An unexpected stochastic dominance: Pareto distributions, dependence, and diversification

Yuyu Chen
Department of Economics, University of Melbourne, Australia, yuyu.chen@unimelb.edu.au

Paul Embrechts
RiskLab, Department of Mathematics & ETH Risk Center, ETH Zurich, Switzerland, embrechts@math.ethz.ch

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

We find the perhaps surprising inequality that the weighted average of independent and identically distributed Pareto random variables with infinite mean is larger than one such random variable in the sense of first-order stochastic dominance. This result holds for more general models including super-Pareto distributions, negative dependence, and triggering events, and yields superadditivity of the risk measure Value-at-Risk for these models.

Key words: Pareto distributions; diversification effect; risk pooling; first-order stochastic dominance

1. Introduction

Pareto distributions are arguably the most important class of heavy-tailed loss distributions, due to their connection to regularly varying tails, Extreme Value Theory (EVT), and power laws in economics and social networks; see, e.g., Embrechts et al. (1997), de Haan and Ferreira (2006) and Gabaix (2009). In quantitative risk management, Pareto distributions are frequently used to model losses from catastrophes such as earthquakes, hurricanes, and wildfires; see, e.g., Embrechts et al. (1999). They are also widely used in economics for wealth distributions (e.g., Taleb (2020)) and modeling the tails of financial asset losses and operational risks (e.g., McNeil et al. (2015)); applications of power laws in economics, finance, and insurance are treated in Ibragimov et al. (2015) and Ibragimov and Prokhorov (2017). Andriani and McKelvey (2007) listed over 80 examples of power laws in diverse fields of applications. By the Pickands-Balkema-de Haan Theorem (Pickands (1975) and Balkema and de Haan (1974)), generalized Pareto distributions are the only possible non-degenerate limiting distributions of the residual lifetime of random variables exceeding a high level.

In the realm of banking and insurance, distributions with infinite mean occur as a possible mathematical model in several contexts. For instance, catastrophic losses, operational losses, large insurance losses, and financial returns from technological innovations are often modelled by Pareto distributions without finite mean; see Embrechts et al. (1997) in the context of extreme value theory, Hofert and Wüthrich (2012) on nuclear power accidents, and the more recent Cheynel et al. (2022) on modeling fraud. In risk management, such infinite-mean models often lead to intriguing phenomena, such as the diversification disaster studied by Ibragimov et al. (2009, 2011).

Stochastic dominance relations are an important tool in economic decision theory, which allows for the analysis of risk preferences for a group of decision makers (Hadar and Russell (1969)). They have been studied in the forms of first and second degrees (Quirk and Saposnik (1962), Hadar and Russell (1969, 1971) and Rothschild and Stiglitz (1970)), larger integer degrees (Whitmore (1970) and Caballé and Pomansky (1996)), and fractional degrees (Müller et al. (2017) and Huang et al. (2020)), and they are widely applied in the expected utility and dual utility theory (Yaari (1987)), behavioural decision models (Chew et al. (1987), Baucells and Heukamp (2006) and Schmidt and Zank (2008)), and risk measures (Föllmer and Schied (2016)). See also Levy (1992, 2016) for the wide applicability of stochastic dominance relations in decision making, and Müller and Stoyan (2002) and Shaked and Shanthikumar (2007) for the mathematics of stochastic dominance.

The strongest form of commonly used stochastic dominance relations is first-order stochastic dominance. For two random variables X and Y representing random losses, we say X is smaller than Y in *first-order stochastic dominance*, denoted by $X \le_{\mathrm{st}} Y$, if $\mathbb{P}(X \le x) \ge \mathbb{P}(Y \le x)$ for all $x \in \mathbb{R}$. The relation $X \le_{\mathrm{st}} Y$ means that all decision makers with an increasing utility function will prefer loss X to loss Y if their expected utilities are finite. In this paper, all terms like "increasing" and "decreasing" are in the non-strict sense.

For iid random variables $X_1, \dots, X_n$ following a Pareto distribution with infinite mean and weights $\theta_1, \dots, \theta_n \ge 0$ with $\sum_{i=1}^n \theta_i = 1$, one consequence of our main result, Theorem 1, is the stochastic dominance relation

$$X_1 \le_{\mathrm{st}} \theta_1 X_1 + \cdots + \theta_n X_n, \tag{1}$$

and (1) is not an equality except for the trivial case that only one of $\theta_1, \dots, \theta_n$ is non-zero. As far as we are aware, (1) is not known in the literature, even in the case that $\theta_1, \dots, \theta_n$ are equal (i.e., they are $1/n$).

To appreciate the nature of (1), we first recall that for any identically distributed random variables $X_1, \dots, X_n$ with finite mean, regardless of their distribution or dependence structure, for $\theta_1, \dots, \theta_n > 0$ with $\sum_{i=1}^n \theta_n = 1$, (1) can only hold if $X_1 = \cdots = X_n$ (almost surely), in which case we have the trivial equality $X_1 = \theta_1 X_1 + \cdots + \theta_n X_n$; see Proposition 2. Therefore, the assumption of infinite mean is very important for (1) to hold.

It is somewhat surprising that, for infinite-mean Pareto losses, inequality (1) holds for a very strong form of risk comparison: For every decision maker with a risk preference favouring less loss over more loss and *well defined* for losses in (1), a diversified portfolio of such iid Pareto losses is less preferred to a non-diversified one. Flipping the sign, diversification is preferred if the Pareto random variables are treated as profits or gains from, for instance, research and development. We call such a stochastic dominance "unexpected" for both its surprising nature and the infinite expectations involved.

The infinite mean assumption comes with a caveat: for a risk-averse expected utility decision maker, due to the concavity of the utility function, (1) does not imply a preference for non-diversification, because losses in both sides of (1) have $-\infty$ expected utility. It does, however, give *superadditivity* of the regulatory risk measure Value-at-Risk (VaR) in banking and insurance sectors; that is, the weighted average of super-Pareto losses gives a larger VaR than the weighted average of VaRs of individual super-Pareto losses. Different from the literature on VaR superadditivity for regularly varying distributions (e.g., Embrechts et al. (2009) and McNeil et al. (2015)), the superadditivity of VaR implied by (1) holds for all probability levels, and this is not in an asymptotic sense.

Observations similar to (1), although with less generality, occur in the literature in different forms. Samuelson (1967) mentioned that having an infinite mean in portfolio diversification may lead to a worse distribution; see also Fama and Miller (1972, p. 271) and Malinvaud (1972). Inequality (1) for $n=2$ and the Pareto tail parameter $\alpha=1/2$ (see Section 2 for the parametrization) has an explicit formula in Example 7 of Embrechts et al. (2002). Simple numerical examples are provided by Embrechts and Puccetti (2010, Figure 5.2) and Bauer and Zanjani (2016, Table 2). Ibragimov (2005) showed that (1) holds for iid positive one-sided stable random variables with infinite mean. Another relevant result of Ibragimov (2009) is that for iid random variables $Z_1,\dots,Z_n$ which follow a convolution of symmetric stable distributions without finite mean, $\mathbb{P}(\theta_1 Z_1+\cdots+\theta_n Z_n \le x) \le \mathbb{P}(Z_1 \le x)$ for $x>0$ but the opposite holds for $x<0$ (and hence first-order stochastic dominance does not hold). The symmetry of distributions is essential for this inequality, and $Z_1,\dots,Z_n$ can take negative values, unlike Pareto losses, which are positive, skewed and more suitable for the modeling of extreme losses.

In Section 2, we first introduce super-Pareto distributions, a class of infinite-mean distributions, and weak negative association, a notion of dependence weaker than negative association (Alam and Saxena (1981) and Joag-Dev and Proschan (1983)). The class of super-Pareto distributions includes all infinite-mean Pareto distributions. Our main result, Theorem 1, shows that (1) holds if $X_1,\dots,X_n$ are weakly negatively associated super-Pareto random variables, and also in case they are triggered by some events. Some discussions on this result are provided after its proof. Section 3 concludes.

We fix some notation. Throughout, random variables are defined on an atomless probability space $(\Omega,\mathcal{F},\mathbb{P})$. Denote by $\mathbb{N}$ the set of all positive integers and $\mathbb{R}_+$ the set of non-negative real numbers. For $n\in\mathbb{N}$, let $[n]=\{1,\dots,n\}$. Denote by Δ_n the standard simplex, that is, $\Delta_n=\{(\theta_1,\dots,\theta_n)\in[0,1]^n:\sum_{i=1}^n\theta_i=1\}$. For $x,y\in\mathbb{R}$, write $x\wedge y=\min\{x,y\}$, $x\vee y=\max\{x,y\}$, and $x_+=\max\{x,0\}$. We write $\mathbf{X}\overset{\mathrm{d}}{=}\mathbf{Y}$ if $\mathbf{X}$ and $\mathbf{Y}$ have the same distribution. We always assume $n\ge 2$. Equalities and inequalities are interpreted component-wise when applied to vectors. For any random variable X, its essential infimum is given by $z_X=\inf\{z\in\mathbb{R}:\mathbb{P}(X>z)>0\}$.

2. Stochastic dominance for super-Pareto risks

2.1. Super-Pareto distributions and weak negative association

We first introduce the Pareto distribution and the super-Pareto distribution. A common parameterization of Pareto distributions is given by, for $\theta,\alpha>0$, the cumulative distribution function

$$P_{\alpha,\theta}(x)=1-\left(\frac{\theta}{x}\right)^{\alpha},\quad x\ge\theta.$$

As θ is a scale parameter, it suffices to study $P_{\alpha,1}$, which we write as Pareto(α). The mean of $P_{\alpha,\theta}$ is infinite if and only if the tail parameter α is in $(0,1]$. We say that the $P_{\alpha,\theta}$ distribution is *extremely heavy-tailed* if $\alpha\le 1$.

Definition 1. A random variable X is *super-Pareto* (or has a super-Pareto distribution) if $X\overset{\mathrm{d}}{=}f(Y)$ for some increasing, convex, and non-constant function f and $Y\sim$ Pareto(1). Moreover, X is *regular* if $f(0)=0$ and $f(1)>0$.

All extremely heavy-tailed Pareto distributions are super-Pareto and regular. By definition, the super-Pareto property is preserved under increasing, convex, and non-constant transforms, including location-scale transforms. Intuitively, increasing convex transforms, such as $x\mapsto x^{\beta}$ for $\beta>1$ and $x\mapsto\exp(x)$, generally make the tail of a random variable heavier. Thus, super-Pareto risks have heavier tails than (or equivalent to) Pareto(1) risks. It is straightforward to check that any super-Pareto random variable has infinite mean.

Some examples of the super-Pareto family include the generalized Pareto distribution when $\xi\ge 1$, specified by

$$\mathbb{P}(X\le x)=1-\left(1+\xi\frac{x}{\beta}\right)^{-1/\xi},\quad x\ge 0,$$

where $\beta>0$, and the Burr distribution when $\alpha,\tau\in(0,1]$, specified by

$$\mathbb{P}(X\le x)=1-\left(\frac{1}{x^{\tau}+1}\right)^{\alpha},\quad x\ge 0.$$

Special cases of the Burr family are the paralogistic ($\alpha=\tau$) and the log-logistic ($\alpha=1$) distributions.

The next proposition gives an equivalent formulation for super-Pareto distributions, which will become useful in proving some results.

Proposition 1. *A random variable X with essential infimum $z_X \in \mathbb{R}$ is super-Pareto if and only if the function $g : x \mapsto 1/\mathbb{P}(X > x)$ is strictly increasing and concave on $[z_X, \infty)$. If further X is regular, then $z_X > 0$ and $g(x) \le x/z_X$ for $x \ge z_X$.*

Next, we introduce a new notion of negative dependence.

Definition 2. A set $S \subseteq \mathbb{R}^k$, $k \in \mathbb{N}$ is *decreasing* if $\mathbf{x} \in S$ implies $\mathbf{y} \in S$ for all $\mathbf{y} \le \mathbf{x}$. Random variables $X_1, \dots, X_n$ are *weakly negatively associated* if for any $i \in [n]$, decreasing set $S \subseteq \mathbb{R}^{n-1}$, and $x \in \mathbb{R}$ with $\mathbb{P}(X_i \le x) > 0$,

$$\mathbb{P}(\mathbf{X}_{-i} \in S \mid X_i \le x) \le \mathbb{P}(\mathbf{X}_{-i} \in S), \tag{2}$$

where $\mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$.

Weak negative association includes independence as a special case. Intuitively, (2) means that knowing X_i is small implies that $\mathbf{X}_{-i}$ is less likely to be small, thus a concept of negative dependence. Moreover, (2) implies, by flipping signs,

$$\mathbb{P}(\mathbf{X}_{-i} > \mathbf{x} \mid X_i > x) \le \mathbb{P}(\mathbf{X}_{-i} > \mathbf{x}) \tag{3}$$

for $i \in [n]$, $x \in \mathbb{R}$, and $\mathbf{x} \in \mathbb{R}^{n-1}$.

Weak negative association is weaker than the popular notion of negative association (Alam and Saxena (1981) and Joag-Dev and Proschan (1983)), hence the name. Examples satisfying negative association, such as normal distributions with non-positive correlations, are studied by Joag-Dev and Proschan (1983). It is also implied by negative regression dependence (Lehmann (1966) and Block et al. (1985)), and it implies negative orthant dependence (Block et al. (1982)); see Remark 3 for more details on these notions of dependence.

In most results, we consider weakly negatively associated and identically distributed (WNAID) super-Pareto random variables. This setting includes the iid Pareto(α) model for $\alpha \in (0, 1]$.

2.2. The main result

For random variables X and Y, we write $X <_{\mathrm{st}} Y$ if $\mathbb{P}(X > x) < \mathbb{P}(Y > x)$ for all $x \in \mathbb{R}$ satisfying $\mathbb{P}(X > x) \in (0, 1)$, and this will be referred to as strict stochastic dominance. Note that this condition is stronger than a different notion of strict stochastic dominance defined by $X \le_{\mathrm{st}} Y$ and $X \not\ge_{\mathrm{st}} Y$. The following theorem is our main result.

Theorem 1. *Suppose that $X_1, \dots, X_n$ are WNAID super-Pareto, $(\theta_1, \dots, \theta_n) \in \Delta_n$, and $X \overset{\mathrm{d}}{=} X_1$.*

(i) Stochastic dominance holds:

$$X \leq_{\mathrm{st}} \sum_{i=1}^{n} \theta_i X_i, \tag{4}$$

and strict stochastic dominance $X <_{\mathrm{st}} \sum_{i=1}^n \theta_i X_i$ holds if $\theta_i > 0$ for at least two $i \in [n]$.

(ii) If X is regular, then for any events $A_1, \dots, A_n$ independent of $(X_1, \dots, X_n)$ and event A independent of X satisfying $\mathbb{P}(A) = \sum_{i=1}^n \theta_i \mathbb{P}(A_i)$, we have

$$X \mathbb{1}_A \leq_{\mathrm{st}} \sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i}. \tag{5}$$

We will say that a *diversification penalty* exists if (4) or (5) holds. To interpret Theorem 1, the left-hand side of (4) can be regarded as the loss of an agent who keeps their own risk, and the right-hand side is the loss of an agent who shares risks with other agents. By pooling among super-Pareto losses, agents expect to suffer less loss when their own loss occurs. However, every agent in the pool will have a higher frequency of bearing losses. Hence, diversification has two competing effects on the loss portfolio: It increases the frequency of losses and decreases the sizes of individual losses. The above results show that the combined effects of diversification of super-Pareto losses lead to a higher default probability at any capital reserve level, that is, $\mathbb{P}(\sum_{i=1}^n \theta_i X_i > x) > \mathbb{P}(X > x)$ for all $x > 1$.

The model in Theorem 1 (ii) reflects that catastrophic losses are large losses but usually occur with very small probabilities. Such losses are usually modelled by $X\mathbb{1}_A$, where X is a positive random loss, and A is an event with a small probability, indicating the occurrence of a catastrophe, such as an earthquake or a flood. Note that in (ii), the events $A_1, \dots, A_n$ are arbitrary, meaning that $X_1, \dots, X_n$ may be caused by either the same or different triggering events. In particular, if $A_1, \dots, A_n$ are the same, then A can also be chosen as the same event, and in this case (5) follows from (4). More generally, the condition $\mathbb{P}(A) = \sum_{i=1}^n \theta_i \mathbb{P}(A_i)$ makes it fair to compare the two sides of (5); for instance, if X has a finite mean instead of being super-Pareto, then both sides of (5) would have the same mean. Although our setting mainly concerns the losses of an agent, it is also applicable to the setting of investment decisions. For instance, $X_1, \dots, X_n$ can represent profits from technology innovations modelled by $A_1, \dots, A_n$; negatively dependent profits may arise in innovation races (see e.g., Shapiro and Varian (1999)).

An immediate consequence of Theorem 1 is that if super-Pareto random variables $X_1, \dots, X_n$ are independent and comparable in first-order stochastic dominance, for $(\theta_1, \dots, \theta_n) \in \Delta_n$, we have $X_{i^*} \leq_{\mathrm{st}} \sum_{i=1}^n \theta_i X_i$ if $X_{i^*} \leq_{\mathrm{st}} X_i$ for all $i \in [n]$.

To better understand the result in Theorem 1, we stress that (4) and (5) cannot be expected if $X_1, \dots, X_n$ have finite mean, regardless of their dependence structure, as summarized in the following proposition.

Proposition 2. *For $\theta_1,\dots,\theta_n>0$ with $\sum_{i=1}^n \theta_n=1$ and identically distributed random variables $X,X_1,\dots,X_n$ with finite mean and any dependence structure,* (4) *holds if and only if $X_1=\cdots=X_n$ almost surely.*

Proposition 2 implies, in particular, that (4) never holds for WNAID non-degenerate random variables $X,X_1,\dots,X_n$ with finite mean. Even if $X,X_1,\dots,X_n$ have an infinite mean, we are not aware of any other distributions for which (4) and (5) hold other than the super-Pareto distributions studied in this paper.

Risk measures are popular tools to quantify the risk of a financial portfolio; see Artzner et al. (1999) and Föllmer and Schied (2016). Value-at-Risk (VaR) is one of the most widely used classes of risk measures in financial regulation. For a random variable X with distribution function F_X, VaR at level $p\in(0,1)$ is defined as the (left) quantile of X at p, that is,

$$\mathrm{VaR}_p(X)=F_X^{-1}(p)=\inf\{t\in\mathbb{R}:F_X(t)\ge p\}.$$

For any random variable X, $\mathrm{VaR}_p(X)$ is always finite, making it suitable for assessing losses with infinite mean. Moreover, for two random variables X and Y, $X\le_{\mathrm{st}}Y$ (resp. $X<_{\mathrm{st}}Y$) if and only if $\mathrm{VaR}_p(X)\le\mathrm{VaR}_p(Y)$ (resp. $\mathrm{VaR}_p(X)<\mathrm{VaR}_p(Y)$) for all $p\in(0,1)$. Noting the equality $\mathrm{VaR}_p(X_1)=\sum_{i=1}^n\theta_i\mathrm{VaR}_p(X_i)$ for iid random variables, Theorem 1 gives superadditivity:

$$\mathrm{VaR}_p\left(\sum_{i=1}^n\theta_iX_i\right)>\sum_{i=1}^n\theta_i\mathrm{VaR}_p(X_i). \tag{6}$$

Inequality (6) and its non-strict version can be intuitively seen as diversification penalty for VaR_p.

2.3. Proof of Theorem 1

We first show in step (a) that the result in (i) holds for two independent Pareto(1) losses. An induction argument yields the result for the sums of n such losses. We then extend the result to WNAID Pareto losses in step (b), and further to super-Pareto losses in step (c). Finally, step (d) proves the result in (ii) by applying (i) and analysis on the combination of indicator functions, where the regularity assumption of X is needed.

(a) Let $Y,Y_1,\dots,Y_n\sim\mathrm{Pareto}(1)$ such that $Y_1,\dots,Y_n$ are independent and $\Delta_n^+=\Delta_n\cap(0,1)^n$. It suffices to show

$$\mathbb{P}(\theta_1Y_1+\cdots+\theta_nY_n\ge x)>\frac{1}{x}\text{ for all }x\in(1,\infty)\text{ and }(\theta_1,\dots,\theta_n)\in\Delta_n^+. \tag{7}$$

Let $\delta=(x-1+\theta_n)/\theta_n$. We will use the following fact: For $(y_1,\dots,y_n)\in(1,\infty)^n$, if $y_n\ge\delta$, then

$$\theta_1y_1+\cdots+\theta_ny_n\ge(\theta_1+\cdots+\theta_{n-1})+\theta_n\delta=1-\theta_n+x-1+\theta_n=x.$$

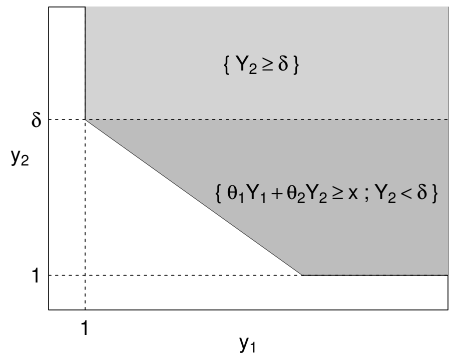

Figure 1 An illustration of $\{\theta_1 Y_1 + \theta_2 Y_2 \geq x\} = \{Y_2 \geq \delta\} \cup \{\theta_1 Y_1 + \theta_2 Y_2 \geq x; Y_2 < \delta\}$, where $x = 5$, $\theta_1 = 2/5$, $\theta_2 = 3/5$, and $\delta = 23/3$

We first show the case of $n = 2$. For all $x \in (1, \infty)$ and $(\theta_1, \theta_2) \in \Delta_2^+$ (see Figure 1),

$$\begin{aligned}
\mathbb{P}(\theta_1 Y_1 + \theta_2 Y_2 \geq x) &= \mathbb{P}(Y_2 \geq \delta) + \mathbb{P}(\theta_1 Y_1 + \theta_2 Y_2 \geq x; Y_2 < \delta) \\
&\geq \mathbb{P}(Y_2 \geq \delta) + \mathbb{P}(\theta_1 Y_1 + \theta_2 \geq x; Y_2 < \delta) \\
&= \frac{1}{\delta} + \frac{\theta_1}{x - \theta_2}\left(1 - \frac{1}{\delta}\right) \\
&> \frac{1}{\delta} + \frac{\theta_1}{x}\left(1 - \frac{1}{\delta}\right) \\
&= \frac{1}{x\delta}(x + \delta - \theta_2 \delta - 1 + \theta_2) = \frac{1}{x}.
\end{aligned} \tag{8}$$

Next, let $n > 2$ and assume (7) holds true for the case of $n - 1$. Then, for all $x \in (1, \infty)$ and $(\theta_1, \dots, \theta_n) \in \Delta_n^+$,

$$\begin{aligned}
\mathbb{P}(\theta_1 Y_1 + \cdots + \theta_n Y_n \geq x) &= \mathbb{P}(Y_n \geq \delta) + \mathbb{P}(\theta_1 Y_1 + \cdots + \theta_n Y_n \geq x; Y_n < \delta) \\
&\geq \mathbb{P}(Y_n \geq \delta) + \mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} \geq x - \theta_n; Y_n < \delta) \\
&= \mathbb{P}(Y_n \geq \delta) + \mathbb{P}\left(\frac{\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1}}{1 - \theta_n} \geq \frac{x - \theta_n}{1 - \theta_n}\right)\mathbb{P}(Y_n < \delta) \\
&\geq \frac{1}{\delta} + \frac{1 - \theta_n}{x - \theta_n}\left(1 - \frac{1}{\delta}\right) \\
&> \frac{1}{\delta} + \frac{1 - \theta_n}{x}\left(1 - \frac{1}{\delta}\right) = \frac{1}{x}.
\end{aligned} \tag{9}$$

By induction, (4) holds for n independent Pareto(1) random variables.

(b) Since $\{(y_1, \dots, y_n) : \theta_1 y_1 + \cdots + \theta_n y_n < t\}$ is a decreasing set, if $Y_1, \dots, Y_n$ are weakly negatively associated, for any $t > 1$ and $n \geq 2$,

$$\mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} \geq t; Y_n < \delta) = \mathbb{P}(Y_n < \delta) - \mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} < t; Y_n < \delta)$$

$$
\begin{aligned}
&\geq \mathbb{P}(Y_n < \delta) - \mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} < t)\mathbb{P}(Y_n < \delta) \\
&= \mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} \geq t)\mathbb{P}(Y_n < \delta).
\end{aligned}
$$

Therefore, if $Y_1, \ldots, Y_n$ are weakly negatively associated, the equal signs in (8) and (9) will become greater-than-or-equal signs and the proof in step (a) can still go through. Hence, (4) holds for weakly negatively associated Pareto(1) random variables.

(c) To show (4) for WNAID super-Pareto random variables, the following two lemmas will be helpful. Their proofs are in Appendix A, and they do not assume super-Pareto distributions.

Lemma 1. *Suppose that $X \overset{\mathrm{d}}{=} f(Y)$ for an increasing convex function f, $X_1 \overset{\mathrm{d}}{=} X$, and $X_1, \ldots, X_n$ are WNAID. There exist WNAID random variables $Y_1, \ldots, Y_n$ with $Y_1 \overset{\mathrm{d}}{=} Y$ such that*

$$
(X_1, \ldots, X_n) \overset{\mathrm{d}}{=} (f(Y_1), \ldots, f(Y_n)).
$$

Lemma 2. *Let $Y, Y_1, \ldots, Y_n$ be any random variables, and $(\theta_1, \ldots, \theta_n) \in \Delta_n$. If $Y \leq_{\mathrm{st}} \sum_{i=1}^n \theta_i Y_i$, then $f(Y) \leq_{\mathrm{st}} \sum_{i=1}^n \theta_i f(Y_i)$ for any increasing convex function f.*

By Lemma 1 and the definition of super-Pareto distribution, for any WNAID super-Pareto random variables $X_1, \ldots, X_n$ with $X_1 \overset{\mathrm{d}}{=} X$, there exist WNAID $Y_1, \ldots, Y_n \sim$ Pareto(1) such that $(X_1, \ldots, X_n) \overset{\mathrm{d}}{=} (f(Y_1), \ldots, f(Y_n))$. Therefore, for (4), it suffices to show

$$
f(Y) \leq_{\mathrm{st}} \sum_{i=1}^n \theta_i f(Y_i), \tag{10}
$$

which follows by combining Lemma 2 and the result in step (b). As it is shown in the previous steps that $Y <_{\mathrm{st}} \sum_{i=1}^n \theta_i Y_i$, it is clear that for $t > f(1)$, $\mathbb{P}\left(\sum_{i=1}^n \theta_i f(X_i) > t\right) > \mathbb{P}\left(f(X) > t\right)$, where $\theta_i > 0$ for at least two $i \in [n]$. Hence, the strictness statement after (4) also holds.

(d) Next we show (ii). Let $A_1, \ldots, A_n$ be events independent of $(X_1, \ldots, X_n)$ and event A be independent of X satisfying $\mathbb{P}(A) = \sum_{i=1}^n \theta_i \mathbb{P}(A_i)$. For $S \subseteq [n]$, let $B_S = \left(\bigcap_{i \in S} A_i\right) \cap \left(\bigcap_{i \in S^c} A_i^c\right)$. For $(\theta_1, \ldots, \theta_n) \in \Delta_n$, we write

$$
\sum_{i=1}^n \theta_i X_i \mathbb{1}_{A_i} = \sum_{S \subseteq [n]} \mathbb{1}_{B_S} \sum_{i \in S} \theta_i X_i.
$$

By (4), $\sum_{i \in S} \theta_i X_i \geq_{\mathrm{st}} \sum_{i \in S} \theta_i X$ for any $S \subseteq [n]$. As first-order stochastic dominance is closed under mixture (e.g., Theorem 1.A.3 (d) of Shaked and Shanthikumar (2007)), $\sum_{i \in S} \theta_i X_i \mathbb{1}_{B_S} \geq_{\mathrm{st}} \sum_{i \in S} \theta_i X \mathbb{1}_{B_S}$ for any $S \subseteq [n]$. Since B_S and B_R are mutually exclusive for any distinct $S, R \subseteq [n]$, we have

$$
\sum_{i=1}^n \theta_i X_i \mathbb{1}_{A_i} = \sum_{S \subseteq [n]} \mathbb{1}_{B_S} \sum_{i \in S} \theta_i X_i \geq_{\mathrm{st}} \sum_{S \subseteq [n]} \sum_{i \in S} \theta_i X \mathbb{1}_{B_S}.
$$

Finally, we need to show $\sum_{S\subseteq[n]}(\sum_{i\in S}\theta_i)X\mathbb{1}_{B_S}\ge_{\mathrm{st}} X\mathbb{1}_A$. For this, we prove the following statement. For mutually exclusive events $B_1,\dots,B_n$ independent of X and $(c_1,\dots,c_n)\in[0,1]^n$, we have

$$X\mathbb{1}_B \le_{\mathrm{st}} \sum_{i=1}^n c_i X\mathbb{1}_{B_i}, \tag{11}$$

where B is an event independent of X satisfying $\mathbb{P}(B)=\sum_{i=1}^n c_i\mathbb{P}(B_i)$. To show this, first note that the statement is clearly true if $c_1=\cdots=c_n=0$. If any components of $(c_1,\dots,c_n)$ are zero, the problem simply reduces its dimension. Hence, we assume that $(c_1,\dots,c_n)\in(0,1]^n$ for the rest of the proof. Let the survival function of X be $\mathbb{P}(X>x)=1/g(x)$ for $x\ge z_X$ and $\mathbb{P}(X>x)=1$ for $x<z_X$. As X is regular, $z_X>0$, the concavity of g, and $g(x)\le x/z_X$ for $x\ge z_X$ together imply $g(t/c)\le g(t)/c$ for $t\ge 0$ and $c\in(0,1]$. For $t\ge z_X$, as $B_1,\dots,B_n$ are mutually exclusive and $c_i\in(0,1]$ for all $i\in[n]$,

$$\mathbb{P}\left(\sum_{i=1}^n c_iX\mathbb{1}_{B_i}\le t\right)=1-\sum_{i=1}^n\frac{\mathbb{P}(B_i)}{g(t/c_i)}\le 1-\frac{1}{g(t)}\sum_{i=1}^n c_i\mathbb{P}(B_i)=1-\frac{\mathbb{P}(B)}{g(t)}=\mathbb{P}(X\mathbb{1}_B\le t).$$

For $t\in[0,z_X)$,

$$\mathbb{P}\left(\sum_{i=1}^n c_iX\mathbb{1}_{B_i}\le t\right)\le\mathbb{P}\left(\sum_{i=1}^n c_iX\mathbb{1}_{B_i}\le z_X\right)\le 1-\frac{\mathbb{P}(B)}{g(z_X)}=\mathbb{P}(X\mathbb{1}_B\le t),$$

where we used $g(z_X)=1$, implied by $1\le g(z_X)\le z_X/z_X$. This yields (11). As $\sum_{i\in S}\theta_i\in[0,1]$ for any $S\subseteq[n]$, and

$$\sum_{S\subseteq[n]}\mathbb{P}(B_S)\sum_{i\in S}\theta_i=\sum_{j=1}^n\theta_j\sum_{S\subseteq[n],j\in S}\mathbb{P}(B_S)=\sum_{j=1}^n\theta_i\mathbb{P}(A_j)=\mathbb{P}(A),$$

$\sum_{S\subseteq[n]}(\sum_{i\in S}\theta_i)X\mathbb{1}_{B_S}\ge_{\mathrm{st}} X\mathbb{1}_A$ follows from (11) and the desired result (5) holds.

2.4. A few remarks

The next few remarks discuss the relation of Theorem 1 to the literature and some technical issues.

Remark 1 (EVT). In the literature of EVT, it has been observed that, for iid extremely heavy-tailed Pareto losses $X_1,\dots,X_n$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i>t\right)\ge\mathbb{P}\left(X>t\right)$$

holds true asymptotically as $t\to\infty$; see, e.g., Kaas et al. (2004), Albrecher et al. (2006), and Embrechts et al. (2009). Theorem 1 implies that the same inequality holds for any $t\in\mathbb{R}$ regardless of whether t is large enough. This gives rise to implications for decision makers whose preferences are not determined purely by the tail behaviour of risks.

Remark 2 (Stable distributions). As an important class of heavy-tailed distributions, stable distributions have frequently appeared in the analysis of portfolio diversification (e.g., Ibragimov (2005), Ibragimov and Walden (2007, 2008)). Using majorization order, Ibragimov (2005) showed that diversification increases the risk of a portfolio which consists of iid stable random variables without finite mean. In particular, if the stable random variables are one-sided on the positive axis, diversification will increase the total loss in first-order stochastic dominance; Ibragimov and Walden (2010) applied this result to study the problem of optimal bundling strategies for extremely heavy-tailed valuations. On the other hand, if the stable random variables are not one-sided, diversification will make the total loss "more spread out", hence different from first-order stochastic dominance. These results were extended to the case when losses are convex transformations of iid infinite-mean stable random variables in Ibragimov and Walden (2008). For iid symmetric infinite-mean stable random variables truncated by a sufficiently large number, diversification still makes the total loss "more spread out", as shown by Ibragimov and Walden (2007).

Remark 3 (Notions of negative dependence). Among the following notions of negative dependence, weak negative association is weaker than (a) and (b) below, and stronger than (c).

(a) A random vector $\mathbf{X}=(X_1,\dots,X_n)$ is *negatively associated* if for every pair of disjoint sets A, B of $[n]$ and any functions f and g both increasing or decreasing coordinate-wise, provided the covariance below exists,
$$\operatorname{cov}(f(\mathbf{X}_A),g(\mathbf{X}_B))\le 0, \tag{12}$$
where $\mathbf{X}_A=(X_k)_{k\in A}$ and $\mathbf{X}_B=(X_k)_{k\in B}$ (Alam and Saxena (1981) and Joag-Dev and Proschan (1983)). It is known that random vectors following multivariate normal distributions with non-positive correlations are negatively associated, and so are those obtained from increasing transforms of such normal random vectors (Joag-Dev and Proschan (1983)). Choosing $A=\{i\}$, $B=[n]\setminus A$, $f(y)=\mathbb{1}_{\{y\le x\}}$ and $g(\mathbf{y})=\mathbb{1}_{\{\mathbf{y}\in S\}}$ in (12) yields (2), and hence weak negative association is implied.

(b) A random vector $\mathbf{X}$ is *negative regression dependent* if for every $i\in[n]$, the random variable $\mathbb{E}[g(\mathbf{X}_{-i})|X_i]$ is a decreasing function of X_i for any coordinate-wise increasing function g such that the conditional expectation exists; see Lehmann (1966), who only formulated the case $n=2$. This notion for general $n>2$ is called negative dependence through stochastic ordering by Block et al. (1985). To check that this notion is stronger than weak negative association, it suffices to take the function $g(\mathbf{x})=-\mathbb{1}_{\{\mathbf{x}\in S\}}$ for a decreasing set $S\subseteq\mathbb{R}^{n-1}$.

(c) A random vector $\mathbf{X}=(X_1,\dots,X_n)$ is *negatively orthant dependent* if for all $\mathbf{x}=(x_1,\dots,x_n)\in\mathbb{R}^n$, $\mathbb{P}(\mathbf{X}\le\mathbf{x})\le\prod_{i=1}^n\mathbb{P}(X_i\le x_i)$ and $\mathbb{P}(\mathbf{X}>\mathbf{x})\le\prod_{i=1}^n\mathbb{P}(X_i>x_i)$. The fact that (2) implies negative orthant dependence follows from (2) and (3). Negative orthant dependence is not sufficient for the proof of Theorem 1, because in the proof (see Section 2.3, step (b)) we need the inequality $\mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} < t, Y_n \le \delta) \le \mathbb{P}(\theta_1 Y_1 + \cdots + \theta_{n-1} Y_{n-1} < t)\mathbb{P}(Y_n \le \delta)$, which holds under weak negative association of $(Y_1, \ldots, Y_n)$ but not under negative orthant dependence.

Remark 4 (Majorization). Let $X_1, \ldots, X_n$ be iid super-Pareto random variables. Inspired by Theorem 1, a question is whether

$$\sum_{i=1}^{n} \eta_i X_i \le_{\mathrm{st}} \sum_{i=1}^{n} \theta_i X_i \tag{13}$$

holds for two vectors $(\theta_1, \ldots, \theta_n) \in \Delta_n$ and $(\eta_1, \ldots, \eta_n) \in \Delta_n$ increasing in majorization order; that is, $\sum_{i=1}^{n} \phi(\theta_i) \le \sum_{i=1}^{n} \phi(\eta_i)$ for all continuous and convex functions ϕ (see Marshall et al. (2011)). Theorem 1 corresponds to the case $(\eta_1, \ldots, \eta_n) = (1, 0, \ldots, 0)$. This question seems to be beyond the current techniques. For results similar to (13) on some distributions, see Proschan (1965) and Ibragimov (2005).

3. Conclusion

Our main result (Theorem 1) establishes that a weighted average of WNAID super-Pareto random variables, possibly triggered by different events, is larger than one such loss in the sense of first-order stochastic dominance. This result implies that the diversification of many super-Pareto losses without finite mean increases the risk assessment of a portfolio, according to the superadditivity of VaR. Some technical questions remain open and are discussed in Remark 4.

A. Proofs of all other results

Proof of Proposition 1 We first show the equivalence statement. For the "$\Leftarrow$" direction, let $\mathbb{P}(X \le x) = 1 - 1/g(x)$ for $x \in [z_X, \infty)$, where g is strictly increasing and concave on $[z_X, \infty)$. Let $f(y) = g^{-1}(y)$ for $y > g(z_X)$ and $f(y) = z_X$ for $1 \le y \le g(z_X)$. It is straightforward to see that for any Pareto(1) random variable Y, $f(Y) \overset{\mathrm{d}}{=} X$. Next, we show the "$\Rightarrow$" direction. For $x < \infty$, the right-continuous generalized inverse of f is $f^{-1+}(x) = \inf\{t : f(t) > x\}$. For $x \ge f(1)$, $\mathbb{P}(f(Y) \le x) = \mathbb{P}(Y \le f^{-1+}(x)) = 1 - 1/f^{-1+}(x)$. As f is increasing, convex, and non-constant, $g := f^{-1+}$ is strictly increasing and concave. Hence, g is concave and strictly increasing on $[z_X, \infty)$.

The statement on regularity follows by observing two facts. First, $f(1) > 0$ implies that $z_X = f(1) > 0$. Moreover, since f is convex and $f(0) = 0$, we have $f(y) \ge y f(1)$ for all $y > 1$, which gives $g(x) \le x/z_X$ for $x \ge z_X$ via $g = f^{-1+}$.

Proof of Proposition 2 Note that (1) implies that $\mathrm{ES}_p(X) \le \mathrm{ES}_p(\sum_{i=1}^{n} \theta_i X_i)$ for all $p \in (0,1)$, where ES_p is defined as

$$\mathrm{ES}_p(X) = \frac{1}{1-p} \int_p^1 \mathrm{VaR}_u(X) \mathrm{d}u.$$

Since ES_p is convex and $X_1,\dots,X_n$ are identically distributed, we have

$$\mathrm{ES}_p(X) \le \mathrm{ES}_p\left(\sum_{i=1}^n \theta_i X_i\right) \le \theta_i \sum_{i=1}^n \mathrm{ES}_p(X_i) = \mathrm{ES}_p(X), \quad p\in(0,1).$$

Using positive homogeneity of ES_p, it follows that the equality $\sum_{i=1}^n \mathrm{ES}_p(\theta_i X_i) = \mathrm{ES}_p(\sum_{i=1}^n \theta_i X_i)$ holds for each $p\in(0,1)$. By Theorem 5 of Wang and Zitikis (2021), this implies that $(\theta_1X_1,\dots,\theta_nX_n)$ is p-concentrated for each p; this result requires $X_1,\dots,X_n$ to have finite mean. Using Theorem 3 of Wang and Zitikis (2021), the above condition implies that $(X_1,\dots,X_n)$ is comonotonic. For definitions of comonotonicity and p-concentration, see Wang and Zitikis (2021). Since $X_1,\dots,X_n$ are identically distributed, comonotonicity further implies that $X_1=\dots=X_n$ almost surely.

Proof of Lemma 1 If f is constant, then there is nothing to show. If f is not constant, then there exists $z\in\mathbb{R}$ such that $f(x)$ is strictly increasing for $x>z$. Denote by $q=\mathbb{P}(X=z_X)$. For $i\in[n]$, let W_i be a uniform transform of X_i, that is, W_i is a standard uniform random variable such that $F_X^{-1}(W_i)=X_i$ a.s. A uniform transform always exists in an atomless probability space; see Lemma A.32 of Föllmer and Schied (2016). For all $i\in[n]$, let $A_i=\{X_i\le z_X\}$ and

$$U_i = qV_i\mathbb{1}_{A_i} + W_i\mathbb{1}_{A_i^c},$$

where $V_1,\dots,V_n$ are iid standard uniform random variables independent of $(X_1,\dots,X_n)$. Note that for each $i\in[n]$, U_i is also a uniform transform of X_i, and the distribution of $(U_1,\dots,U_n)$ is one possible copula of $(X_1,\dots,X_n)$. Let $Y_i=F_Y^{-1}(U_i)$ for $i\in[n]$. Note that $F_X^{-1}=f\circ F_Y^{-1}$ and f is strictly increasing for $x>z$. Hence Y_i is a strictly increasing function of X_i given V_i for each $i\in[n]$. Moreover, $(X_1,\dots,X_n)\overset{\mathrm{d}}{=}(f(Y_1),\dots,f(Y_n))$ as they have the same copula and marginal distributions. It remains to show that $Y_1,\dots,Y_n$ are weakly negatively associated. For any decreasing set $A\subseteq\mathbb{R}^{n-1}$ and $x\in\mathbb{R}$ with $F_Y(x)>0$, denote by $\beta=1/F_Y(x)$, and we have

$$\begin{aligned}
\mathbb{P}((Y_1,\dots,Y_{n-1})\in A|Y_n\le x) &= \mathbb{P}((Y_1,\dots,Y_{n-1})\in A, Y_n\le x)\beta\\
&= \mathbb{E}[\mathbb{P}((Y_1,\dots,Y_{n-1})\in A, Y_n\le x|V_1,\dots,V_n)]\beta\\
&\le \mathbb{E}[\mathbb{P}((Y_1,\dots,Y_{n-1})\in A|V_1,\dots,V_n)\mathbb{P}(Y_n\le x|V_1,\dots,V_n)]\beta\\
&= \mathbb{E}[\mathbb{P}((Y_1,\dots,Y_{n-1})\in A|V_1,\dots,V_{n-1})\mathbb{P}(Y_n\le x|V_n)]\beta\\
&= \mathbb{E}[\mathbb{P}((Y_1,\dots,Y_{n-1})\in A|V_1,\dots,V_{n-1})]\mathbb{E}[\mathbb{P}(Y_n\le x|V_n)]\beta\\
&= \mathbb{P}((Y_1,\dots,Y_{n-1})\in A)\mathbb{P}(Y_n\le x)\beta = \mathbb{P}((Y_1,\dots,Y_{n-1})\in A),
\end{aligned}$$

where the inequality holds because for each $i\in[n]$, conditional on V_i, Y_i is a strictly increasing function of X_i, and $X_1,\dots,X_n$ are weakly negatively associated. Therefore, $Y_1,\dots,Y_n$ are also weakly negatively associated.

Proof of Lemma 2 As f is convex and increasing, we have $f(Y) \leq_{\rm st} f(\sum_{i=1}^n \theta_i Y_i) \leq \sum_{i=1}^n \theta_i f(Y_i)$, where the first inequality holds as (4) holds for $Y_1, \dots, Y_n$ and the second inequality is due to the convexity of f.

Acknowledgments

The authors thank an Area Editor, an Associate Editor, and three anonymous referees for various relevant comments on the paper. In particular, the inclusion of negative dependence was inspired by a comment from an Associate Editor, and the open question in Remark 4 was inspired by comments from the referees. The authors thank Yuming Wang and Wenhao Zhu for ideas that led to a preliminary version of Theorem 1. We also thank Hansjörg Albrecher, Jan Dhaene, John Ery, Taizhong Hu, Massimo Marinacci, Alexander Schied, and Qihe Tang for helpful comments on a previous version of the paper. Ruodu Wang is supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-03823 and CRC-2022-00141).

References

Alam, K. and Saxena, K. M. L. (1981). Positive dependence in multivariate distributions. *Communications in Statistics-Theory and Methods*, 10(12):1183–1196.

Albrecher, H., Asmussen, S. and Kortschak, D. (2006). Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes*, 9(2):107–130.

Andriani, P. and McKelvey, B. (2007). Beyond Gaussian averages: Redirecting international business and management research toward extreme events and power laws. *Journal of International Business Studies*, 38(7):1212–1230.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Balkema, A. and de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2:792–804.

Baucells, M. and Heukamp, F. H. (2006). Stochastic dominance and cumulative prospect theory. *Management Science*, 52(9):1409–1423.

Bauer, D. and Zanjani, G. (2016). The marginal cost of risk, risk measures, and capital allocation. *Management Science*, 62(5):1431–1457.

Block, H. W., Savits, T. H. and Shaked, M. (1982). Some concepts of negative dependence. *Annals of Probability*, 10(3): 765–772.

Block, H. W., Savits, T. H. and Shaked, M. (1985). A concept of negative dependence using stochastic ordering. *Statistics & Probability Letters*, 3(2):81–86.

Caballé, J. and Pomansky, A. (1996). Mixed risk aversion. *Journal of Economic Theory*, 71(2):485–513.

Cheynel, E., Cianciaruso, D. and Zhou, F. (2022). Fraud power laws. *SSRN*: 4292259.

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

de Haan L. and Ferreira A. (2006). *Extreme Value Theory: An Introduction*. Springer.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg.

Embrechts, P., Resnick, S. I. and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal*, 3(2):30–41.

Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In *Risk Management: Value at Risk and Beyond* (Eds: Dempster) pp. 176–223. Cambridge University Press.

Embrechts, P., Lambrigger, D. and Wüthrich, M. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, 12(2):107–127.

Embrechts, P. and Puccetti, G. (2010). Risk aggregation. In *Copula Theory and its Applications* (Eds: Jaworski et al.) pp. 111–126. Springer, Heidelberg.

Fama, E. F. and Miller, M. H. (1972). *The Theory of Finance*. Dryden Press, Hinsdale.

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

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294.

Hadar, J. and Russell, W. R. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, 59(1):25–34.

Hadar, J. and Russell, W. R. (1971). Stochastic dominance and diversification. *Journal of Economic Theory*, 3(3):288–305.

Hofert, M. and Wüthrich, M. V. (2012). Statistical review of nuclear power accidents. *Asia-Pacific Journal of Risk and Insurance*, 7(1).

Huang, R. J., Tzeng, L. Y. and Zhao, L. (2020). Fractional degree stochastic dominance. *Management Science*, 66(10): 4630–4647.

Ibragimov, M., Ibragimov, R. and Walden, J. (2015). *Heavy-Tailed Distributions and Robustness in Economics and Finance*, Vol. 214 of Lecture Notes in Statistics, Springer.

Ibragimov, R. (2005). New majorization theory in economics and martingale convergence results in econometrics. Ph.D. dissertation, Yale University, New Haven, CT.

Ibragimov, R. (2009). Portfolio diversification and value at risk under thick-tailedness. *Quantitative Finance*, 9(5):565–580.

Ibragimov, R., Jaffee, D. and Walden, J. (2009). Non-diversification traps in markets for catastrophic risk. *Review of Financial Studies*, 22:959–993.

Ibragimov, R., Jaffee, D. and Walden, J. (2011). Diversification disasters. *Journal of Financial Economics*, 99(2):333–348.

Ibragimov, R. and Prokhorov, A. (2017). *Heavy Tails and Copulas: Topics in Dependence Modelling in Economics and Finance*. World Scientific.

Ibragimov, R. and Walden, J. (2007). The limits of diversification when losses may be large. *Journal of Banking and Finance*, 31(8):2551–2569.

Ibragimov, R. and Walden, J. (2008). Portfolio diversification under local and moderate deviations from power laws. *Insurance: Mathematics and Economics*, 42(2):594–599.

Ibragimov, R. and Walden, J. (2010). Optimal bundling strategies under heavy-tailed valuations. *Management Science*, 56(11):1963–1976.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *Annals of Statistics*, 286–295.

Kaas, R., Goovaerts, M. and Tang, Q. (2004). Some useful counterexamples regarding comonotonicity. *Belgian Actuarial Bulletin*, 4(1):1–4.

Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(5):1137–1153.

Levy, H. (1992). Stochastic dominance and expected utility: Survey and analysis. *Management Science*, 38(4):555–593.

Levy, H. (2016). *Stochastic Dominance. Investment Decision Making under Uncertainty*. Third Edition. Springer.

Malinvaud, E. (1972). The allocation of individual risks in large markets. *Journal of Economic Theory*, 4(2):312–328.

Marshall, A. W., Olkin, I. and Arnold, B. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer, 2nd edition.

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

Müller, A., Scarsini, M., Tsetlin, I. and Winkler. R. L. (2017). Between first and second-order stochastic dominance. *Management Science*, 63(9):2933–2947.

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

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131.

Proschan, F. (1965). Peakedness of distributions of convex combinations. *Annals of Mathematical Statistics*, 36(6):1703–1706.

Quirk, J. P. and Saposnik, R. (1962). Admissibility and measurable utility functions. *The Review of Economic Studies*, 29(2):140–146.

Rothschild, M. and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, 2(3):225–243.

Samuelson, P. A. (1967). General proof that diversification pays. *Journal of Financial and Quantitative Analysis*, 2(1):1–13.

Schmidt, U. and Zank, H. (2008). Risk aversion in cumulative prospect theory. *Management Science*, 54:208–216.

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

Shapiro, C. and Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Taleb, N. (2020). *Statistical Consequences of Fat Tails*. STEM Academic Press.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, 67:1413–1429.

Whitmore, G. A. (1970). Third-degree stochastic dominance. *The American Economic Review*, 60(3):457–459.

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