate objects to study, as extending from L^1 to L^0 may lead to undefinedness, and there is no natural notion of continuity on L^0 to help establish \leq_{cx} -consistency. Below, we consider a popular class of risk measures, the class of distortion risk measures (e.g., Wang et al. (1997), McNeil et al. (2015)), which we define on L^0 . They are defined as

$$\rho(X) = \int_0^\infty h(\mathbb{P}(X > x)) \, dx + \int_{-\infty}^0 (h(\mathbb{P}(X > x)) - 1) \, dx, \quad X \in L^0, \quad (17)$$

where $h : [0, 1] \rightarrow [0, 1]$ is an increasing function satisfying $h(0) = 0$ and $h(1) = 1$ such that (17) is well-defined for all $X \in L^0$. The function h is called the distortion function of ρ . It is well known that on L^∞ , ρ is convex if and only if h is concave, which is again equivalent to \leq_{cx} -consistency; see Wang et al. (2020b, Theorem 3). For ρ to be well-defined on L^0 , some condition on h needs to be imposed; see Wang et al. (2020a) for some examples.

Theorem 8. *For a distortion risk measure $\rho : L^0 \rightarrow [-\infty, \infty]$ with distortion function h , the following are equivalent:*

- (i) ρ is \leq_{cx} -consistent on L^0 ;
- (ii) ρ is \leq_{cx} -consistent on L^∞ ;
- (iii) ρ is convex on L^0 ;
- (iv) ρ is convex on L^∞ ;

(v) h is concave.

In the usual framework where the domain of \leq_{cx} is limited to L^1 , one would have to exclude infinite-mean risks from comparison, even if ES, ER, and some other distortion risk measures are well-defined for such risks. In Sections 2 and 3, we discussed that there is no need to restrain the set of loss functions or the set of compared risks for the framework given by \leq_{cx} to make sense. The convex order \leq_{cx} remains well-behaved and, although it no longer satisfies antisymmetry, it continues to satisfy its principal properties and is well-suited to compare risks in $L^0 \setminus L^1$. Let us note that risks having infinite means do not necessarily imply infinite values of \leq_{cx} -consistent risk measures, as we will see in Example 5 below.

One of our main results, Theorem 4, pertains to risk aggregation, stipulating that comonotonicity and counter-monotonicity yield bounds on riskiness, as discerned by \leq_{cx} , of the aggregate risk, and thus

$$\rho(X^{\text{ct}} + Y^{\text{ct}}) \leq \rho(X + Y) \leq \rho(X^{\text{co}} + Y^{\text{co}}), \quad \text{for all } \leq_{\text{cx}}\text{-consistent } \rho, \quad (18)$$

assuming that these are well-defined. Theorem 8 allows us to apply (18) to convex distortion risk measures on L^0 . The relation (18) was previously inaccessible for risks outside L^1 . The following example illustrates.

Example 5. Let X and Y be identically distributed random variables following a negated Pareto distribution with shape parameter 0.8 and scale parameter 1, written as $-X, -Y \sim \text{Pareto}(0.8)$, that is,

$$\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x) = (1 - x)^{-0.8}, \quad x \leq 0.$$

Note that $X, Y \in L^0 \setminus L^1$. From (18) and the \leq_{cx} -consistency of ES_p and ER_β , we can immediately obtain chains of inequalities for ES_p and ER_β . Numerical calculations confirm this for $p \in \{0.75, 0.9\}$ and $\beta \in \{0.1, 0.5\}$, as shown in Table 1. None of these values are infinite. Although the patterns in Table 1 are intuitive, especially for ES_p since its right-tailedness naturally eliminates the negative-infinite mean, previous formulations of (12) would not have allowed for such deductions since they exclude X and Y from comparison.

Even if risk measurements end up being infinite under some dependence schemes, they remain ordered according to (18); we provide two simple examples in Table 2. From Theorems 5 and 6, similar conclusions to the ones above translate to portfolios of multiple risks.

On the contrary, in some settings of decision making, risks may not aggregate linearly and each matters, by its own nature (e.g., health and wealth), giving rise to cross-preferences (see, e.g., Eeckhoudt et al. (2007)). In such settings, a decision-maker whose multivariate loss function

	ES _{0.75}	ES _{0.90}	ER _{0.1}	ER _{0.5}
$X^{\text{ct}} + Y^{\text{ct}}$	-3.223	-1.278	-7.07	-4.56
$X^{\perp} + Y^{\perp}$	-0.987	-0.497	-6.02	-2.98
$X^{\text{co}} + Y^{\text{co}}$	-0.386	-0.135	-4.52	-2.12

Table 1: Risk measures ES_p , $p \in \{0.75, 0.90\}$, and ER_β , $\beta \in \{0.1, 0.5\}$, of the sum of two identically-distributed negated-Pareto risks X, Y such that $-X, -Y \sim \text{Pareto}(0.8)$ with various dependence relations: $(X^{\text{ct}}, Y^{\text{ct}})$ is a counter-monotonic version of (X, Y) , (X^{\perp}, Y^{\perp}) is an independent version, and $(X^{\text{co}}, Y^{\text{co}})$ is a comonotonic version.

	Cauchy		Pareto(0.8)	
	ES _p	ER _β	ES _p	ER _β
$X^{\text{ct}} + Y^{\text{ct}}$	0	0	∞	∞
$X^{\perp} + Y^{\perp}$	∞	∞	∞	∞
$X^{\text{co}} + Y^{\text{co}}$	∞	∞	∞	∞

Table 2: Convex-order-consistent measurements of risk, although possibly infinite, are indeed ordered according to Theorem 3. Here, the values $p \in (0, 1)$ and $\beta > 0$ are irrelevant.

is supermodular and unbounded might prefer a portfolio of comonotonic risks over some other dependence relations, as we have shown in Example 1.

6.2 Diversification for infinite-mean risks

Recently, the literature on risk aggregation has seen a lot of interest been addressed to diversification for risks with infinite means (Chen et al., 2025a,b,c; Chen and Schneer, 2025; Arab et al., 2025; Müller, 2025; Vincent, 2025). All of these papers discuss stochastic dominance of the form

$$(\theta_1 + \theta_2)X \leq_{\text{st}} \theta_1 X + \theta_2 Y \quad \text{for } \theta_1, \theta_2 > 0, \quad (19)$$

where X, Y both follow some infinite-mean distribution, and generalizations to n random variables. In particular, Chen et al. (2025b) established that (19) holds for weakly negatively-associated identically-distributed infinite-mean Pareto risks; Müller (2025), for iid super-Cauchy risks; and Arab et al. (2025), for iid InvSub risks.

A risk-management interpretation of (19) is that diversification of the risk portfolio is harmful for these infinite-mean risks. Note that the left-hand side of (19) may be rewritten as $\theta_1 X^{\text{co}} + \theta_2 Y^{\text{co}}$, meaning comonotonicity is the less risky dependence scheme in these cases (less than negative dependence, even, for Pareto-distributed risks, see Chen et al. (2025b, Theorem 1)). This notion of “less riskiness” is modelled by the stochastic order instead of the convex order.

Such a statement seems to contrast with our discussion from the previous subsection. Lemma 1 part (i) makes clear that Theorem 3 is actually compatible with such orderings of the form (19); we first discuss this through the following example.

Example 6. Suppose X and Y are iid following a Pareto distribution with parameter $\alpha \in (0, 1]$, that is, $\mathbb{P}(X > x) = x^{-\alpha}$ for $x \in [1, \infty)$. Then $X, Y \in L^0 \setminus L^1$. For any $\theta_1, \theta_2 > 0$, Theorem 3 implies

$$\theta_1 X + \theta_2 Y \leq_{\text{cx}} (\theta_1 X)^{\text{co}} + (\theta_2 Y)^{\text{co}} \stackrel{\text{d}}{=} (\theta_1 + \theta_2) X,$$

which, by Lemma 1 part (i), amounts to

$$\theta_1 X + \theta_2 Y \leq_{\text{dcx}} (\theta_1 + \theta_2) X. \tag{20}$$

Meanwhile, Theorem 1 of Chen et al. (2025b) states that (19) holds, which is clearly stronger than (20). While (19) holds for some (but not all) other classes of distributions without finite mean, the relation (20) holds for all distributions of X with infinite $\mathbb{E}[X_+]$.

As we explained in Section 3, the convex order can loosely be interpreted as indicating deviation from the mean. This interpretation brings clarity as to why a reverse ordering of riskiness with respect to stochastic dominance, as in (19), may hold simultaneously to (12): a distribution deviating more from the infinite mean may end up stochastically smaller. Thus follows a shift in our perception of risk: since the mean is infinite, we now *want* to deviate from it as much as possible; deviation from the mean is no longer seen as risky. The conclusions of the previously cited papers, that diversification may produce more risk, actually stem from this shift in perception. Random variables in L^0 still behave the same when summed, and comonotonicity still produces the most deviation from the mean.

In general, \leq_{st} is seen as much stronger in terms of comparing risks and perhaps most will agree that risk preferences should follow \leq_{st} , even if they do not follow \leq_{cx} . Therefore, \leq_{cx} -consistent preferences may not adequately capture most's conception of riskiness on L^0 .

6.3 Limitations

The results of this paper were mainly developed in the perspective of better understanding convex order for infinite-mean risks; as such, their practical applications have some limitations.

First, as argued above, the convex order perhaps does not encapsulate adequately what constitutes “risk” for infinite mean random variables. In other words, while our results provide clarifications on the diversification behavior of infinite-mean risk, reconciling the convex ordering (12) with (19) observed for many distributions in $L^0 \setminus L^1$, they lose some part of applicability due to these very clarifications.

Second, one could also argue about the limited practicality of convex orderings when evaluations of risk are infinite. As highlighted in Table 2, ES and ER of the sum of Pareto-distributed

risks are infinite, no matter the dependence relations between these risks. Any expected losses with increasing convex loss function would be infinite under all dependence schemes as well. It may be conceptually irrelevant to ask a decision-maker to choose between risks they inherently value as infinite: if truly they assess the riskiness as infinite under each dependence relation, the decision-maker should be indifferent to it — they are in utter despair in every case. Thus, any possible convex ordering does not matter.

Our intent was mainly to provide mathematical insights and a more in-depth understanding of convex order for risks outside of L^1 . We humbly leave to more philosophically inclined minds the considerations on whether a truly infinite level of risk may exist and, if so, be manageable in any sorts.

7 Proofs of Theorems 1–3 and 6–8

The proofs of Theorems 1 and 2 rely on Lemma 1 in Section 3, which we prove below.

Proof of Lemma 1. Note that for any $u \in \mathcal{U}_{\text{cx}}$, if $u(y) > u(x)$ for some $y > x$, then u dominates a function of the form $z \mapsto az + b$ for some $a > 0$ and $b \in \mathbb{R}$. Under the assumption $\mathbb{E}[Y_+] = \infty$ in item (i), this u produces an infinite or undefined $\mathbb{E}[u(Y)]$ so that $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ holds necessarily whenever it is well defined. Hence, remaining to establish convex ordering is to verify the inequality for decreasing u , that is, $u \in \mathcal{U}_{\text{dcox}}$. To prove item (ii), one puts forth the same reasoning symmetrically. For item (iii), it follows from items (i) and (ii), because $\mathcal{U}_{\text{dcox}} \cap \mathcal{U}_{\text{icx}}$ contains only constant functions. In this case, $Y \leq_{\text{cx}}^{\dagger} X$ also hold because all $u \in \mathcal{U}_{\text{dcox}} \cup \mathcal{U}_{\text{icx}}$ are overlooked in establishing comparison as they produce infinite expectations. \square

Proof of Theorem 1. First, it is straightforward to verify that (ii) and (iii) are equivalent. For instance, when $\mathbb{E}[X_+]$ and $\mathbb{E}[Y_+]$ are finite, the equivalence between the second condition in (10) and that in (11) is known from the proof of Denuit et al. (2005, Proposition 3.4.7) or Shaked and Shanthikumar (2007, Equations (4.A.4)–(4.A.8)). When $\mathbb{E}[X_+]$ or $\mathbb{E}[Y_+]$ is infinite, it is clear that the second condition in (10) and that in (11) are equivalent.

The direction (i) \Rightarrow (iii) follows immediately, because both $x \mapsto (x - w)_+$ and $x \mapsto (x - w)_-$ are convex functions for all $w \in \mathbb{R}$. Next, we show (iii) \Rightarrow (i). Since (10) is equivalent to (9) when $X, Y \in L^1$, we only need to argue the cases when at least one of X, Y is in $L^0 \setminus L^1$. Clearly, if X is in $L^0 \setminus L^1$, so is Y ; we will therefore assume $Y \in L^0 \setminus L^1$. If both $\mathbb{E}[Y_+] = \infty$ and $\mathbb{E}[Y_-] = \infty$, then $X \leq_{\text{cx}} Y$ is always true by Lemma 1, item (iii). Next assume $\mathbb{E}[Y_+] < \infty$ and $\mathbb{E}[Y_-] = \infty$. Note that both expectations in the second inequality of (11) are finite, and it is well-known that

functions of the form $x \mapsto (x - w)_+$ then suffice to generate \leq_{icx} on L^1 , that is, for $Z, W \in L^1$,

$$\mathbb{E}[(Z - w)_+] \leq \mathbb{E}[(W - w)_+] \text{ for all } w \in \mathbb{R} \implies Z \leq_{\text{icx}} W; \quad (21)$$

see, e.g., [Müller and Stoyan \(2002, Theorem 1.5.7\)](#). Note that for any $m, w \in \mathbb{R}$,

$$((X \vee m) - w)_+ = \begin{cases} (X - w)_+ & m \leq w; \\ (X - m)_+ + (m - w) & m > w. \end{cases}$$

Hence, the second condition in (11) implies $\mathbb{E}[(X \vee m - w)_+] \leq \mathbb{E}[(Y \vee m - w)_+]$. Using (21), we get $(X \vee m) \leq_{\text{icx}} (Y \vee m)$ for all $m \in \mathbb{R}$. For any $f \in \mathcal{U}_{\text{icx}}$ with $\mathbb{E}[f(X)]$ and $\mathbb{E}[f(Y)]$ well-defined, the monotone convergence theorem guarantees that $\mathbb{E}[f(X \vee m)] \rightarrow \mathbb{E}[f(X)]$ and $\mathbb{E}[f(Y \vee m)] \rightarrow \mathbb{E}[f(Y)]$ as $m \rightarrow -\infty$. Using $\mathbb{E}[f(X \vee m)] \leq \mathbb{E}[f(Y \vee m)]$, we get $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$. This further gives $X \leq_{\text{cx}} Y$ by using Lemma 1, item (ii). The case where $\mathbb{E}[Y_+] = \infty$ and $\mathbb{E}[Y_-] < \infty$ is covered by a symmetric argument, involving Lemma 1, item (i). \square

Proof of Theorem 2. The equivalence between (ii) and (iii) is straightforward by equivalence of corresponding inequalities in (10) and (11). As for Theorem 1, the direction (i) \Rightarrow (ii) follows by noting that $x \mapsto (x - w)_+$ and $x \mapsto (x - w)_-$ are convex for any $w \in \mathbb{R}$. If $X, Y \in L^1$, then item (ii) boils down to (9), and \leq_{cx} and \leq_{cx}^\dagger are equivalent by Proposition 1.

We argue (iii) \Rightarrow (i) for cases when at least one of X, Y has an infinite mean. Suppose that $\mathbb{E}[X_-]$ is infinite. Then all $u \in \mathcal{U}_{\text{dcx}}$, other than the constant functions, will produce an infinite $\mathbb{E}[u(X)]$. Hence, all non-constant $u \in \mathcal{U}_{\text{dcx}}$ are excluded when checking $X \leq_{\text{cx}}^\dagger Y$ according to the definition of \leq_{cx}^\dagger . We therefore only need to assess $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for $u \in \mathcal{U}_{\text{icx}}$; the second inequality in (11) suffices as discussed in the proof of Theorem 1. The same argument also applies if rather $\mathbb{E}[Y_-]$ is infinite, and it is symmetric if $\mathbb{E}[X_+] = \infty$ or $\mathbb{E}[Y_+] = \infty$. If $\mathbb{E}[X]$ is not well-defined, then neither inequality in (11) needs to be checked, and $X \leq_{\text{cx}}^\dagger Y$ by Lemma 1, item (iii). \square

The proof of Theorem 3 relies on extending the desired relations (12) from L^∞ , as stated in the following lemma, to L^0 .

Lemma 2 (Corollary 3.28 of [Rüschendorf \(2013\)](#), restated on L^∞). *For $X, Y \in L^\infty$, (12) holds.*

Proof of Theorem 3. The main idea of the proof is that appropriately truncating the random variables allows to apply Lemma 2 on L^∞ , and then to exploit monotone convergence to extend it to the original random variables. We present the developments for the lower bound in (12).

By changing “ct” to “co” and flipping the inequalities, one will immediately obtain the proof for the upper bound.

For $m > 0$ and a random variable Z , let $Z_{(m)} = \max\{\min\{Z, m\}, -m\}$, that is, Z truncated at $\pm m$. The terms $X_{(m)}$, $Y_{(m)}$, $X_{(m)}^{\text{ct}}$ and $Y_{(m)}^{\text{ct}}$ are truncations in the above sense. Note that $X_{(m)}^{\text{ct}}$ and $Y_{(m)}^{\text{ct}}$ are counter-monotonic since truncation is a monotone transform. Moreover, $X_{(m)} \stackrel{\text{d}}{=} X_{(m)}^{\text{ct}}$ and $Y_{(m)} \stackrel{\text{d}}{=} Y_{(m)}^{\text{ct}}$. Therefore, Lemma 2 implies

$$\mathbb{E}[u(X_{(m)}^{\text{ct}} + Y_{(m)}^{\text{ct}})] \leq \mathbb{E}[u(X_{(m)} + Y_{(m)})] \quad \text{for all } u \in \mathcal{U}_{\text{cx}}, \quad (22)$$

as soon as the two expectations are well-defined.

Let $\mathcal{U}_{\text{cx}}^*$ be the set of all $u \in \mathcal{U}_{\text{cx}}$ such that u is monotone (i.e., either increasing or decreasing) on $[0, \infty)$ and on $(-\infty, 0]$. Note that any convex function $u \in \mathcal{U}_{\text{cx}}$ can change monotonicity at most once, say at x_0 . Then, the function $v : x \mapsto u(x - x_0)$ is in $\mathcal{U}_{\text{cx}}^*$. If u does not change monotonicity, then $u \in \mathcal{U}_{\text{cx}}^*$.

Next, we make the simple, but nontrivial, observation that $(X_{(m)} + Y_{(m)})_+$ and $(X_{(m)} + Y_{(m)})_-$ are increasing in m . To see this, take $(X_{(m)} + Y_{(m)})_+$ for instance. The lower truncation at $-m$ does not matter for the sum, and one has

$$(X_{(m)} + Y_{(m)})_+ = (\min\{X, m\} + \min\{Y, m\})_+.$$

This observation was made by [Simons \(1977\)](#).

Denote by $A_m = \{X_{(m)} + Y_{(m)} \geq 0\}$ and A_m^c its complement. Take any $u \in \mathcal{U}_{\text{cx}}^*$ such that $\mathbb{E}[u(X + Y)]$ is well-defined. We have

$$\begin{aligned} \mathbb{E}[u(X_{(m)} + Y_{(m)})] &= \mathbb{E}[u(X_{(m)} + Y_{(m)})\mathbb{1}_{A_m} + u(X_{(m)} + Y_{(m)})\mathbb{1}_{A_m^c}] \\ &= \mathbb{E}[u((X_{(m)} + Y_{(m)})_+)\mathbb{1}_{A_m} + u(-(X_{(m)} + Y_{(m)})_-)\mathbb{1}_{A_m^c}] \\ &= \mathbb{E}[u((X_{(m)} + Y_{(m)})_+) - u(0)\mathbb{1}_{A_m^c} + u(-(X_{(m)} + Y_{(m)})_-) - u(0)\mathbb{1}_{A_m}] \\ &= \mathbb{E}[u((X_{(m)} + Y_{(m)})_+) + u(-(X_{(m)} + Y_{(m)})_-)] - u(0). \end{aligned}$$

Similarly,

$$\mathbb{E}[u(X + Y)] = \mathbb{E}[u((X + Y)_+) + u(-(X + Y)_-)] - u(0).$$

Because $(X_{(m)} + Y_{(m)})_+ \uparrow (X + Y)_+$ and $(X_{(m)} + Y_{(m)})_- \uparrow (X + Y)_-$ as $m \rightarrow \infty$, the monotone convergence theorem yields $\mathbb{E}[u(X_{(m)} + Y_{(m)})] \rightarrow \mathbb{E}[u(X + Y)]$ as $m \rightarrow \infty$. Applying this to $(X^{\text{ct}}, Y^{\text{ct}})$, we get $\mathbb{E}[u(X_{(m)}^{\text{ct}} + Y_{(m)}^{\text{ct}})] \rightarrow \mathbb{E}[u(X^{\text{ct}} + Y^{\text{ct}})]$ as $m \rightarrow \infty$ if $\mathbb{E}[u(X^{\text{ct}} + Y^{\text{ct}})]$ is well-defined. Using (22), we obtain $\mathbb{E}[u(X^{\text{ct}} + Y^{\text{ct}})] \leq \mathbb{E}[u(X + Y)]$.

To summarize, we have shown

$$\mathbb{E}[u(X^{\text{ct}} + Y^{\text{ct}})] \leq \mathbb{E}[u(X + Y)] \quad \text{for all } X, Y \in L^0 \text{ and all } u \in \mathcal{U}_{\text{cx}}^*, \quad (23)$$

as soon as the expectations are well-defined.

For general $u \in \mathcal{U}_{\text{cx}}$ but not in $\mathcal{U}_{\text{cx}}^*$, we can find $x_0 \in \mathbb{R}$ and $v : x \mapsto u(x - x_0)$ such that $v \in \mathcal{U}_{\text{cx}}^*$, which we argued above. For $X, Y \in L^0$ with well-defined expectations $\mathbb{E}[u(X + Y)]$ and $\mathbb{E}[u(X^{\text{ct}} + Y^{\text{ct}})]$, take $Z = Y + x_0$. Note that $Z^{\text{ct}} = Y^{\text{ct}} + x_0$. Using the result (23) applied to (v, X, Z) , we get

$$\begin{aligned} \mathbb{E}[u(X^{\text{ct}} + Y^{\text{ct}})] &= \mathbb{E}[u(X^{\text{ct}} + Z^{\text{ct}} - x_0)] \\ &= \mathbb{E}[v(X^{\text{ct}} + Z^{\text{ct}})] \\ &\leq \mathbb{E}[v(X + Z)] = \mathbb{E}[u(X + Z - x_0)] = \mathbb{E}[u(X + Y)]. \end{aligned}$$

Therefore, we obtain the desired result that (23) holds for all $u \in \mathcal{U}_{\text{cx}}$, and thus $X^{\text{ct}} + Y^{\text{ct}} \leq_{\text{cx}} X + Y$. \square

Note that the proof of Theorem 3 does not generalize to the multivariate setting, so Theorems 5 and 6 require separate proofs. The proof of Theorem 5 was presented in Section 4.

Proof of Theorem 6. The case $d = 1$ is trivial and $d = 2$ is covered by Theorem 3. The cases where at most 2 random variables are non-degenerate are then also covered because constant shifts do not affect the comparison $\mathbb{E}[u(\sum_{i=1}^d X_i^{\text{ct}})] \leq \mathbb{E}[u(\sum_{i=1}^d X_i)]$, $u \in \mathcal{U}_{\text{cx}}$. For all other cases where $d \geq 3$ and a counter-monotonic version of (X_1, \dots, X_n) exists, the result ensues directly from Theorem 7, whose proof follows below, and the supermodularity of functions of the form $(x, y) \mapsto u(x + y)$, $u \in \mathcal{U}_{\text{cx}}$. \square

Proof of Theorem 7. By Theorem 3 of Dall'Aglio (1972), for $d \geq 3$, a d -dimensional random vector with at least 3 non-degenerate components has a counter-monotonic version if and only if

$$\text{either } \sum_{i=1}^d \mathbb{P}(X_i > \text{ess-inf} X_i) \leq 1, \quad (24)$$

$$\text{or } \sum_{i=1}^d \mathbb{P}(X_i < \text{ess-sup} X_i) \leq 1. \quad (25)$$

We assume (24), and the other case (25) is symmetric and omitted. Note that (24) implies that $\text{ess-inf} X_i$ is finite for each $i \in [d]$, and without loss of generality we can assume $\text{ess-inf} X_i = 0$ for $i \in [d]$ because constant shifts do not affect the comparison $\mathbb{E}[\varphi(\mathbf{X}^{\text{ct}})] \leq \mathbb{E}[\varphi(\mathbf{X})]$, $\varphi \in \Phi_{\text{sm}}^d$.

Take $\varphi \in \Phi_{\text{sm}}^d$ such that both $\mathbb{E}[\varphi(\mathbf{X}^{\text{ct}})]$ and $\mathbb{E}[\varphi(\mathbf{X}^{\text{ct}})]$ are well-defined, and without loss of generality assume $\varphi(0, \dots, 0) = 0$. It is straightforward to verify that the supermodular function φ with $\varphi(0, \dots, 0) = 0$ satisfies

$$\varphi(x_1, \dots, x_d) \geq \sum_{i=1}^d \varphi(x_i \mathbf{e}_i),$$

where $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 at the i th position. Note that $\varphi(X_i \mathbf{e}_i)$ is identically distributed as $\varphi(X_i^{\text{ct}} \mathbf{e}_i)$ for each $i \in [d]$. Therefore, $\mathbb{E}[\varphi(X_i \mathbf{e}_i)] = \mathbb{E}[\varphi(X_i^{\text{ct}} \mathbf{e}_i)]$ as long as they are well-defined. Moreover,

$$\sum_{i=1}^d \varphi(X_i^{\text{ct}} \mathbf{e}_i) = \varphi(\mathbf{X}^{\text{ct}}) \quad \text{almost surely,}$$

because \mathbf{X}^{ct} is counter-monotonic; see [Dall'Aglio \(1972, Lemma 2\)](#) or [Puccetti and Wang \(2015, Proposition 3.2\)](#). Using this, for $\mathbb{E}[\varphi(\mathbf{X}^{\text{ct}})]$ to be well-defined, it holds that $\mathbb{E}[\varphi(X_i^{\text{ct}} \mathbf{e}_i)] < \infty$ for all $i \in [d]$ or $\mathbb{E}[\varphi(X_i^{\text{ct}} \mathbf{e}_i)] > -\infty$ for all $i \in [d]$. Therefore, we get

$$\mathbb{E}[\varphi(\mathbf{X})] \geq \sum_{i=1}^d \mathbb{E}[\varphi(X_i \mathbf{e}_i)] = \sum_{i=1}^d \mathbb{E}[\varphi(X_i^{\text{ct}} \mathbf{e}_i)] = \mathbb{E} \left[\sum_{i=1}^d \varphi(X_i^{\text{ct}} \mathbf{e}_i) \right] = \mathbb{E}[\varphi(\mathbf{X}^{\text{ct}})],$$

where we used the verified fact that all expectations are well-defined. \square

Proof of Theorem 8. The chains (i) \Rightarrow (ii) \Rightarrow (v) and (iii) \Rightarrow (iv) \Rightarrow (v) follow from the definition and [Wang et al. \(2020b, Theorem 3\)](#). It remains to show (v) \Rightarrow (i) and (v) \Rightarrow (iii). Take a concave h and a corresponding distortion risk measure ρ in (17). Since h is increasing and concave, it is left-continuous. Using Lemma 1 of [Wang et al. \(2020a\)](#), we have

$$\rho(X) = \int_0^1 F_X^{-1}(1-t) dh(t), \quad X \in L^0.$$

Let $g(t) = 1 - h(1-t)$ for $t \in [0, 1]$, with its right derivative denoted by g' ; since g is convex, g' is increasing. The right derivative of g' is given by g'' , which exists almost everywhere and it is positive. Further, let $F_X^{-1}(1) = \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) = 1\}$, and $\delta = g(1) - \lim_{t \uparrow 1} g(t)$. It follows

that

$$\begin{aligned}
\rho(X) &= \int_0^1 F_X^{-1}(t) dg(t) \\
&= \int_0^1 F_X^{-1}(t) g'(t) dt + \delta F_X^{-1}(1) \\
&= \int_0^1 \int_0^t F_X^{-1}(t) g''(s) ds dt + \delta F_X^{-1}(1) \\
&= \int_0^1 g''(s) \int_s^1 F_X^{-1}(t) dt ds + \delta F_X^{-1}(1).
\end{aligned}$$

Since $X \mapsto \int_s^1 F_X^{-1}(t) dt$ is \leq_{cx} -consistent on L^0 for $s \in (0, 1)$ by using Theorem 1-(ii) and $X \mapsto F_X^{-1}(1)$ is \leq_{cx} -consistent on L^0 by using Theorem 1-(iii), we conclude that ρ is \leq_{cx} -consistent on L^0 , and thus (i) holds. Similarly, both $X \mapsto \int_s^1 F_X^{-1}(t) dt$ and $X \mapsto F_X^{-1}(1)$ are convex due to the convexity of ES on L^0 , and hence (iii) holds. \square

Acknowledgments

The authors thank an editor, two anonymous referees, Alfred Müller, and Giovanni Puccetti for helpful comments on an early version of the paper. Ruodu Wang is supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823, RGPAS-2018-522590) and Canada Research Chairs (CRC-2022-00141).

References

- Arab, I., Lando, T., and Oliveira, P. E. (2025). Convex combinations of random variables stochastically dominate the parent for a new class of heavy tailed distributions. *Electronic Communications in Probability*, **30**, 1–11.
- Bernard, C., Jiang, X., and Wang, R. (2014). Risk aggregation with dependence uncertainty. *Insurance: Mathematics and Economics*, **54**, 93–108.
- Beiglböck, M., Henry-Labordère, P., and Penkner, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, **17**(3), 477–501.
- Chen, Y., Embrechts, P., and Wang, R. (2025a). Risk exchange under infinite-mean Pareto models. *Insurance: Mathematics and Economics*, **124**, 103131.
- Chen, Y., Embrechts, P. and Wang, R. (2025b). Technical note—an unexpected stochastic dominance: Pareto distributions, dependence, and diversification. *Operations Research*, **73**(3), 1336–1344.

- Chen, Y., Hu, T., Shneer, S., and Zou, Z. (2025c). Stochastic dominance for linear combinations of infinite-mean risks. *arXiv:2505.01739*.
- Chen, Y. and Shneer, S. (2025). Risk aggregation and stochastic dominance for a class of heavy-tailed distributions. *ASTIN Bulletin*, forthcoming.
- Chen, Y. and Wang, R. (2025). Infinite-mean models in risk management: Discussions and recent advances. *Risk Sciences*, 100003.
- Cheung, K. C., Dhaene, J., Lo, A. and Tang, Q. (2014). Reducing risk by merging counter-monotonic risks. *Insurance: Mathematics and Economics*, **54**, 58–65.
- Cheung, K. C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance: Mathematics and Economics*, **55**, 180–190.
- Dall’Aglío, G. (1972). Fréchet classes and compatibility of distribution functions. *Symposia Mathematica*, **9**, 131–150.
- Denuit, M. and Dhaene, J. (2012). Convex order and comonotonic conditional mean risk sharing. *Insurance: Mathematics and Economics*, **51**(2), 265–270.
- Denuit, M., Dhaene, J., Goovaerts and M.J., Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Wiley, Chichester.
- Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R. and Vynche, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, **31**(1), 3–33.
- Durrett, R. (2019). *Probability: Theory and Examples*. Cambridge University Press. Fifth Edition.
- Eeckhoudt, L., Rey, B., and Schlesinger, H. (2007). A good sign for multivariate risk taking. *Management Science*, **53**(1), 117–124.
- Elton, J. and Hill, T. P. (1992). Fusions of a probability distribution. *The Annals of Probability*, **20**(1), 421–454.
- Filipović, D. and Svindland, G. (2012). The canonical model space for law-invariant convex risk measures is L^1 . *Mathematical Finance*, **22**(3), 585–589.
- Föllmer, H. and Schied, A. (2016). *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter. Fourth Edition.
- Hirsch, F., Profeta, C., Roynette, B. and Yor, M. (2011). *Peacocks and Associated Martingales, with Explicit Constructions*. Springer Science and Business Media.
- Kaas, R., Dhaene, J. and Goovaerts, M. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, **27**(2), 151–168.
- Lorentz, G. G. (1953). An inequality for rearrangements. *American Mathematical Monthly*,

- 60**(3), 176–179.
- Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.
- McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Revised Edition. Princeton, NJ: Princeton University Press.
- Müller, A. (1997a). Stochastic orders generated by integrals: a unified study. *Advances in Applied Probability*, **29**(2), 414–428.
- Müller, A. (1997b). Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics*, **21**(3), 219–223.
- Müller, A. (2025). Some remarks on the effect of risk sharing and diversification for infinite mean risks. *ASTIN Bulletin*, forthcoming.
- Müller, A. and Scarsini, M. (2000). Some remarks on the supermodular order. *Journal of Multivariate Analysis*, **73**(1), 107–119.
- Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. Wiley.
- Puccetti, G., Rigo, P., Wang, B. and Wang, R. (2019). Centers of probability measures without the mean. *Journal of Theoretical Probability*, **32**(3), 1482–1501.
- Puccetti, G. and Wang R. (2015). Extremal dependence concepts. *Statistical Science*, **30**(4), 485–517.
- Rachev, S. T. and Rüschendorf, L. (2006). *Mass Transportation Problems, Volume 1: Theory*. Springer, New York.
- Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.
- Rüschendorf, L. (2013). *Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios*. Springer, Heidelberg.
- Simons, G. (1977). An unexpected expectation. *Annals of Probability*, **5**(1), 157–158.
- Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer, New York.
- Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, **36**(2), 423–439.
- Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probability*, **8**(4), 814–827.
- Torgersen, E. N. (1991). *Comparison of Statistical Experiments*. Cambridge University Press, Cambridge.
- Villani, C. (2009). *Optimal Transport: Old and New*. Springer, Heidelberg.
- Vincent, L. (2025). Diversification and stochastic dominance: When all eggs are better put in

one basket. *arXiv:2507.16265*.

Wang, Q., Wang, R. and Wei, Y. (2020a). Distortion riskmetrics on general spaces. *ASTIN Bulletin*, **50**(4), 827–851.

Wang, R., Wei, Y. and Willmot, G. E. (2020b). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.

Wang, S., Young, V. R. and Panjer, H. H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, **21**(2), 173–183.

Whitt, W. (1986). Stochastic comparisons for non-Markov processes. *Mathematics of Operations Research*, **11**(4), 608–618.