An unexpected stochastic dominance: Pareto distributions, catastrophes, and risk exchange

Yuyu Chen∗ Paul Embrechts† Ruodu Wang‡

October 6, 2023

Abstract

We show the perhaps surprising inequality that the weighted average of negatively dependent super-Pareto random variables, possibly caused by triggering events, is larger than one such random variable in the sense of first-order stochastic dominance. The class of super-Pareto distributions is extremely heavy-tailed and it includes the class of infinite-mean Pareto distributions. We discuss several implications of this result via an equilibrium analysis in a risk exchange market. First, diversification of super-Pareto losses increases portfolio risk, and thus a diversification penalty exists. Second, agents with super-Pareto losses will not share risks in a market equilibrium. Third, transferring losses from agents bearing super-Pareto losses to external parties without any losses may arrive at an equilibrium which benefits every party involved. The empirical studies show that our new inequality can be observed empirically for real datasets that fit well with extremely heavy tails.

Keywords: Pareto distributions; diversification effect; risk pooling; equilibrium; 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

∗Department of Economics, University of Melbourne, Australia. yuyu.chen@unimelb.edu.au
†RiskLab, Department of Mathematics & ETH Risk Center, ETH Zurich, Switzerland. embrechts@math.ethz.ch
‡Department of Statistics and Actuarial Science, University of Waterloo, Canada. wang@uwaterloo.ca
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 after careful statistical analysis 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; Section 1.1 below collects some examples and related literature. In risk management, such infinite-mean models often lead to intriguing phenomena. The following question is of particular interest to us.

Q. Suppose that there is a pool of identically distributed extremely heavy-tailed losses, possibly statistically dependent. An agent needs to decide whether to diversify on these risks or to concentrate on one of them. Without knowing the preference of the agent, what can we know about the optimal decision?

This question will be used to motivate our model and main result. The multiple-agent version of this question is addressed in the equilibrium model in Section 4.

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 (Hadar and Russell (1969) 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, which implies essentially all other forms. For two random variables $X$ and $Y$ representing random losses, we say $X$ is smaller than $Y$ in first-order stochastic dominance, denoted by
$X \leq_{st} Y$, if $\mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x)$ for all $x \in \mathbb{R}$. The relation $X \leq_{st} Y$ means that all decision makers with an increasing utility function will prefer the loss $X$ to the loss $Y$, as studied by Quirk and Saposnik (1962) and Hadar and Russell (1969, 1971). In this paper, all terms like “increasing” and “decreasing” are in the non-strict sense.

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

$$X_1 \leq_{st} \theta_1 X_1 + \cdots + \theta_n X_n,$$

and the inequality is strict in a natural sense. As far as we are aware, inequality (1) is not known in the literature, even in the case that $\theta_1, \ldots, \theta_n$ are equal (i.e., they are $1/n$). It is somewhat surprising that, for infinite-mean losses, inequality (1) holds for the strongest form of risk comparison: For every monotone decision maker (with precise definition in Section 3), 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.

To appreciate the remarkable nature of (1), we first remark that for any identically distributed random variables $X_1, \ldots, X_n$ with finite mean, regardless of their distribution or dependence structure, for $\theta_1, \ldots, \theta_n > 0$ with $\sum_{i=1}^{n} \theta_i = 1$, (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.

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 $Z_1, \ldots, Z_n$ which follow a convolution of symmetric stable distributions without finite mean, $\mathbb{P}(\theta_1 Z_1 + \cdots + \theta_n Z_n \leq x) \leq \mathbb{P}(Z_1 \leq x)$ 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 $Z_1, \ldots, Z_n$ can take negative values, unlike Pareto losses, which are positive, skewed and more suitable for the modeling of
extreme losses.

In risk management, inequality (1) yields superadditivity of the regulatory risk measure Value-at-Risk (VaR) in banking and insurance sectors; that is, the weighted average of Pareto losses without finite mean gives a larger VaR than that given by an individual Pareto loss. 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 just in some asymptotic sense.

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 $X_1, \ldots, X_n$ are weakly negatively associated super-Pareto random variables, and also in case they are triggered by some events.

We discuss in Section 3 the implications of Theorem 1 and related inequalities on the risk management decision of a single agent. For super-Pareto losses, the action of diversification increases the risk uniformly for all risk preferences, such as VaR, expected utilities, and distortion risk measures, as long as the risk preferences are monotone and well defined. The increase of the portfolio risk is strict, and it provides an important implication in decision making: For an agent who faces super-Pareto losses and aims to minimize their risk by choosing a position across these losses, the optimal decision is to take only one of the super-Pareto losses (i.e., no diversification).

We proceed to study the equilibria of a risk exchange market for super-Pareto losses in Section 4. As individual agents do not benefit from diversification in a risk exchange market where super-Pareto losses are present, we may expect that agents will not share their losses with each other. Indeed, if each agent in the market is associated with an initial position in one of these super-Pareto losses, the agents will merely exchange the entire loss position instead of risk sharing in an equilibrium model (Theorem 2). The situation becomes quite different if the agents with initial losses are allowed to transfer their losses to external parties, or if the losses have finite mean; in either case, agents are willing to share or transfer risks among themselves. Analyses of these cases are put in Appendix D. The above results are consistent with the observations made in Ibragimov et al. (2011) based on a different model.

The class of distributions satisfying (1) beyond Theorem 1 is discussed in Section 5. In Section 6, numerical and real data examples are presented to illustrate the presence of extremely heavy tails in two real datasets in which the phenomenon of inequality (1) can be empirically observed.
We proceed to study the diversification effects of extremely heavy-tailed Pareto losses with different tail indices. Section 7 concludes the paper. Some background on risk measures is put in Appendix A. Proofs of all results, except for Theorem 1, are put in Appendix B. Some generalizations of our main result to other models are presented in Appendix C. Appendix D contains generalizations of the risk exchange market studied in Section 4. Some details for the numerical results in Section 6 are put in Appendix E.

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 \([n] = \{1, \ldots, n\}\). Denote by \(\Delta_n\) the standard simplex, that is, \(\Delta_n = \{(\theta_1, \ldots, \theta_n) \in [0,1]^n : \sum_{i=1}^n \theta_i = 1\}\). For \(x, y \in \mathbb{R}\), write \(x \land y = \min\{x, y\}\), \(x \lor y = \max\{x, y\}\), and \(x_+ = \max\{x, 0\}\). We write \(X \overset{d}{=} Y\) if \(X\) and \(Y\) have the same distribution. We always assume \(n \geq 2\). Equalities and inequalities are interpreted component-wise when applied to vectors.

### 1.1 Infinite-mean Pareto-type models

The key assumption of our paper is that losses follow super-Pareto distributions, which have infinite mean. Whereas statistical models with some divergent higher moments are ubiquitous throughout the risk management literature, the infinite mean case needs more specific motivation. For power-tail data, a standard approach for the estimation of the underlying tail parameters is the Peaks Over Threshold (POT) methodology from EVT; see Embrechts et al. (1997). Other estimation methods include the classic Hill estimators (see Embrechts et al. (1997)) and the log-log rank-size estimation (see Ibragimov et al. (2015)). It is known that the Hill estimator may be sensitive to the dependence in the data and small sample sizes, and the log-log rank-size estimation can be biased in small samples; see Gabaix and Ibragimov (2011) for an improved version of log-log rank-size estimation. As we will discuss in Proposition A.2 and Section 6.2, our results apply to the case of the generalized Pareto distribution which is the basic model for the POT set-up. Below we discuss some examples from the literature leading to extremely heavy-tailed Pareto models; extra data examples are provided in Section 6.2.\(^1\)

In the parameterization used in Section 2, a tail parameter \(\alpha \leq 1\) corresponds to an infinite-mean Pareto model. Ibragimov et al. (2009) used standard seismic theory to show that the tail indices \(\alpha\) of earthquake losses lie in the range \([0.6, 1.5]\). Estimated by Rizzo (2009), the tail indices \(\alpha\) for some wind catastrophic losses are around 0.7. Hofert and Wüthrich (2012) showed that the tail indices \(\alpha\) of losses caused by nuclear power accidents are around \([0.6, 0.7]\); similar observations

\(^1\)For these examples, it turns out that infinite-mean models yield a better overall fit than finite-mean ones, although one can never say for sure that any real world dataset is generated by an infinite-mean model.
can be found in Sornette et al. (2013). Based on data collected by the Basel Committee on Banking Supervision, Moscadelli (2004) reported the tail indices $\alpha$ of (over 40000) operational losses in 8 different business lines to lie in the range $[0.7, 1.2]$, with 6 out of the 8 tail indices being less than 1, with 2 out of these 6 significantly less than 1 at a 95% confidence level. For a detailed discussion on the risk management consequences in this case, see Neslehová et al. (2006). Losses from cyber risk have tail indices $\alpha \in [0.6, 0.7]$; see Eling and Wirfs (2019), Eling and Schnell (2020) and the references therein. In a standard Swiss Solvency Test document (FINMA (2021, p. 110)), most major damage insurance losses are modelled by a Pareto distribution with default parameter $\alpha$ in the range $[1, 2]$, with $\alpha = 1$ attained by some aircraft insurance. As discussed by Beirlant et al. (1999), some fire losses collected by the reinsurance broker AON Re Belgium have tail indices $\alpha$ around 1. Biffis and Chavez (2014) showed that several large commercial property losses collected from two Lloyd’s syndicates have tail indices $\alpha$ considerably less than 1. Silverberg and Verspagen (2007) concluded that the tail indices $\alpha$ are less than 1 for financial returns from some technological innovations. The tail part of cost overruns in information technology projects can have tail indices $\alpha \leq 1$; see Flyvbjerg et al. (2022). Besides large financial losses and returns, the number of deaths in major earthquakes and pandemics modelled by Pareto distributions also has infinite mean; see Clark (2013) and Cirillo and Taleb (2020). City sizes and firm sizes follow Zipf’s law ($\alpha \approx 1$); see Gabaix (1999) and Axtell (2001). In the model of Cheynel et al. (2022), which considers managers’ fraudulent behavior, detected fraud size behaves like a power law with tail index 1. Heavy-tailed to extremely heavy-tailed models also occur in the realm of climate change and environmental economics. Weitzman’s Dismal Theorem (see Weitzman (2009)) discusses the breakdown of standard economic thinking like cost-benefit analysis in this context. This led to an interesting discussion with William Nordhaus, a recipient of the 2018 Nobel Memorial Prize in Economic Sciences; see Nordhaus (2009).

The above references exemplify the occurrence of infinite mean models. Our perspective on these examples and discussions is that if these models are the result of some careful statistical analyses, then the practicing modeler has to take a step back and carefully reconsider the risk management consequences. Of course, in practice, there are several methods available to avoid such extremely heavy-tailed models, like cutting off the loss distribution model at some specific level, or tapering (concatenating a light-tailed distribution far in the tail of the loss distribution). In examples like those referred to above, such corrections often come at the cost of a great variability depending on the methodology used. It is in this context that our results add to the existing literature and modeling practice in cases where power-tail data play an important role.
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, used throughout the paper. A common parameterization of Pareto distributions is given by, for $\theta, \alpha > 0$,

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

As $\theta$ is a scale parameter, it suffices to study $P_{\alpha,1}$, which we write as Pareto($\alpha$). The mean of $P_{\alpha,\theta}$ is infinite if and only if the tail parameter $\alpha$ is in $(0, 1]$. We say that the $P_{\alpha,\theta}$ distribution is extremely heavy-tailed if $\alpha \leq 1$, and it is moderately heavy-tailed if $\alpha > 1$. The essential infimum of $X$ is given by $z_X = \inf\{z \in \mathbb{R} : P(X > z) > 0\}$.

**Definition 1.** A random variable $X$ with essential infimum $z_X \in \mathbb{R}$ is super-Pareto (or has a super-Pareto distribution) if the function $g : x \mapsto 1/P(X > x)$ is strictly increasing and concave on $[z_X, \infty)$. Moreover, $X$ is regular if $z_X > 0$ and $g(x) \leq x/z_X$ for $x \geq z_X$.

For $\alpha \in (0, 1]$, the function $g : x \mapsto 1/(1 - P_{\alpha,\theta}(x)) = (x/\theta)^\alpha \vee 1$ is strictly increasing, concave, and bounded by $x/\theta$ on $[\theta, \infty)$. Hence, all extremely heavy-tailed Pareto distributions are super-Pareto and regular. The same holds for the generalized Pareto distribution when $\xi \geq 1$, given by

$$ G_{\xi,\beta}(x) = 1 - \left( 1 + \xi \frac{x}{\beta} \right)^{-1/\xi}, \quad x \geq 0, \quad (2) $$

where $\beta > 0$. The next proposition gives an alternative formulation of super-Pareto distributions, which can be checked directly.

**Proposition 1.** A random variable $X$ is super-Pareto if and only if $X \overset{d}{=} f(Y)$ for some increasing, convex, and non-constant function $f$ and $Y \sim$ Pareto(1). Any super-Pareto random variable has infinite mean.

An immediate consequence of Proposition 1 is that 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 a heavier tail than (or equivalent to) Pareto(1) risks.

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 $x \in S$ implies $y \in S$ for all $y \leq x$. Random variables $X_1, \ldots, X_n$ are weakly negatively associated if for any $i \in [n]$, decreasing set $S \subseteq \mathbb{R}^{n-1}$, and $x \in \mathbb{R}$ with $\mathbb{P}(X_i \leq x) > 0$,

$$\mathbb{P}(X_{-i} \in S \mid X_i \leq x) \leq \mathbb{P}(X_{-i} \in S),$$

(3)

where $X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

Weak negative association includes independence as a special case. Intuitively, (3) means that knowing $X_i$ is small implies that $X_{-i}$ is less likely to be small, thus a concept of negative dependence. Moreover, (3) implies, by flipping signs,

$$\mathbb{P}(X_{-i} > x \mid X_i > x) \leq \mathbb{P}(X_{-i} > x)$$

(4)

for $i \in [n]$, $x \in \mathbb{R}$, and $x \in \mathbb{R}^{n-1}$. Interpreting $X$ as a vector of profits, inequality (4) reflects on the “winner-takes-all” phenomenon relevant in innovation races, that is, a single firm will gain a significant amount of profits over its competitors (see e.g., Shapiro and Varian (1999)).

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($\alpha$) model for $\alpha \in (0, 1]$.

2.2 The main result

For random variables $X$ and $Y$, we write $X <_{st} Y$ if $\mathbb{P}(X > x) < \mathbb{P}(Y > x)$ for all $x > z_X$. The following theorem is our main result.

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

(i) Stochastic dominance holds:

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

(5)

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

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

$$X \mathbb{1}_A \leq_{st} \sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i}. \quad (6)$$

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

The model in Theorem 1 (ii) reflects on the fact that catastrophic losses are large losses but usually occur with very small probabilities. Note that in (ii), the events $A_1, \ldots, A_n$ are arbitrary, meaning that $X_1, \ldots, X_n$ may be caused by either the same or different triggering events. Although our setting mainly concerns the losses of an agent, it is also applicable to the setting of investment decisions. For instance, $X_1, \ldots, X_n$ can be treated as profits from technology innovations modelled by $A_1, \ldots, A_n$; negatively dependent profits may arise in innovation races as mentioned above.

To interpret Theorem 1, the left-hand side of (5) 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 having losses and decreases the sizes of the 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.

Some special cases of Theorem 1 can be useful in different contexts. Applying (5) with a convex transform $x \mapsto (x - m)_+$, we get $(X_1 - m)_+ \leq_{st} \sum_{i=1}^{n} \theta_i (X_i - m)_+$. Here, $(X_1 - m)_+, \ldots, (X_n - m)_+$ can be interpreted as the losses assumed by a reinsurer who issues excess-of-loss reinsurance contracts with retention level $m$. Similarly, $X_1 \lor m \leq_{st} \sum_{i=1}^{n} \theta_i (X_i \lor m)$. By letting $X = X_1$, $\mathbb{P}(A_1) = \cdots = \mathbb{P}(A_n)$, and $A = A_1$ in (6), we get $X_1 \mathbb{1}_{A_1} \leq_{st} \sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i}$, which has a similar interpretation to (5).

We will say that a diversification penalty exists if (5) or (6) holds, which is naturally interpreted as that having exposures in multiple super-Pareto losses is worse than having just one super-Pareto loss of the same total exposure. Therefore, for an agent who wants to minimize the default probability of their loss portfolio, the optimal strategy is to not diversify. In fact, as long as the risk preference of the agent is monotone, non-diversification is always optimal; this is implied by
Theorem 1 and will be made more clear in Section 3. Such a diversification penalty holds for many distributions that are more heavy-tailed than Pareto(1) distribution. On the other hand, if the super-Pareto random variables in the previous discussion are considered as profits, (5) and (6) will suggest a diversification benefit for investors.

To better understand the result in Theorem 1, we stress that (5) and (6) cannot be expected if \( X_1, \ldots, X_n \) have finite mean, regardless of their dependence structure, as summarized in the following proposition.

**Proposition 2.** For \( \theta_1, \ldots, \theta_n > 0 \) with \( \sum_{i=1}^{n} \theta_n = 1 \) and identically distributed random variables \( X, X_1, \ldots, X_n \) with finite mean and any dependence structure, (5) holds if and only if \( X_1 = \cdots = X_n \) almost surely.

Proposition 2 implies, in particular, that (5) never holds for WNAID non-degenerate random variables \( X, X_1, \ldots, X_n \) with finite mean. As such, Theorem 1 yields a clear and elegant methodological distinction between the two modeling environments; the difference between finite and infinite mean acts as a kind of phase-type transition concerning diversification. Even if \( X, X_1, \ldots, X_n \) have an infinite mean, we are not aware of any other distributions for which (5) and (6) hold other than the ones in this paper.

The next few remarks discuss the relation of Theorem 1 to the literature and some technical issues. More general models for which diversification penalty (5) holds are discussed in Section 5.

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

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

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; see Sections 3 and 4.

**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); a diversification penalty also holds for truncated super-Pareto distributions (Proposition 5 below).

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 $\mathbf{X} = (X_1, \ldots, X_n)$ 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,

$$\text{cov}(f(\mathbf{X}_A), g(\mathbf{X}_B)) \leq 0,$$

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 $A = \{i\}$, $B = [n] \setminus A$, $f(y) = \mathbb{I}_{\{y \leq x\}}$ and $g(y) = \mathbb{I}_{\{y \in S\}}$ in (7) yields (3), 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 $g(x) = -\mathbb{I}_{\{x \in S\}}$ for a decreasing set $S \subseteq \mathbb{R}^{n-1}$.

(c) A random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is negatively orthant dependent if for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $\mathbb{P}(\mathbf{X} \leq \mathbf{x}) \leq \prod_{i=1}^n \mathbb{P}(X_i \leq x_i)$ and $\mathbb{P}(\mathbf{X} > \mathbf{x}) \leq \prod_{i=1}^n \mathbb{P}(X_i > x_i)$. The fact that (3) implies negative orthant dependence can be shown by induction.

2.3 Proof of Theorem 1

(i) By Proposition 1, there exists $Y \sim \text{Pareto}(1)$ and an increasing, convex, and non-constant function $f$ such that $X \overset{d}{=} f(Y)$. If $f$ is strictly increasing (corresponding to $X$ having a continuous
distribution), then, since weak negative association is preserved under increasing marginal transforms, there exist WNAID \(Y_1, \ldots, Y_n \sim \text{Pareto}(1)\) such that \((X_1, \ldots, X_n) \overset{d}{=} (f(Y_1), \ldots, f(Y_n))\). This also holds when \(X\) does not have a continuous distribution, which requires some analysis presented in Lemma 1 below with its proof in Appendix B. Therefore, to show (5), it suffices to show

\[
f(Y) \leq_{st} \sum_{i=1}^{n} \theta_i f(Y_i). \tag{8}
\]

We first prove

\[
Y \leq_{st} \sum_{i=1}^{n} \theta_i Y_i. \tag{9}
\]

For \((u_1, \ldots, u_n) \in (0,1)^n\) and \(\theta = (\theta_1, \ldots, \theta_n) \in \Delta_n\), define the generalized weighted average \(M_{r,\theta}(u_1, \ldots, u_n) = (\theta_1 u_1^r + \cdots + \theta_n u_n^r)^{\frac{1}{r}}\) where \(r \in \mathbb{R}\). Note that \(Y \leq_{st} \sum_{i=1}^{n} \theta_i Y_i\) can be equivalently written as

\[
M_{-1,\theta}(U_1, \ldots, U_n) \leq_{st} U, \tag{10}
\]

where \(U = 1 - P_{1,1}(Y)\) and \(U_i = 1 - P_{1,1}(Y_i), i \in [n]\). The random variables \(U, U_1, \ldots, U_n\) are uniformly distributed on \((0,1)\). Denote by \(R_{\theta}(p) = \mathbb{P}(M_{-1,\theta}(U_1, \ldots, U_n) \leq p)\) for \(\theta \in \Delta_n\) and \(p \in [0,1]\). If some of \(\theta_1, \ldots, \theta_n\) are 0, we can reduce the dimension of the problem. Denote by \(\Delta_n^+ = \Delta_n \cap (0,1)^n\). For \(n \geq 2\), we will prove by induction that

\[
R_{\theta}(p) > p \text{ for all } p \in (0,1) \text{ and } \theta \in \Delta_n^+. \tag{11}
\]

Assume that \(n \geq 3\). Suppose (9) holds true for the case of \(n-1\). As \(Y_1, \ldots, Y_n\) are weakly negatively associated, so are \(Y_1, \ldots, Y_{n-1}\). By assumption, we have \(R_\nu(p) \geq p\) for all \(p \in [0,1]\) and all \(\nu \in \Delta_{n-1}\). For a fixed \(p \in [0,1]\), let \(\delta = \theta_n/(p^{-1} - 1 + \theta_n)\). For \((u_1, \ldots, u_n) \in (0,1)^n\), if \(u_n \leq \delta\), then

\[
\theta_1 u_1^{-1} + \cdots + \theta_n u_n^{-1} \geq (\theta_1 + \cdots + \theta_{n-1}) + \theta_n \delta^{-1} = 1 - \theta_n + p^{-1} - 1 + \theta_n = p^{-1}.
\]

Hence, \(M_{-1,\theta}(u_1, \ldots, u_n) \leq p\) if \(u_n \leq \delta\).

Since \(Y_1, \ldots, Y_n\) are weakly negatively associated and \(U_1, \ldots, U_n\) are strictly monotone transforms of them,

\[
\mathbb{P} \left( \left\{ \sum_{i=1}^{n-1} \theta_i U_i^{-1} \geq t \right\} \cap \{U_n \leq \delta\} \right) \leq \delta \mathbb{P} \left( \sum_{i=1}^{n-1} \theta_i U_i^{-1} \geq t \right). \tag{12}
\]
For $t > 0$, denote by $A_t = \{\sum_{i=1}^{n-1} \theta_i U_i^{-1} \geq t\}$. We have

$$
P( A_t \cap \{U_n > \delta\} ) = P(A_t) - P(A_t \cap \{U_n \leq \delta\})
\geq P( A_t ) - \delta P( A_t ) = (1 - \delta) R_\psi \left( (1 - \theta_n) t^{-1} \right) \geq (1 - \delta)(1 - \theta_n)t^{-1},
$$
where $\nu = (\theta_1, \ldots, \theta_{n-1})/(1 - \theta_n) \in \Delta_{n-1}$. It follows that

$$
R_\theta(p) = P \left( \sum_{i=1}^{n} \theta_i U_i^{-1} \geq p^{-1} \right)
\geq P(U_n \leq \delta) + P(A_{1/p-\theta_n} \cap \{U_n > \delta\})
\geq \delta + (1 - \delta)(1 - \theta_n)(p^{-1} - \theta_n)^{-1}
\geq \delta + (1 - \delta)(1 - \theta_n)p = (1 - \theta_n)p + \delta(1 - p + \theta_np) = p.
$$

Thus, (11) holds. It remains to verify the case of $n = 2$. Note that if $n = 1$, $R_\theta(p) = p$ holds for all $p \in [0, 1]$. In the above argument, we do not need $R_\nu(p) > p$; indeed, $R_\nu(p) \geq p$ is sufficient, which holds for $n = 1$. By induction, (11) holds and so does (9).

As $f$ is convex and increasing, we have $f(Y) \leq \sum_{i=1}^{n} \theta_i f(Y_i) \leq \sum_{i=1}^{n} \theta_i f(Y_i)$, where the first inequality follows from (9) and the second inequality is due to the convexity of $f$. Hence, (8) holds.

By (11), it is clear that for $t > f(1)$, $P(\sum_{i=1}^{n} \theta_i f(X_i) > t) > P(f(X) > t)$ where $\theta_i > 0$ for at least two $i \in [n]$. Hence, the strictness statement after (5) holds.

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

$$
\sum_{i=1}^{n} \theta_i X_i 1_{A_i} = \sum_{S \subseteq [n]} 1_{B_S} \sum_{i \in S} \theta_i X_i.
$$
By (5), $\sum_{i \in S} \theta_i X_i \geq \sum_{i \in S} \theta_i X$ for any $S \subseteq [n]$. As $A_1, \ldots, A_n$ are independent of $(X_1, \ldots, X_n)$, by Theorem 1.A.14 of Shaked and Shanthikumar (2007), $\sum_{i \in S} \theta_i X_i 1_{B_S} \geq \sum_{i \in S} \theta_i X 1_{B_S}$ for any $S \subseteq [n]$. Since $B_S$ and $B_R$ are mutually exclusive for any distinct $S, R \subseteq [n]$, we have

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

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

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

\[ X \mathbb{1}_B \leq_{st} \sum_{i=1}^{n} c_i X \mathbb{1}_{B_i}, \quad (13) \]

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 \( (c_1, \ldots, c_n) \) are zero, the problem simply reduces its dimension. Hence, we assume that \( (c_1, \ldots, c_n) \in (0, 1]^n \) for the rest of the proof. Let the survival function of \( X \) be \( \mathbb{P}(X > x) = 1/g(x) \) for \( x \geq z_X \) and \( \mathbb{P}(X > x) = 1 \) for \( x < z_X \). As \( X \) is regular, \( z_X > 0 \), the concavity of \( g \), and \( g(x) \leq x/z_X \) for \( x \geq z_X \) together imply \( g(t/c) \leq g(t)/c \) for \( t \geq 0 \) and \( c \in (0, 1] \). For \( t \geq z_X \), as \( B_1, \ldots, B_n \) are mutually exclusive and \( c_i \in (0, 1] \) for all \( i \in [n] \),

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

For \( t \in [0, z_X] \),

\[ \mathbb{P} \left( \sum_{i=1}^{n} c_i X \mathbb{1}_{B_i} \leq t \right) \leq \mathbb{P} \left( \sum_{i=1}^{n} c_i X \mathbb{1}_{B_i} \leq z_X \right) \leq 1 - \frac{\mathbb{P}(B)}{g(z_X)} = \mathbb{P}(X \mathbb{1}_B \leq t), \]

where we used \( g(z_X) = 1 \), implied by \( 1 \leq g(z_X) \leq z_X/z_X \). This yields (13). As \( \sum_{i \in S} \theta_i \in [0, 1] \) for any \( S \subseteq [n] \), the desired result (6) follows from (13).

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

\( (X_1, \ldots, X_n) \overset{d}{=} (f(Y_1), \ldots, f(Y_n)) \).
3 Risk management decisions of a single agent

3.1 No diversification for a monotone agent

As hinted by Theorem 1, in a pool of WNAID super-Pareto losses, an agent who can choose between keeping their own loss or sharing losses with other agents has no incentive to enter the risk sharing pool, because it will increase their total risk. In this section, we make this observation rigorous by formally considering risk preference models.

Some further notation will be useful. Let \( X \) be the set of all random variables, and let \( L^1 \subseteq X \) be the set of random variables with finite mean. For \( X \in \mathcal{X} \), denote by \( F_X \) the distribution function. Denote by \( F_X^{-1} \) the (left) quantile function of \( X \), that is,

\[
F_X^{-1}(p) = \inf\{t \in \mathbb{R} : F_X(t) \geq p\}, \quad p \in (0, 1].
\]

For vectors \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) and \( y = (y_1, \ldots, y_n) \in \mathbb{R}^n \), their dot product is \( x \cdot y = \sum_{i=1}^n x_i y_i \) and we denote by \( \|x\| = \sum_{i=1}^n |x_i| \).

Measuring the risk of a financial portfolio is a crucial task in both the banking and insurance sectors, and it is typically done by calculating the value of a risk measure which maps the portfolio loss to a real number. A risk measure is a functional \( \rho : \mathcal{X}_\rho \to \mathbb{R} := [-\infty, \infty] \), where the domain \( \mathcal{X}_\rho \subseteq \mathcal{X} \) is a set of random variables representing financial losses. We will assume that an agent uses a risk measure \( \rho \) for their preference, in the sense that the agent prefers a smaller value of \( \rho \). Our notion of a risk measure is quite broad, and it includes not only risk measures in the sense of Artzner et al. (1999) and Föllmer and Schied (2016) but also decision models such as the expected utility by flipping the sign. However, we need to be clear that most classic expected utility models or convex risk measures (standard properties of risk measures are collected in Appendix A) in the literature are not suitable for our setting, because super-Pareto losses do not have a finite mean, and most expected utility functions and convex risk measures will take infinite values when evaluating these losses. Nevertheless, we will soon see that there are still many useful examples of risk measures conforming with our setting.

To interpret our main results, we only need minimal assumptions of monotonicity on \( \rho \), in the following two forms.

(a) Weak monotonicity: \( \rho(X) \leq \rho(Y) \) for \( X, Y \in \mathcal{X}_\rho \) if \( X \preceq_{st} Y \).

(b) Mild monotonicity: \( \rho \) is weakly monotone and \( \rho(X) < \rho(Y) \) if \( F_X^{-1} < F_Y^{-1} \) on \( (0, 1) \).
It is well-known that for any functional $\rho : \mathcal{X}_\rho \to \mathbb{R}$ that is law-invariant (i.e., $\rho(X) = \rho(Y)$ if $X \overset{d}{=} Y$), weak monotonicity is equivalent to the standard monotonicity (e.g., Theorem 6.28 in Shapiro et al. (2021)): $\rho(X) \leq \rho(Y)$ for $X, Y \in \mathcal{X}_\rho$ if $X \leq Y$. This property is satisfied by all commonly used risk measures, as well as expected utilities and hence they are weakly monotone. As such, stochastic dominance is a natural notion to consider for all preference models.

Many commonly used preference models are not just weakly monotone but mildly monotone; we highlight some examples. First, for an increasing utility function $u$, the expected utility agent can be represented by a risk measure $E_v$, namely

$$E_v(X) = \mathbb{E}[v(X)], \quad X \in \mathcal{X}_v := \{Y \in \mathcal{X} : \mathbb{E}[|v(Y)|] < \infty\},$$

where $v(x) = -u(-x)$ is also increasing. It is clear that $E_v$ is mildly monotone if $v$ or $u$ is strictly increasing. The next examples are the two widely used regulatory risk measures in insurance and finance, Value-at-Risk (VaR) and Expected Shortfall (ES). For $X \in \mathcal{X}$ and $p \in (0, 1)$, VaR is defined as

$$\text{VaR}_p(X) = F_X^{-1}(p) = \inf\{t \in \mathbb{R} : F_X(t) \geq p\},$$

and ES is defined as

$$\text{ES}_p(X) = \frac{1}{1-p} \int_p^1 \text{VaR}_u(X) du.$$  

For $X \notin L^1$, such as super-Pareto losses, $\text{ES}_p(X)$ can be $\infty$, whereas $\text{VaR}_p(X)$ is always finite. VaR is mildly monotone on $\mathcal{X}$, whereas ES is mildly monotone only on $L^1$.

In Theorem 1, we have established a diversification penalty for super-Pareto losses. We consider model A, where $Y_1, \ldots, Y_n$ are WNAID super-Pareto and $Y \overset{d}{=} Y_1$, and model B, where $Y = X_1 1_A$ and $Y_i = X_i 1_{A_i}$ for $i \in [n]$ in Theorem 1 (ii). From now on, we will assume that $\mathcal{X}_\rho$ contains the random variables in models A and B (this puts some restrictions on $v$ for $E_v$ since super-Pareto random variables do not have finite mean). The following result on the diversification penalty of super-Pareto losses for a monotone agent follows directly from Theorem 1.

**Proposition 3.** For $(\theta_1, \ldots, \theta_n) \in \Delta_n$ and a weakly monotone risk measure $\rho : \mathcal{X}_\rho \to \mathbb{R}$, for both models A and B, we have

$$\rho \left( \sum_{i=1}^n \theta_i Y_i \right) \geq \rho(Y).$$  

In model A, the inequality in (15) is strict if $\rho$ is mildly monotone and $\theta_i > 0$ for at least two $i \in [n]$. 

16
We distinguish strict and non-strict inequalities in (15) because a strict inequality has stronger implications on the optimal decision of an agent. As an important consequence of Proposition 3, for \( p \in (0, 1) \) and \((\theta_1, \ldots, \theta_n) \in \Delta^n\), in models A and B, 

\[
\text{VaR}_p \left( \sum_{i=1}^n \theta_i Y_i \right) \geq \text{VaR}_p(Y),
\]

and in model A, if \( \theta_i > 0 \) for at least two \( i \in [n] \),

\[
\text{VaR}_p \left( \sum_{i=1}^n \theta_i Y_i \right) > \sum_{i=1}^n \theta_i \text{VaR}_p(Y_i).
\]

Inequality (17) and its non-strict version will be referred to as the diversification penalty for \( \text{VaR}_p \).

Since all commonly used preference models are mildly monotone, Proposition 3 and (17) suggest that diversification of super-Pareto losses is detrimental for the agent.

Proposition 3 further leads to the following optimal decision for an agent in a market where several WNAID super-Pareto losses are present. Suppose that the agent needs to decide on a position \( w \in \mathbb{R}^+_n \) across these losses to minimize the total risk. The agent faces a total loss \( w \cdot Y - g(\|w\|) \) where the function \( g \) represents a compensation that depends on \( w \) through \( \|w\| \), and \( Y \) is as in model A or B. The agent’s optimization problem then becomes

\[
\text{to minimize } \rho(w \cdot Y - g(\|w\|)) \text{ subject to } w \in \mathbb{R}^+_n \text{ and } \|w\| = w \text{ with given } w > 0,
\]

or

\[
\text{to minimize } \rho(w \cdot Y - g(\|w\|)) \text{ subject to } w \in \mathbb{R}^+_n.
\]

For \( i \in [n] \), let \( e_{i,n} \) be the \( i \)th column vector of the \( n \times n \) identity matrix, and \( E_w = \{we_{i,n} : i \in [n]\} \) for \( w \geq 0 \), which represents the positions of taking only one loss with exposure \( w \).

**Proposition 4.** Let \( \rho : X_\rho \to \mathbb{R} \) be weakly monotone and \( g : \mathbb{R} \to \mathbb{R} \).

(i) For model A, if \( \rho \) is mildly monotone, then the set of minimizers of (18) is \( E_w \), and that of (19) is contained in \( \bigcup_{w \in \mathbb{R}^+_n} E_w \).

(ii) For models A and B, if (18) has an optimizer, then it has an optimizer in \( E_w \); if (19) has an optimizer, then it has an optimizer in \( \bigcup_{w \in \mathbb{R}^+_n} E_w \).

Remarkably, there are almost no restrictions on \( \rho \) and \( g \) in Proposition 4 other than monotonicity of \( \rho \), and hence this result can be applied to many economic decision models. We are now
able to answer the question raised in the Introduction: By Proposition 4, as long as the agent’s risk preference is monotone, an agent should not diversify in a risk sharing pool of extremely heavy-tailed losses.

**Remark 4.** Since $ES_p$ is $\infty$ for the losses in models A and B, Proposition 4 applied to $ES$ gives the trivial statement that every position has infinite risk. The main context of application for Proposition 4 should be risk measures which are finite for losses in models A and B, such as VaR, $E_v$ with some sublinear $v$, and Range Value-at-Risk (RVaR); see Appendix A for the definition of RVaR.

### 3.2 A model of excess-of-loss reinsurance coverage

Next, we assume the agent is an insurance company. In practice, insurers seek reinsurance coverage to transfer their losses. One of the most popular catastrophe reinsurance coverages is the excess-of-loss coverage; see OECD (2018). Therefore, it is important to consider heavy-tailed losses bounded at some thresholds. Catastrophe excess-of-loss coverage can be provided on per-loss or aggregate basis. We will see that the result in Proposition 3 holds if the excess-of-loss coverage is provided on either per-loss basis with high thresholds or aggregate basis.

We first discuss the case that the excess-of-loss coverage is provided on a per-loss basis, where non-diversification traps may exist for insurers; see Ibragimov et al. (2009). For WNAID super-Pareto losses $X_1, \ldots, X_n$, take $Y_i = X_i \wedge c_i$, where $c_i > z_X$ is the threshold, $i = 1, \ldots, n$. Note that each $Y_i$ is bounded. Since $Y_i$ has a finite mean, we cannot expect (16) or (17) to hold for all $p \in (0, 1)$. Nevertheless, we will see from the following result that for a given $p$ and large $c_1, \ldots, c_n$, (17) holds, and thus there exists a diversification penalty for $\text{VaR}_p$.

**Proposition 5.** Let $X, X_1, \ldots, X_n$ be WNAID super-Pareto random variables and $Y_i = X_i \wedge c_i$ where $c_i \geq z_X$ for each $i \in [n]$. Suppose that $(\theta_1, \ldots, \theta_n) \in \Delta_n$ such that $\theta_i > 0$ for at least two $i \in [n]$, and denote by $c = \min\{c_1\theta_1, \ldots, c_n\theta_n\}$. We have

$$
\mathbb{P} \left( \sum_{i=1}^n \theta_i Y_i > t \right) = \mathbb{P} \left( \sum_{i=1}^n \theta_i X_i > t \right) > \mathbb{P} \left( X > t \right) = \mathbb{P} \left( Y_i > t \right), \quad i \in [n]
$$

for $t \in (z_X, c]$, and $\text{VaR}_p \left( \sum_{i=1}^n \theta_i Y_i \right) > \sum_{i=1}^n \theta_i \text{VaR}_p(Y_i)$ for $p \in (0, \mathbb{P}(X \leq c))$.

Proposition 5 says that for a fixed weight vector $(\theta_1, \ldots, \theta_n)$ and a fixed threshold $t$, if the thresholds $c_1, \ldots, c_n$ are high enough, implying that $c$ is large, then $\mathbb{P} \left( \sum_{i=1}^n \theta_i Y_i > t \right)$ coincides with $\mathbb{P} \left( \sum_{i=1}^n \theta_i X_i > t \right)$, and (17) holds for $p \in (0, 1)$ not too large.
If the excess-of-loss coverage is provided on an aggregate basis, then stochastic dominance holds as \( X \land c \leq_{st} (\sum_{i=1}^{n} \theta_i X_i) \land c \) where \( c > z_X \) is the threshold; indeed the inequality is preserved under a monotone transform. Hence, for any weakly monotone risk measure \( \rho : \mathcal{X} \to \mathbb{R} \), we have \( \rho(X \land c) \leq \rho((\sum_{i=1}^{n} \theta_i X_i) \land c) \), and a diversification penalty exists for \( \rho \). Unlike the situation of model \( A \) in Proposition 3, strict inequality may not hold for \( \rho = \text{VaR}_p \) because \( X \land c \) and \( (\sum_{i=1}^{n} \theta_i X_i) \land c \) have the same \( p \)-quantile \( c \) for large \( p \). Nevertheless, for the expected utility preference \( E_v \), we have
\[
E[v(X \land c)] < E[v((\theta_1 X_1 + \cdots + \theta_n X_n) \land c)],
\]
for \( c > z_X \) and \( v \) strictly increasing on \([z_X, c]\). This is because \( E_v \) is strictly monotone in the sense that for \( X \leq_{st} Y \) taking values in \([z_X, c]\) and \( X \neq Y \), we have \( E_v(X) < E_v(Y) \).

4 Equilibrium analysis in a risk exchange economy

Suppose that there are \( n \geq 2 \) agents in a risk exchange market. Let \( X = (X_1, \ldots, X_n) \), where \( X_1, \ldots, X_n \) are WNAID super-Pareto random variables. The \( i \)th agent faces a loss \( a_i X_i \), where \( a_i > 0 \) is the initial exposure. In other words, the initial exposure vector of agent \( i \) is \( \mathbf{a}^i = a_i \mathbf{e}_{i,n} \), and the corresponding loss can be written as \( \mathbf{a}^i \cdot X = a_i X_i \).

In a risk exchange market, each agent decides whether and how to share the losses with the other agents. For \( i \in [n] \), let \( p_i \geq 0 \) be the premium (or compensation) for one unit of loss \( X_i \); that is, if an agent takes \( b \geq 0 \) units of loss \( X_i \), it receives the premium \( b p_i \), which is linear in \( b \). Denote by \( \mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n_+ \) the (endogenously generated) premium vector. Let \( \mathbf{w}^i \in \mathbb{R}^n_+ \) be the exposure vector of the \( i \)th agent from \( X \) after risk sharing. Then the loss of agent \( i \in [n] \) after risk sharing is
\[
L_i(\mathbf{w}^i, \mathbf{p}) = \mathbf{w}^i \cdot X - (\mathbf{w}^i - \mathbf{a}^i) \cdot \mathbf{p}.
\]

For each \( i \in [n] \), assume that agent \( i \) is equipped with a risk measure \( \rho_i : \mathcal{X} \to \mathbb{R} \), where \( \mathcal{X} \) contains the convex cone generated by \( \{X\} \cup \mathbb{R}^n \). Moreover, there is a cost associated with taking a total risk position \( \|\mathbf{w}^i\| \) different from the initial total exposure \( \|\mathbf{a}^i\| \). The cost is modelled by \( c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|) \), where \( c_i \) is a non-negative convex function satisfying \( c_i(0) = 0 \). Some examples of \( c_i \) are \( c_i(x) = 0 \) (no cost), \( c_i(x) = \lambda_i |x| \) (linear cost), \( c_i(x) = \lambda_i x^2 \) (quadratic cost), and \( c_i(x) = \lambda_i x_+ \) (cost only for excess risk taking), where \( \lambda_i > 0 \). We denote by \( c'_{i-}(x) \) and \( c'_{i+}(x) \) the left and right derivatives of \( c_i \) at \( x \in \mathbb{R} \), respectively.

The above setting is called a super-Pareto risk sharing market. In this risk sharing market, the goal of each agent is to choose an exposure vector so that their own risk is minimized, i.e.,
minimizing \( \rho_i(L_i(w^i, p)) + c_i(\|w^i\| - \|\alpha^i\|) \) over \( w^i \in \mathbb{R}_+^n, i \in [n] \). An \textit{equilibrium} of the market is a tuple \( (p^*, w^1, \ldots, w^n) \in (\mathbb{R}_+^n)^{n+1} \) if the following two conditions are satisfied.

(a) Individual optimality:

\[
\begin{align*}
 w^{i*} &\in \arg \min_{w^i \in \mathbb{R}_+^n} \{ \rho_i(L_i(w^i, p^*)) + c_i(\|w^i\| - \|\alpha^i\|) \}, \quad \text{for each } i \in [n]. \\
\end{align*}
\]  

(20)

(b) Market clearance:

\[
\sum_{i=1}^n w^{i*} = \sum_{i=1}^n \alpha^i. 
\]  

(21)

In this case, the vector \( p^* \) is an \textit{equilibrium price}, and \( (w^1, \ldots, w^n) \) is an \textit{equilibrium allocation}.

Some of our results rely on a popular class of risk measures, many of which can be applied to super-Pareto losses. A distortion risk measure is defined as \( \rho : \mathcal{X}_\rho \to \mathbb{R} \), via

\[
\rho(Y) = \int_{-\infty}^{0} (h(P(Y > x)) - 1)dx + \int_{0}^{\infty} h(P(Y > x))dx, 
\]  

(22)

where \( h : [0, 1] \to [0, 1] \), called the distortion function, is a nondecreasing function with \( h(0) = 0 \) and \( h(1) = 1 \). The distortion risk measure \( \rho \), up to sign change, coincides with the dual utility of Yaari (1987) in decision theory. As a class of risk measures, it includes VaR, ES, and RVaR as special cases, and almost all distortion risk measures are mildly monotone (see Proposition A.1).

We assume that \( \mathcal{X}_\rho \) contains the convex cone generated by \( \{X\} \cup \mathbb{R}^n \); this always holds in case \( \rho \) is VaR or RVaR, but it does not hold for \( \rho \) being ES as super-Pareto losses do not have finite mean.

As anticipated from Proposition 4, each agent's optimal strategy is to not share with the other agents if their risk measure is mildly monotone and the losses are super-Pareto. This observation is made rigorous in the following result, where we obtain a necessary condition for all possible equilibria in the market, as well as two different conditions in the case of distortion risk measures.

As before, let \( X \overset{d}{=} X_1 \).

**Theorem 2.** In a super-Pareto risk sharing market, suppose that \( \rho_1, \ldots, \rho_n \) are mildly monotone.

(i) All equilibria \( (p^*, w^1*, \ldots, w^n*) \) (if they exist) satisfy \( p^* = (p, \ldots, p) \) for some \( p \in \mathbb{R}_+ \) and \( (w^1, \ldots, w^n) \) is an \( n \)-permutation of \( (\alpha^1, \ldots, \alpha^n) \).

(ii) Suppose that \( \rho_1, \ldots, \rho_n \) are distortion risk measures on \( \mathcal{X} \). The tuple \( ((p, \ldots, p), (\alpha^1, \ldots, \alpha^n)) \) is an equilibrium if \( p \) satisfies

\[
c_{i+}'(0) \geq p - \rho_i (X) \geq c_{i-}'(0) \quad \text{for } i \in [n].
\]  

(23)
(iii) Suppose that $\rho_1, \ldots, \rho_n$ are distortion risk measures on $X$. If $(p, \ldots, p)$ is an equilibrium price, then
\[
\max_{j \in [n]} c'_{i+}(a_j - a_i) \geq p - \rho_i(X) \geq \min_{j \in [n]} c'_{i-}(a_j - a_i) \quad \text{for } i \in [n].
\] (24)

Theorem 2 (i) states that, even if there is some risk exchange in an equilibrium, the agents merely exchange positions entirely instead of sharing a pool. This observation is consistent with Theorem 1, which implies that diversification among multiple super-Pareto losses increases risk in a uniform sense. As there is no diversification in the optimal allocation for each agent, taking any of these WNAID losses is equivalent for the agent, and the equilibrium price should be identical across losses. Part (ii) suggests that if $c_i$ has a kink at 0, i.e., $c'_i(0^+)>0>c'_i(0^-)$, then $p$ can be an equilibrium price if it is very close to $\rho_i(X)$ in the sense of (23). Conversely, in part (iii), if $p$ is an equilibrium price, then it needs to be close to $\rho_i(X)$ for $i \in [n]$ in the sense of (24). This observation is quite intuitive because by (i), the agents will not share losses but rather keep one of them in an equilibrium. If the price of taking one unit of the loss is too far away from an agent’s assessment of the loss, it may have an incentive to move away, and the equilibrium is broken.

**Example 1** (Equilibrium for Pareto losses and VaR agents with no costs). Suppose that $c_i = 0$ for $i \in [n]$ and $X_1, \ldots, X_n \sim \text{Pareto}(\alpha), \alpha \in (0, 1]$. Let $\rho_i = \text{VaR}_q$, $q \in (0, 1)$, $i \in [n]$. The tuple $(p^*, w^{1*}, \ldots, w^{n*})$ is an equilibrium where $p^* = ((1-q)^{-1/\alpha}, \ldots, (1-q)^{-1/\alpha})$, and $(w^{1*}, \ldots, w^{n*})$ is an $n$-permutation of $(a^1, \ldots, a^n)$. For $i \in [n]$, $\rho_i(L_i(w^{i*}, p^*)) = \text{VaR}_q(a_iX) = a_i(1-q)^{-1/\alpha}$.

Further results on the super-Pareto risk sharing market are put in Appendix D. Section D.1 contains some further discussions on Theorem 2.

We present in Section D.2 a market with external agents. It turns out that in the presence of the external market, the picture is drastically different from the one in this section: All agents will benefit from transferring some losses to an external market; see Theorem A.1.

Although the agents will not benefit from losses with infinite mean like super-Pareto losses, the situation changes if the losses have finite mean, which will be discussed in Section D.3.

### 5 Models for diversification penalty

Given the many implications of (5), one naturally wonders whether and how it can be generalized beyond the WNAID super-Pareto model. That is, whether the diversification penalty
\[
X \leq_{st} \sum_{i=1}^n \theta_i X_i \quad \text{for all } (\theta_1, \ldots, \theta_n) \in \Delta_n, \text{ where } X, X_1, \ldots, X_n \text{ are identically distributed},
\] (25)
holds under models other than those covered in Theorem 1. We consider two directions of generalization, one on the marginal distributions and one on the dependence structure. Below, \( n \geq 2 \) is fixed. First, let

\[ \mathcal{F}_{\text{IN}} = \{ \text{distribution of } X : (25) \text{ holds for all independent } X_1, \ldots, X_n \}. \]

The sets \( \mathcal{F}_{\text{WNA}} \) and \( \mathcal{F}_{\text{NA}} \) are defined similarly, where the subscript (W)NA means (weak) negative association, instead of independence, among \( X_1, \ldots, X_n \). Clearly, \( \mathcal{F}_{\text{WNA}} \subseteq \mathcal{F}_{\text{NA}} \subseteq \mathcal{F}_{\text{IN}} \), and each of them contains all super-Pareto distributions by Theorem 1.

**Proposition 6.** The set \( \mathcal{F}_{\text{IN}} \) is closed under convolution. All of \( \mathcal{F}_{\text{WNA}}, \mathcal{F}_{\text{NA}}, \) and \( \mathcal{F}_{\text{IN}} \) are closed under strictly increasing convex transforms.

Second, we consider possible dependence structures for (25). Copulas are useful tools for modeling dependence structures; see Nelsen (2006) for an overview. A **copula** is a distribution function with standard uniform (i.e., on \([0, 1]\)) marginal distributions. For a random vector \( X \) with distribution function \( F \) and marginal distributions \( F_1, \ldots, F_n \), by Sklar’s Theorem (e.g., Theorem 7.3 of McNeil et al. (2015)), there exists a copula \( C \) satisfying \( F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \) for \( (x_1, \ldots, x_n) \in \mathbb{R}^n \), and such \( C \) is called a copula of \( X \), which is unique when \( X \) has continuous marginal distributions. The **copula for independence** and the **copula for comonotonicity** are given by \( u \mapsto \prod_{i=1}^n u_i \) and \( u \mapsto \min(u) \) for \( u = (u_1, \ldots, u_n) \in [0, 1]^n \), respectively. A **copula for weak negative association** is any copula of weakly negatively associated standard uniform random variables. As their names suggest, these copulas represent the corresponding dependence structures. The set of dependence structures satisfying the diversification penalty is then represented by

\[ \mathcal{C}_{\text{DP}} = \{ \text{copula } C : (25) \text{ holds for all super-Pareto } X_1, \ldots, X_n \text{ with copula } C \}. \]

**Proposition 7.** The set \( \mathcal{C}_{\text{DP}} \) is closed under mixture, and it contains all copulas for comonotonicity, independence, and weak negative association.

By Proposition 7, (25) holds under some particular forms of positive or mixed dependence, in addition to negative dependence as studied throughout the paper. Nevertheless, we did not find a natural model of positive dependence, other than a mixture of independence and comonotonicity, that yields (25).
6 Numerical examples

6.1 Diversification effects as \( n \) increases

For an equally weighted pool of \( k \) iid super-Pareto losses, we wonder whether enlarging \( k \) increases the risk in first-order stochastic dominance, that is, for iid super-Pareto random variables \( X_1, \ldots, X_\ell \), whether it holds that

\[
\frac{1}{k} \sum_{i=1}^{k} X_i \leq_{st} \frac{1}{\ell} \sum_{i=1}^{\ell} X_i \quad \text{for } k, \ell \in \mathbb{N} \text{ and } k \leq \ell. \tag{26}
\]

The case of \( k = 1 \) in (26) corresponds to (5) with equal weights \( \theta_1, \ldots, \theta_n \). Here, independence is assumed so that the dependence structure is the same across dimensions. Inequality (26) means that the more we diversify super-Pareto losses, the higher the penalty. The next result gives this inequality for the case that \( \ell \) is a multiple of \( k \).

**Proposition 8.** For \( m, n \in \mathbb{N} \), let \( X_1, \ldots, X_{mn} \) be iid super-Pareto random variables. We have

\[
\frac{1}{m} \sum_{i=1}^{m} X_i \leq_{st} \frac{1}{mn} \sum_{i=1}^{mn} X_i.
\]

For \( \alpha \in (0, 1] \) and \( p \in (0, 1) \), we compute \( \text{VaR}_p(\sum_{i=1}^{n} Y_i/n) \) for \( n = 2, \ldots, 6 \), where \( Y_1, \ldots, Y_n \) are iid Pareto(\( \alpha \)). From Figure 1, we observe that \( \text{VaR}_p(\sum_{i=1}^{n} Y_i/n) \) increases as \( n \) increases. The difference between the curves for different \( n \) becomes more pronounced as \( \alpha \) becomes smaller, i.e., the tail of the Pareto losses becomes heavier. From these numerical results, we may expect (26) to hold, although we were only able to show the case where \( \ell \) is a multiple of \( k \) in Proposition 8.

![Figure 1: VaR_p((Y_1 + \cdots + Y_n)/n) for n = 2, \ldots, 6 and p \in (0.9, 0.96).](image-url)
6.2 Examples of extremely heavy-tailed Pareto losses

In addition to the many examples mentioned in Section 1.1, we provide two further data examples: the first one on marine losses, and the second one on suppression costs of wildfires. The marine losses dataset, from the insurance data repository CASdatasets, was originally collected by a French private insurer and comprises 1,274 marine losses (paid) between January 2003 and June 2006. The wildfire dataset contains 10,915 suppression costs in Alberta, Canada from 1983 to 1995. For the purpose of this section, we only provide the Hill estimates of these two datasets although a more detailed EVT analysis is available (see McNeil et al. (2015)). The tail indices $\alpha$ for the marine losses and wildfire suppression costs are estimated as 0.916 and 0.847 with 95% confidence intervals being $(0.674, 1.158)$ and $(0.776, 0.918)$, respectively; the Hill plots are contained in Appendix E. Thus, these losses/costs have infinite mean if they follow Pareto distributions in their tails regions.

These observations suggest that the two loss datasets may have similar tail parameters. Applying Theorem 1, if two loss random variables $X_1$ and $X_2$ are independent and follow generalized Pareto distributions with the same tail parameter $\alpha = 1/\xi < 1$ (see (2)), then, for all $p \in (0, 1)$,

$$\text{VaR}_p(X_1 + X_2) > \text{VaR}_p(X_1) + \text{VaR}_p(X_2).$$

(27)

Even if $X_1$ and $X_2$ are not Pareto distributed, as long as their tails are Pareto, (27) may hold for $p$ relatively large, as suggested by Proposition A.2 in Appendix C.

We will verify (27) on our datasets to show how the implication of our main results holds for real data; a detailed analysis is presented in Appendix E. Let $\hat{F}_1$ be the empirical distribution of the marine losses (renormalized\(^4\)) and $\hat{F}_2$ be the empirical distribution of the wildfire suppression costs. Take independent random variables $\hat{Y}_1 \sim \hat{F}_1$ and $\hat{Y}_2 \sim \hat{F}_2$. We observe from Figure 2 that

$$\text{VaR}_p(\hat{Y}_1 + \hat{Y}_2) > \text{VaR}_p(\hat{Y}_1) + \text{VaR}_p(\hat{Y}_2)$$

(28)

holds unless $p$ is greater than 0.9847 (marked by the vertical line in Figure 2). Since the quantiles are directly computed from data, thus from distributions with bounded supports, for $p$ close enough to 1 it must hold that $\text{VaR}_p(\hat{Y}_1 + \hat{Y}_2) \leq \text{VaR}_p(\hat{Y}_1) + \text{VaR}_p(\hat{Y}_2)$; see the similar observation made in Proposition 2. Nevertheless, we observe (28) for most values of $p \in (0, 1)$. Note that the observation

\(^2\)Available at http://cas.uqam.ca/.

\(^3\)See https://wildfire.alberta.ca/resources/historical-data/historical-wildfire-database.aspx.

\(^4\)The data are multiplied by 500. This normalization does not matter for (27) and is made only for better visualization in our analysis.
of (28) is entirely empirical and it does not use the fitted models.

### 6.3 Aggregation of Pareto risks with different parameters

As mentioned above, for independent losses $Y_1, \ldots, Y_n$ following generalized Pareto distributions with the same tail parameter $\alpha = 1/\xi < 1$, it holds that

$$\sum_{i=1}^{n} \text{VaR}_p(Y_i) \leq \text{VaR}_p \left( \sum_{i=1}^{n} Y_i \right),$$

usually with strict inequality.  \hfill (29)

Inspired by the results in Section 6.2, we are interested in whether (29) holds for losses following generalized Pareto distributions with different parameters. To make a first attempt on this problem, we look at the 6 operational losses of different business lines with infinite mean in Table 5 of Moscadelli (2004), where the operational losses are assumed to follow generalized Pareto distributions. Denote by $Y_1, \ldots, Y_6$ the operational losses corresponding to these 6 generalized Pareto distributions. The estimated parameters in Moscadelli (2004) for these losses are presented in Table 1; they all have infinite mean.

| $i$ | 1  | 2  | 3  | 4   | 5    | 6    |
|-----|----|----|----|-----|------|------|
| $\xi_i$ | 1.19 | 1.17 | 1.01 | 1.39 | 1.23  | 1.22  |
| $\beta_i$ | 774  | 254  | 233  | 412  | 107   | 243   |

**Table 1:** The estimated parameters $\xi_i$ and $\beta_i$, $i \in [6]$.

For the purpose of this numerical example, we assume that $Y_1, \ldots, Y_6$ are independent and plot $\sum_{i=1}^{6} \text{VaR}_p(Y_i)$ and $\text{VaR}_p(\sum_{i=1}^{6} Y_i)$ for $p \in (0.95, 0.99)$ in Figure 3. We can see that $\text{VaR}_p(\sum_{i=1}^{6} Y_i)$
is larger than $\sum_{i=1}^{6} \text{VAR}_p(Y_i)$, and the gap between the two values gets larger as the level $p$ approaches 1. This observation further suggests that even if the extremely heavy-tailed Pareto losses have different tail parameters, a diversification penalty may still exist. We conjecture that this is true for any generalized Pareto losses $Y_1, \ldots, Y_n$ with shape parameters $\xi_1, \ldots, \xi_n \in [1, \infty)$, although we do not have a proof. Similarly, we may expect that $\sum_{i=1}^{n} \theta_i \text{VAR}_p(X_i) \leq \text{VAR}_p(\sum_{i=1}^{n} \theta_i X_i)$ holds for any Pareto losses $X_1, \ldots, X_n$ with tail parameters $\alpha_1, \ldots, \alpha_n \in (0, 1]$.

![Figure 3: Curves of $\text{VAR}_p(\sum_{i=1}^{n} Y_i)$ and $\sum_{i=1}^{n} \text{VAR}_p(Y_i)$ for $n = 6$ generalized Pareto losses with parameters in Table 1 and $p \in (0.95, 0.99)$.](image)

From a risk management point of view, the message from Sections 6.2 and 6.3 is clear. If a careful statistical analysis leads to statistical models in the realm of infinite means, then the risk manager at the helm should take a step back and question to what extent classical diversification arguments can be applied. Though we mathematically analyzed the case of identically distributed losses, we conjecture that these results hold more widely in the heterogeneous case. As a consequence, it is advised to hold on to only one such super-Pareto risk. Of course, the discussion concerning the practical relevance of infinite mean models remains. When such underlying models are methodologically possible, then one should think carefully about the applicability of standard risk management arguments; this brings us back to Weitzman's Dismal Theorem as discussed towards the end of Section 1. From a methodological point of view, we expect that the results from Sections 3 and 4 carry over to the above heterogeneous setting.
7 Concluding remarks

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. Our results provide an important implication in risk management, answering the question in the Introduction. That is, the diversification of many catastrophic losses without finite mean increases the risk assessment of a portfolio, uniformly for all commonly used risk preferences.

The equilibrium of a risk exchange model is analyzed in Theorem 2, where agents can take extra super-Pareto losses with compensations. In particular, if every agent is associated with an initial position of a super-Pareto loss, the agents can merely exchange their entire position with each other. On the other hand, if some external agents are not associated with any initial losses, it is possible that all agents can reduce their risks by transferring the losses from the agents with initial losses to those without initial losses (Theorem A.1 in Appendix D).

In the case of Pareto($\alpha$) distributions, our main results rely on that $\alpha \leq 1$ is known. In practice, the estimation of $\alpha$ is not an easy task. If $\alpha$ is estimated to be close to 1, then we expect that a diversification penalty holds for VaR$_p$ at most (but not all) levels $p \in (0, 1)$ if the true (unknown) value of $\alpha$ is slightly larger than 1.

The diversification effects are investigated by numerical studies where two open technical questions arise. The first question is whether (26) holds, that is,

$$\frac{1}{k} \sum_{i=1}^{k} X_i \leq_{\text{st}} \frac{1}{\ell} \sum_{i=1}^{\ell} X_i,$$

holds for all $k, \ell \in \mathbb{N}$ such that $k \leq \ell$, where $X_1, \ldots, X_l$ are iid super-Pareto losses. The statement is true if $\ell$ is a multiple of $k$, as shown in Proposition 8. The second question is whether

$$\text{VaR}_p \left( \sum_{i=1}^{n} \theta_iX_i \right) \geq \sum_{i=1}^{n} \theta_i \text{VaR}_p (X_i)$$

holds for $(\theta_1, \ldots, \theta_n) \in \Delta_n$ and independent extremely heavy-tailed Pareto losses $X_1, \ldots, X_n$ with possibly different tail parameters. From the numerical results in Section 6, both (30) and (31) are anticipated to hold; a proof seems to be beyond the current techniques.

We conclude the paper with a conjecture stronger than (30). For two vectors $(\theta_1, \ldots, \theta_n)$ and $(\eta_1, \ldots, \eta_n)$ in $\mathbb{R}^n$, we say that $(\theta_1, \ldots, \theta_n)$ is dominated by $(\eta_1, \ldots, \eta_n)$ in majorization order, denoted by $(\theta_1, \ldots, \theta_n) \preceq_{\text{m}} (\eta_1, \ldots, \eta_n)$, if $\sum_{i=1}^{n} \phi(\theta_i) \leq \sum_{i=1}^{n} \phi(\eta_i)$ for all continuous and convex
functions $\phi$; see Marshall et al. (2011) for an introduction to majorization order. If $(\theta_1, \ldots, \theta_n) \leq_m (\eta_1, \ldots, \eta_n)$, the components of $(\theta_1, \ldots, \theta_n)$ are less spread out than those of $(\eta_1, \ldots, \eta_n)$; e.g., $(1/3, 1/3, 1/3) \leq_m (1/2, 1/2, 0) \leq_m (1, 0, 0)$. Therefore, it is natural to regard a portfolio with a smaller exposure vector in majorization order as more diversified. Let $(\theta_1, \ldots, \theta_n) \in \mathbb{R}_+^n$ and $(\eta_1, \ldots, \eta_n) \in \mathbb{R}_+^n$ such that $(\theta_1, \ldots, \theta_n) \leq_m (\eta_1, \ldots, \eta_n)$. Proschan (1965) showed that for iid random variables $Z_1, \ldots, Z_n$ following symmetric log-concave distributions,

$$\mathbb{P} \left( \sum_{i=1}^n \theta_i Z_i > t \right) \leq \mathbb{P} \left( \sum_{i=1}^n \eta_i Z_i > t \right) \text{ for } t \geq 0.$$ \hfill (32)

It is shown by Ibragimov (2005) that (32) continues to hold for iid stable random variables with finite mean and it flips if the stable random variables have infinite mean. In particular, by Theorem 1.2.4 of Ibragimov (2005), for iid infinite-mean one-sided positive stable random variables $Y_1, \ldots, Y_n$,

$$\sum_{i=1}^n \eta_i Y_i \leq_{st} \sum_{i=1}^n \theta_i Y_i.$$ \hfill (33)

Some applications of (33) are studied by Ibragimov and Walden (2010) and Ibragimov et al. (2015). Inspired by Theorem 1 and Proposition 8, we conjecture that (33) also holds for iid super-Pareto risks; note that this is stronger than (30). We leave this question for future research.

Acknowledgements

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).

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.

Axtell, R. L. (2001). Zipf distribution of U.S. firm sizes. *Science*, 293:1818–1820.

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

Barrett, G. F. and Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, 71(1):71–104.

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*, 71(1):71–104.

Beirlant, J., Dierckx, G., Goegebeur, Y. and Matthys, G. (1999). Tail index estimation and an exponential regression model. *Extremes*, 2(2):177–200.

Biffis, E. and Chavez, E. (2014). Tail risk in commercial property insurance. *Risks*, 2(4):393–410.

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. Available at SSRN: https://ssrn.com/abstract=4292259.

Cirillo, P. and Taleb, N. N. (2020). Tail risk of contagious diseases. *Nature Physics*, 16(6):606–613.

Clark, D. R. (2013). A note on the upper-truncated Pareto distribution. *Casualty Actuarial Society E-Forum*, Winter, 2013, Volume 1, pp. 1–22.

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

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.

Eling, M. and Schnell, W. (2020). Capital requirements for cyber risk and cyber risk insurance: An analysis of Solvency II, the US risk-based capital standards, and the Swiss Solvency Test. *North American Actuarial Journal*, 24(3):370–392.

Eling, M. and Wirfs, J. (2019). What are the actual costs of cyber risk events? *European Journal of Operational Research*, 272(3):1109–1119.

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., Liu, H. and Wang, R. (2018). Quantile-based risk sharing. *Operations Research*, 66(4):936–949.

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.

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.

FINMA (2021). Standardmodell Versicherungen (Standard Model Insurance): Technical description for the SST standard model non-life insurance (in German), October 31, 2021, [www.finma.ch](http://www.finma.ch).

Flyvbjerg, B., Budzier, A., Lee, J. S., Keil, M., Lunn, D. and Bester, D. W. (2022). The empirical reality of IT project cost overruns: Discovering a power-law distribution. *Journal of Management Information Systems*, 39(3):607–639.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447.

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

Gabaix, X. (1999). Zipf’s law and the growth of cities. *American Economic Review*, 89(2):129–132.

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

Gabaix, X. and Ibragimov, R. (2011). Rank-1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29(1):24–39.

Guan, Y., Jiao, Z. and Wang, R. (2022). A reverse ES (CVaR) optimization formula. *arXiv*: 2203.02599.

Hadar, J. and Russell, W. R. (1969). Rules for ordering uncertain prospects. *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, 11*(1):286–295.

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

Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. 4th Edition. John Wiley & Sons.

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.

Moscadelli, M. (2004). The modelling of operational risk: Experience with the analysis of the data collected by the Basel committee. Technical Report 517. *SSRN: 557214*.

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.

Nešlehová, J., Embrechts, P. and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for
operational risk. *Journal of Operational Risk, 1*(1):3–25.

Nordhaus, W. D. (2009). An analysis of the Dismal Theorem. Yale University: Cowles Foundation Discussion Paper 1686.

Nelsen, R. (2006). *An Introduction to Copulas*. Second Edition. Springer, New York.

OECD (2018). *The Contribution of Reinsurance Markets to Managing Catastrophe Risk*. Available at www.oecd.org/finance/the-contribution-of-reinsurance-markets-to-managing-catastrophe-risk.pdf.

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.

Rizzo, M. L. (2009). New goodness-of-fit tests for Pareto distributions. *ASTIN Bulletin*, 39(2):691–715.

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, A., Dentcheva, D. and Ruszczynski, A. (2021). *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.

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

Silverberg, G. and Verspagen, B. (2007). The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. *Journal of Econometrics*, 139(2):318–339.

Sornette, D., Maillart, T. and Kröger, W. (2013). Exploring the limits of safety analysis in complex technological systems. *International Journal of Disaster Risk Reduction*, 6:59–66.

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

Tasche, D. (2000). Conditional expectation as quantile derivative. *arXiv: 0104190*.

Wang, Q., Wang, R. and Wei, Y. (2020). Distortion riskmetrics on general spaces. *ASTIN Bulletin*, 50(3):827–851.

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

Weitzman, M. L. (2009). On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics*, 91(1):1-19.

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

32
Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, 55(1):95–115.
A Background on risk measures

Recall that $\mathcal{X}_\rho$ is a convex cone of random variables representing losses faced by financial institutions. We first present commonly used properties of a risk measure $\rho : \mathcal{X}_\rho \rightarrow \mathbb{R}$:

(c) Translation invariance: $\rho(X + c) = \rho(X) + c$ for $c \in \mathbb{R}$.

(d) Positive homogeneity: $\rho(aX) = a\rho(X)$ for $a \geq 0$.

(e) Convexity: $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda)\rho(Y)$ for $X, Y \in \mathcal{X}_\rho$ and $\lambda \in [0, 1]$.

A risk measure that satisfies (a) weak monotonicity, (c) translation invariance, and (e) convexity is a **convex risk measure** (Föllmer and Schied, 2002). It is well-known that ES is a convex risk measure. The convexity property means that diversification will not increase the risk of the loss portfolio, i.e., the risk of $\lambda X + (1 - \lambda)Y$ is less than or equal to that of the weighted average of individual losses. However, the canonical space for law-invariant convex risk measures is $L^1$ (see Filipović and Svindland (2012)) and hence convex risk measures are not useful for losses without finite mean.

For losses without finite mean, such as extremely heavy-tailed Pareto losses, it is natural to consider VaR or Range Value-at-Risk (RVaR), which includes VaR as a limiting case. For $X \in \mathcal{X}$ and $0 \leq p < q < 1$, the RVaR is defined as

$$RVaR_{p,q}(X) = \frac{1}{q - p} \int_p^q \text{VaR}_u(X) du.$$ 

For $p \in (0, 1)$, $\lim_{q \downarrow p^+} RVaR_{p,q}(X) = \text{VaR}_p(X)$. The class of RVaR is proposed by Cont et al. (2010) as robust risk measures; see Embrechts et al. (2018) for its properties and risk sharing results. VaR, ES and RVaR, as well as essential infimum (ess-inf) and essential supremum (ess-sup), belong to the family of distortion risk measures as defined in (22). For $X \in \mathcal{X}$, ess-inf and ess-sup are defined as

$$\text{ess-inf}(X) = \sup \{ x : F_X(x) = 0 \} \quad \text{and} \quad \text{ess-sup}(X) = \inf \{ x : F_X(x) = 1 \}.$$ 

The distortion functions of ess-inf and ess-sup are given as $h(t) = \mathbb{1}_{\{t = 1\}}$ and $h(t) = \mathbb{1}_{\{0 < t \leq 1\}}$, $t \in [0, 1]$, respectively; see Table 1 of Wang et al. (2020). Distortion risk measures satisfy (a), (c) and (d). Almost all the useful distortion risk measures are mildly monotone, as shown by the following proposition.
Proposition A.1. Any distortion risk measure is mildly monotone unless it is a mixture of ess-sup and ess-inf.

Proof. Let $\rho_h$ be a distortion risk measure with distortion function $h$. Suppose that $\rho_h$ is not mildly monotone. Then there exist $X, Y \in \mathcal{X}$ satisfying $F_X^{-1}(p) < F_Y^{-1}(p)$ for all $p \in (0, 1)$ and $\rho(X) = \rho(Y)$. Suppose that there exist $b \in (0, 1)$ such that $h(1-a) < h(1-b)$ for all $a > b$. For $x \in (F_X^{-1}(b), F_Y^{-1}(b))$, we have $F_X(x) \geq b > F_Y(x)$; see e.g., Lemma 1 of Guan et al. (2022). Hence, we have $h(1 - F_X(x)) \leq h(1 - b) < h(1 - F_Y(x))$ for $x \in (F_X^{-1}(b), F_Y^{-1}(b))$. Since $h(1 - F_X(x)) - h(1 - F_Y(x)) \leq 0$ for all $x \in \mathbb{R}$, by (22) we get

$$
\rho(X) - \rho(Y) = \int_{-\infty}^{\infty} (h(1 - F_X(x)) - h(1 - F_Y(x))) \, dx < 0.
$$

This contradicts $\rho(X) = \rho(Y)$. Hence, there is no $b \in (0, 1)$ such that $h(1-a) < h(1-b)$ for all $a > b$. Using a similar argument with the left quantiles replaced by right quantiles, we conclude that there is no $b \in (0, 1)$ such that $h(1-a) > h(1-b)$ for all $a < b$. Therefore, for every $b \in (0, 1)$, there exists an open interval $I_b$ such that $b \in I_b$ and $h$ is constant on $I_b$. For any $\epsilon > 0$, the interval $[\epsilon, 1-\epsilon]$ is compact. Hence, there exists a finite collection $\{I_b : b \in B\}$ which covers $[\epsilon, 1-\epsilon]$. Since the open intervals in $\{I_b : b \in B\}$ overlap, we know that $h$ is constant on $[\epsilon, 1-\epsilon]$. Letting $\epsilon \downarrow 0$ yields that $h$ takes a constant value on $(0,1)$, denoted by $\lambda \in [0,1]$. Together with $h(0) = 0$ and $h(1) = 1$, we get that $h(t) = \lambda 1_{\{0 < t \leq 1\}} + (1 - \lambda) 1_{\{t = 1\}}$ for $t \in [0,1]$, which is the distortion function of $\rho_h = \lambda \text{ess-inf} + (1 - \lambda) \text{ess-sup}$. \qed

As a consequence, for any set $\mathcal{X}$ containing a random variable unbounded from above and one unbounded from below, such as the $L^q$-space for $q \in [0, \infty)$, a real-valued distortion risk measure on $\mathcal{X}$ is mildly monotone.

B Proofs of technical results

Proof of Proposition 1. For the “$\Rightarrow$” direction, let $\mathbb{P}(X \leq x) = 1 - 1/g(x)$ 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 $1 \leq y \leq g(z_X)$. It is straightforward to see that for any Pareto(1) random variable $Y$, $f(Y) \overset{d}{=} X$. Next, we show the “$\Leftarrow$” direction. For $x < \infty$, the right-continuous generalized inverse of $f$ is $f^{-1+}(x) = \inf\{t : f(t) > x\}$. For $x \geq f(1)$, $\mathbb{P}(f(Y) \leq x) = \mathbb{P}(Y \leq f^{-1+}(x)) = 1 - 1/f^{-1+}(x)$. As $f$ is increasing, convex, and non-constant, $f^{-1+}$ is strictly increasing and concave. Hence, $f(Y)$ is super-Pareto with essential infimum $f(1)$. The statement on infinite mean follows because a strictly increasing convex function $f$ satisfies, for some $a > 0$ and $b \in \mathbb{R}$, $f(x) \geq ax + b$ for all $x \in \mathbb{R}$. \qed
Proof of Proposition 2. Note that (1) implies that $\text{ES}_p(X) \leq \text{ES}_p(\sum_{i=1}^n \theta_i X_i)$ for all $p \in (0, 1)$, where $\text{ES}_p$ is defined in Section 3. Since $\text{ES}_p$ is convex and $X_1, \ldots, X_n$ are identically distributed, we have

$$\text{ES}_p(X) \leq \text{ES}_p \left( \sum_{i=1}^n \theta_i X_i \right) \leq \theta_1 \sum_{i=1}^n \text{ES}_p(X_i) = \text{ES}_p(X), \quad p \in (0, 1).$$

Using positive homogeneity of $\text{ES}_p$, it follows that the equality $\sum_{i=1}^n \text{ES}_p(\theta_i X_i) = \text{ES}_p(\sum_{i=1}^n \theta_i X_i)$ holds for each $p \in (0, 1)$. By Theorem 5 of Wang and Zitikis (2021), this implies that $(\theta_1 X_1, \ldots, \theta_n X_n)$ is $p$-concentrated for each $p$; this result requires $X_1, \ldots, X_n$ to have finite mean. Using Theorem 3 of Wang and Zitikis (2021), the above condition implies that $(X_1, \ldots, X_n)$ is comonotonic. For definitions of comonotonicity and $p$-concentration, see Wang and Zitikis (2021). Since $X_1, \ldots, X_n$ are identically distributed, comonotonicity further implies that $X_1 = \cdots = X_n$ 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 $q = \mathbb{P}(X = z)$. 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 $F_{X_i}^{-1}(W_i) = X_i$ 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 $A_i = \{X_i \leq z\}$ and

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

where $V_1, \ldots, V_n$ are iid standard uniform random variables independent of $(X_1, \ldots, X_n)$. Note that for each $i \in [n]$, $U_i$ is also a uniform transform of $X_i$, and the distribution of $(U_1, \ldots, U_n)$ is one possible copula of $(X_1, \ldots, X_n)$. Let $Y_i = F_{Y_i}^{-1}(U_i)$ for $i \in [n]$. Note that $F_{X_i}^{-1} = f \circ F_{Y_i}^{-1}$ 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, $(X_1, \ldots, X_n) \overset{d}{=} (f(Y_1), \ldots, f(Y_n))$ as they have the same copula and marginal distributions. It remains to show that $Y_1, \ldots, Y_n$ 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

$$\mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A \mid Y_n \leq x) = \mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A, Y_n \leq x) \beta$$

$$= \mathbb{E}[\mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A, Y_n \leq x \mid V_1, \ldots, V_n)] \beta$$

$$\leq \mathbb{E} [\mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A \mid V_1, \ldots, V_n) \mathbb{P}(Y_n \leq x \mid V_1, \ldots, V_n)] \beta$$

$$= \mathbb{E}[\mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A \mid V_1, \ldots, V_{n-1}) \mathbb{P}(Y_n \leq x \mid V_n)] \beta$$

$$= \mathbb{E}[\mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A \mid V_1, \ldots, V_{n-1}) \mathbb{E}[\mathbb{P}(Y_n \leq x \mid V_n)] \beta$$

$$= \mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A) \mathbb{P}(Y_n \leq x) \beta = \mathbb{P}((Y_1, \ldots, Y_{n-1}) \in A),$$

36
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 \( X_1, \ldots, X_n \) are weakly negatively associated. Therefore, \( Y_1, \ldots, Y_n \) are also weakly negatively associated.

**Proof of Proposition 3.** This proposition follows directly from Theorem 1.

**Proof of Proposition 4.** The proof of (i) follows directly from Theorem 1. Statement (ii) follows from Theorem 1 by noting that there exists \( j \in [n] \) such that \( \mathbb{P}(A_j) \leq \mathbb{P}(A) \), and hence,

\[
w X_j \mathbb{I}_{A_j} \leq_{st} n \sum_{i=1}^n w_i X_i \mathbb{I}_{A_i},
\]

where \( X \) and \( A \) are as in model B.

**Proof of Proposition 5.** For \( t \in (z_X, c] \), we have

\[
\mathbb{P} \left( \sum_{i=1}^n \theta_i Y_i \leq t \right) = \mathbb{P} \left( \sum_{i=1}^n \theta_i (X_i \land c_i) \leq t \right) = \mathbb{P} \left( \sum_{i=1}^n \theta_i X_i \leq t \right).
\]

We also have \( \mathbb{P}(Y_i \leq t) = \mathbb{P}(X_i \land c_i \leq t) = \mathbb{P}(X_i \leq t), i \in [n] \). By the strictness statement in Theorem 1, we obtain the probability inequality. To show the quantile inequality, note that for \( p \in (0, \mathbb{P}(X \leq c)) \), we have \( \text{VaR}_p(X) < c \). Using Theorem 1 and the definition of \( Y_1, \ldots, Y_n \), we get

\[
\sum_{i=1}^n \theta_i \text{VaR}_p(Y_i) \leq \text{VaR}_p(X) < \text{VaR}_p \left( \sum_{i=1}^n \theta_i X_i \right) \land c
\]

\[
\leq \text{VaR}_p \left( \left( \sum_{i=1}^n \theta_i X_i \right) \land c \right) \leq \text{VaR}_p \left( \sum_{i=1}^n \theta_i Y_i \right).
\]

This gives the desired inequality.

**Proof of Theorem 2.** (i) Suppose that \( (p^*, w^1, \ldots, w^n) \) forms an equilibrium. We let \( p = \max_{j \in [n]} \{p_j\} \) and \( S = \arg \max_{j \in [n]} \{p_j\} \). For the \( i \)th agent, by writing \( w = \|w^i\| \), using Theorem 1 and the fact that \( \rho_i \) is mildly monotone, we have that for any \( w^i \in [0, 1]^n \),

\[
\rho_i(L_i(w^i, p^*)) = \rho_i(w^i \cdot (X - p^*) + a^i \cdot p^*)
\]

\[
\geq \rho_i(w^i \cdot X - wp + a^i \cdot p^*) \geq \rho_i(w X_1 - wp + a^i \cdot p^*).
\]

By the last statement of Theorem 1, the last inequality is strict if \( w^i \) contains at least two non-zero components. Moreover, \( c(\|w^i\| - \|a^i\|) = c(w - \|a^i\|) \). Therefore, we know that the
optimization \( \mathbf{w}^* = (w_1^*, \ldots, w_n^*) \) to (20) has at most one non-zero component \( w_j^* \), and \( j \in S \). Hence, \( w_k^* = 0 \) if \( k \in [n] \setminus S \) and this holds for each \( i \in [n] \). Using \( \sum_{i=1}^{n} w_i^* = \sum_{i=1}^{n} a_i^\prime \), which have all positive components, we know that \( S = [n] \), which further implies that \( \mathbf{p}^* = (p, \ldots, p) \) for \( p \in \mathbb{R}_+^n \). Next, as each \( \mathbf{w}^* \) has only one positive component, \( (\mathbf{w}_1^*, \ldots, \mathbf{w}_n^*) \) has to be an \( n \)-permutation of \( (a_1^\prime, \ldots, a_n^\prime) \) in order to satisfy the clearance condition (21).

(ii) The clearance condition (21) is clearly satisfied. Note that distortion risk measures are translation invariant and positive homogeneous (see Appendix A for properties of risk measures). Using these two properties and Proposition 4, for \( i \in [n] \),

\[
\min_{\mathbf{w}^i \in \mathbb{R}_+^n} \{ \rho_i (L_i(\mathbf{w}^i, \mathbf{p}^*)) + c_i(\|\mathbf{w}^i\| - \|a_i^\prime\|) \} \\
= \min_{\mathbf{w}^i \in \mathbb{R}_+^n} \{ \rho_i (\mathbf{w}^i \cdot \mathbf{X} - (\mathbf{w}^i - a_i^\prime) \cdot \mathbf{p}^*) + c_i(\|\mathbf{w}^i\| - \|a_i^\prime\|) \} \\
= \min_{\|\mathbf{w}^i\| \in \mathbb{R}_+^n} \{ (\rho_i (\|\mathbf{w}^i\|) - (\|\mathbf{w}^i\| - a_i) p) + c_i(\|\mathbf{w}^i\| - \|a_i^\prime\|) \} \\
= \min_{\mathbf{w} \in \mathbb{R}_+^n} \{ w(\rho_i(\mathbf{X}) - p) + a_i p + c_i(w - a_i) \}.
\]  

(A.1)

Note that \( w \mapsto w(\rho_i(\mathbf{X}) - p) + a_i (w - a_i) \) is convex and with condition (23), its minimum is attained at \( w = a_i \). Therefore, \( \mathbf{w}^* = a_i^\prime \) is an optimizer to (20), which shows the desired equilibrium statement.

(iii) By (i), \( (\mathbf{w}_1^*, \ldots, \mathbf{w}_n^*) \) is an \( n \)-permutation of \( (a_1^\prime, \ldots, a_n^\prime) \). It means that for any \( i \in [n] \), there exists \( j \in [n] \) such that \( a_j \) is the minimizer of (A.1). As \( c_i \) is convex, we have

\[
c_i^*(a_j - a_i) \geq p - \rho_i(\mathbf{X}) \geq c_i^*(a_j - a_i), \quad \text{for each} \quad i \in [n].
\]

Hence, we obtain (24). \( \square \)

**Proof of Proposition 6.** To show that \( \mathcal{F}_\text{IN} \) is closed under convolution, note that first-order stochastic dominance is closed under convolution; see Theorem 1.A.3 of Shaked and Shanthikumar (2007). Therefore, under independence,

\[
X_{1j} \leq_{st} \sum_{i=1}^{n} \theta_i X_{ij} \quad \text{for} \quad j = 1, 2 \implies \sum_{j=1}^{2} X_{1j} \leq_{st} \sum_{j=1}^{2} \sum_{i=1}^{n} \theta_i X_{ij} = \sum_{i=1}^{n} \theta_i \sum_{j=1}^{2} X_{ij}.
\]

To show that all \( \mathcal{F}_\text{IN}, \mathcal{F}_\text{NA} \) and \( \mathcal{F}_\text{WNA} \) are closed under strictly increasing convex transforms \( f \), we note that if \( Y \leq_{st} \sum_{i=1}^{n} \theta_i Y_i \), then \( f(Y) \leq_{st} f(\sum_{i=1}^{n} \theta_i Y_i) \leq \sum_{i=1}^{n} \theta_i f(Y_i) \), where the first
inequality follows since \( \leq \) is preserved under increasing transforms, and the second inequality is due to convexity of \( f \). Moreover, the strictly increasing transform \( f \) does not affect the dependence structure of \((Y_1, \ldots, Y_n)\).

**Proof of Proposition 7.** Copulas for independence and weak negative association are in \( \mathcal{C}_{\text{DP}} \) because of Theorem 1. The copula for comonotonicity is in \( \mathcal{C}_{\text{DP}} \) because \( X_1 = \sum_{i=1}^{n} \theta_i X_i \) almost surely in case of comonotonicity. Denote by \( C \) a copula of \((X_1, \ldots, X_n)\). Let \( X \overset{d}{=} X_1 \). Then, there exists a random vector \((U_1, \ldots, U_n) \sim C\) such that \((F_X^{-1}(U_1), \ldots, F_X^{-1}(U_n)) \overset{d}{=} (X_1, \ldots, X_n)\). Note that for \( p \in (0, 1) \), \( \mathbb{P}(\sum_{i=1}^{n} \theta_i F_X^{-1}(U_i) \leq p) \) is linear in the distribution of \((U_1, \ldots, U_n)\). Therefore, if \( \mathbb{P}(\sum_{i=1}^{n} \theta_i X_i \leq p) \leq \mathbb{P}(X \leq p) \) for all \( p \in (0, 1) \) holds for two different copulas, it also holds for their mixture.

**Proof of Proposition 8.** Let \( Y_j = (\sum_{i=n(j-1)+1}^{n(j)} X_i)/n, j = 1, \ldots, m \). By Theorem 1, \( X'_j \leq_{\text{st}} Y_j \) for \( j = 1, \ldots, m \), where \( X'_1, \ldots, X'_m \) are iid super-Pareto. Note that \( Y_1, \ldots, Y_m \) are also independent. As first-order stochastic dominance is closed under convolutions (e.g., Theorem 1.A.3 (a) of Shaked and Shanthikumar (2007)), we obtain

\[
X_1 + \cdots + X_m \overset{d}{=} X'_1 + \cdots + X'_m \leq_{\text{st}} Y_1 + \cdots + Y_m = \frac{X_1 + \cdots + X_{mn}}{n}.
\]

Dividing both sides by \( m \) yields the desired inequality.

**C**  
**Generalizations of the model**

Generalizations of (5) to a tail super-Pareto risk model and a classic insurance risk model are presented below.

**C.1 Tail super-Pareto distributions**

As reflected by the Pickands-Balkema-de Haan Theorem (see Theorem 3.4.13 (b) in Embrechts et al. (1997)), many losses have a power-like tail, but their distributions may not be power-like over the full support. Therefore, it is practically useful to assume that a random loss has a Pareto distribution only in the tail region; see the examples in Section 1.1. Let \( X \) be a super-Pareto random variable. We say that \( Y \) is distributed as \( X \) beyond \( x \) if \( \mathbb{P}(Y > t) = \mathbb{P}(X > t) \) for \( t \geq x \). Our next result suggests that, under an extra condition, stochastic dominance also holds in the tail region for such distributions.
Proposition A.2. Let $X$ be a super-Pareto random variable, $Y_1, \ldots, Y_n$ be iid random variables distributed as $X$ beyond $x \geq z_X$, and $Y \overset{d}{=} Y_1$. Assume that $Y \geq_{st} X$. For $(\theta_1, \ldots, \theta_n) \in \Delta_n$ and $t \geq x$, we have $P(\sum_{i=1}^n \theta_i Y_i > t) \geq P(Y > t)$, and the inequality is strict if $t > z_X$ and $\theta_i > 0$ for at least two $i \in [n]$.

In Proposition A.2, the assumption $Y \geq_{st} X$, that is, $P(Y > t) \geq P(X > t)$ for $t \in [z_X, x]$, is not dispensable. Here we cannot allow the distribution of $Y$ on $[z_X, x]$ to be arbitrary; the entire distribution is relevant in order to establish the inequality $P(\sum_{i=1}^n \theta_i Y_i > t) \geq P(Y > t)$, even for $t$ in the tail region.

C.2 A classic model in insurance

Random weights and a random number of risks are, for instance, common in modeling portfolios of insurance losses; see Klugman et al. (2012). Let $N$ be a counting random variable (i.e., it takes values in $\{0, 1, 2, \ldots\}$), and $W_i$ and $X_i$ be positive random variables for $i \in \mathbb{N}$. We consider an insurance portfolio where each policy incurs a loss $W_i X_i$ if there is a claim, and $N$ is the total number of claims in a given period of time. If $W_1 = W_2 = \cdots = 1$ and $X_1, X_2, \ldots$ are iid, then this model recovers the classic collective risk model. The total loss of a portfolio of insurance policies is given by $\sum_{i=1}^N W_i X_i$, and its average loss across claims is $\left(\sum_{i=1}^N W_i X_i\right) / \left(\sum_{i=1}^N W_i\right)$ where both terms are 0 if $N = 0$.

Proposition A.3. Let $X_1, X_2, \ldots$ be WNAID super-Pareto random variables, $X \overset{d}{=} X_1, W_1, W_2, \ldots$ be positive random variables, and $N$ be a counting random variable, such that $X, \{X_i\}_{i \in \mathbb{N}}, \{W_i\}_{i \in \mathbb{N}},$ and $N$ are independent. We have

$$X \mathbb{1}_{\{N \geq 1\}} \leq_{st} \frac{\sum_{i=1}^N W_i X_i}{\sum_{i=1}^N W_i} \quad \text{and} \quad \sum_{i=1}^N W_i X_i \leq_{st} \sum_{i=1}^N W_i X_i. \quad (A.2)$$

If $P(N \geq 2) \neq 0$, then for $t > z_X$, $P(\sum_{i=1}^N W_i X_i / \sum_{i=1}^N W_i \leq t) < P(X \mathbb{1}_{\{N \geq 1\}} \leq t)$.

If $W_1 = W_2 = \cdots = 1$ as in the classic collective risk model, then, under the assumptions of Proposition A.3, we have

$$X \mathbb{1}_{\{N \geq 1\}} \leq_{st} \frac{1}{N} \sum_{i=1}^N X_i \quad \text{and} \quad NX \leq_{st} \sum_{i=1}^N X_i.$$

To interpret the above inequalities, the sum of a randomly counted sequence of WNAID super-Pareto losses is stochastically larger than the sum of a randomly counted sequence of identical
super-Pareto losses. Therefore, building an insurance portfolio for WNAID super-Pareto claims does not reduce the total risk. In this setting, it is less risky to insure identical policies than to insure weakly negatively associated policies of the same type of super-Pareto loss and thus the basic principle of insurance does not apply to super-Pareto losses.

C.3 Proofs of Propositions A.2 and A.3

Proof of Proposition A.2. Let $X_1, \ldots, X_n$ be iid super-Pareto random variables. Note that for $t \geq x$, by using Theorem 1 and $Y \geq_{st} X$, we have

$$
\mathbb{P} \left( \sum_{i=1}^{n} \theta_i Y_i > t \right) \geq \mathbb{P} \left( \sum_{i=1}^{n} \theta_i X_i > t \right) \geq \mathbb{P} (X > t) = \mathbb{P} (Y > t).
$$

The statement on strictness also follows from Theorem 1. □

Proof of Proposition A.3. By Theorem 1 and the law of total expectation, it is easy to verify that, for $n = 2, 3, \ldots$, $\mathbb{P} (\sum_{i=1}^{n} W_i X_i / \sum_{i=1}^{n} W_i \leq t) < \mathbb{P} (X \leq t)$ for $t > z_X$. As $N$ is independent of $\{W_i X_i\}_{i \in \mathbb{N}}$, for $t > z_X$,

$$
\mathbb{P} \left( \frac{\sum_{i=1}^{N} W_i X_i}{\sum_{i=1}^{N} W_i} \leq t \right) = \mathbb{P} (N = 0) + \sum_{n=1}^{\infty} \mathbb{P} \left( \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i} \leq t \right) \mathbb{P} (N = n)
$$

$$
\leq \mathbb{P} (N = 0) + \mathbb{P} (N \geq 1) \mathbb{P} (X \leq t) = \mathbb{P} (X I_{\{N \geq 1\}} \leq t).
$$

It is obvious that the inequality is strict if $\mathbb{P} (N \geq 2) \neq 0$. To show the second inequality in (A.2), note that for each realization of $N = n$ and $(W_1, \ldots, W_N) = (w_1, \ldots, w_n) \in \mathbb{R}^n$, $\sum_{i=1}^{n} w_i X_i \leq_{st} \sum_{i=1}^{n} w_i X_i$ holds by Theorem 1. Hence, the second inequality in (A.2) holds. □

D Equilibrium analysis in a risk exchange economy

D.1 Discussions on Theorem 2

We discuss the results in Theorem 2 in this section. The equilibrium price $p$ should be very close to the individual risk assessments, and hence the risk sharing mechanism does not benefit the agents. Indeed, in (ii), the equilibrium allocation is equal to the original exposure, and there is no welfare gain. This is drastically different from the market in Section D.2 below.

In general, (23) and (24) are not equivalent, but in the two cases below, they are: (a) $a_1 = \cdots = a_n$; (b) $c_1 = \cdots = c_n = 0$. In either case, both (23) and (24) are a necessary and sufficient condition
for \((p, \ldots, p)\) to be an equilibrium price. Hence, the tuple \((p^*, \mathbf{w}^1, \ldots, \mathbf{w}^n)\) is an equilibrium if and only if (23) holds and \((\mathbf{w}^1, \ldots, \mathbf{w}^n)\) is an \(n\)-permutation of \((a^1, \ldots, a^n)\), which can be checked by Theorem 2 (i). In case (a), \(p\) cannot be too far away from \(\rho_i(X)\) for each \(i \in [n]\). In case (b), \(p = \rho_1(X) = \cdots = \rho_n(X)\), and an equilibrium can only be achieved when all agents agree on the risk of one unit of the loss and use this assessment for pricing.

We offer a few further technical remarks on Theorem 2.

1. Theorem 2 (ii) and (iii) remain valid for all mildly monotone, translation invariant, and positively homogeneous risk measures.

2. If the range of \(w^i = (w^i_1, \ldots, w^i_n)\) in (20) is constrained to \(0 \leq w^i_j \leq a_j\) for \(j \in [n]\), then \(((p, \ldots, p), a^1, \ldots, a^n)\) in Theorem 2 (ii) is still an equilibrium under the condition (23). However, the characterization statement in (i) is no longer guaranteed, which can be seen from the proof of Theorem 2 in Appendix B. As a result, (iii) cannot be obtained either.

3. The super-Pareto risk sharing market is closely related to model A in Section 3. Since model B has similar properties to model A in Proposition 4, we can check that the equilibrium in Theorem 2 (ii) still holds if we replace model A by model B, where the triggering events have the same probability of occurrence (i.e., \(\mathbb{P}(A_1) = \cdots = \mathbb{P}(A_n)\)). However, we cannot guarantee that all equilibria for model B have the form in (i) since holding one of the super-Pareto risks may not be the only optimal strategy for agents in model B; see Proposition 4.

D.2 A market with external risk transfer

In the setting of Section 4, we have considered risk exchange within the group of \(n\) agents, each of which has an initial loss. Next, we consider an extended market with external agents to which risk can be transferred with compensation from the internal agents.

As we have seen from Theorem 2, agents cannot reduce their risks by sharing extremely heavy-tailed losses within the group. As such, they may seek to transfer their risks to other parties external to the group. In this context, the internal agents are risk bearers, and the external agents are institutional investors without initial position of super-Pareto losses.

Consider a super-Pareto risk sharing market with \(n\) internal agents and \(m \geq 1\) external agents equipped with the same risk measure \(\rho_E : \mathcal{X} \rightarrow \mathbb{R}\). Let \(u^j \in \mathbb{R}^n_+\) be the exposure vector of the \(j^{th}\) external agent after sharing the risks of the internal agents, \(j \in [m]\). For the \(j^{th}\) external agent, the loss for taking position \(u^j\) is

\[
L_E(u^j, p) = u^j \cdot X - u^j \cdot p,
\]
where \( \mathbf{p} = (p_1, \ldots, p_n) \) is the premium vector. Like the internal agents, the goal of the external agents is to minimize their risk plus cost. That is, for \( j \in [m] \), external agent \( j \) minimizes 
\[
\rho_E \left( L_E(\mathbf{u}^j, \mathbf{p}) \right) + c_E(\|\mathbf{u}^j\|),
\]
where \( c_E \) is a non-negative cost function satisfying \( c_E(0) = 0 \).

For tractability, we will also make some simplifying assumptions on the internal agents. We assume that the internal agents have the same risk measure \( \rho_I \) and the same cost function \( c_I \). Assume that \( c_I \) and \( c_E \) are strictly convex and continuously differentiable except at 0, and \( \rho_I \) and \( \rho_E \) are mildly monotone distortion risk measures defined on \( \mathcal{X} \). In addition, all internal agents have the same amount \( a > 0 \) of initial loss exposures, i.e., \( a = a_1 = \cdots = a_n \). Finally, we consider the situation where the number of external agents is larger than the number of internal agents by assuming that \( m = kn \), where \( k \) is a positive integer, possibly large.

An equilibrium of this market is a tuple \((\mathbf{p}^*, \mathbf{w}^1^*, \ldots, \mathbf{w}^n^*, \mathbf{u}^1^*, \ldots, \mathbf{u}^m^*) \in (\mathbb{R}^n_+)^{n+m+1}\) if the following two conditions are satisfied.

(a) Individual optimality:
\[
\mathbf{w}^i^* \in \arg \min_{\mathbf{w}^i \in \mathbb{R}^n_+} \left\{ \rho_I \left( L_I(\mathbf{w}^i, \mathbf{p}^*) \right) + c_I(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|) \right\}, \quad \text{for each } i \in [n]; \tag{A.3}
\]
\[
\mathbf{u}^j^* \in \arg \min_{\mathbf{u}^j \in \mathbb{R}^n_+} \left\{ \rho_E \left( L_E(\mathbf{u}^j, \mathbf{p}^*) \right) + c_E(\|\mathbf{u}^j\|) \right\}, \quad \text{for each } j \in [m]. \tag{A.4}
\]

(b) Market clearance:
\[
\sum_{i=1}^n \mathbf{w}^i^* + \sum_{j=1}^m \mathbf{u}^j^* = \sum_{i=1}^n \mathbf{a}^i. \tag{A.5}
\]

The vector \( \mathbf{p}^* \) is an equilibrium price, and \((\mathbf{w}^1^*, \ldots, \mathbf{w}^n^*)\) and \((\mathbf{u}^1^*, \ldots, \mathbf{u}^m^*)\) are equilibrium allocations for the internal and external agents, respectively. Before identifying the equilibria in this market, we first make some simple observations. Let
\[
L_E(b) = c_E'(b) + \rho_E(X) \quad \text{and} \quad L_I(b) = c_I'(b) + \rho_I(X), \quad b \in \mathbb{R}.
\]

We will write \( L_I^-(0) = c_I^-(0) + \rho_I(X) \) and \( L_I^+(0) = c_I^+(0) + \rho_I(X) \) to emphasize that the left and right derivative of \( c_I \) may not coincide at 0; this is particularly relevant in Theorem 2 (ii). On the other hand, \( L_E(0) \) only has one relevant version since the allowed position is non-negative. Note that both \( L_E \) and \( L_I \) are continuous except at 0 and strictly increasing.

If an external agent takes only one source of loss (intuitively optimal from Proposition 4) among \( X_1, \ldots, X_n \) (we use the generic variable \( X \) for this loss), then \( L_E(b) \) is the marginal cost of further increasing their position at \( bX \). As a compensation, this agent will also receive \( p \).
Therefore, the external agent has incentives to participate in the risk sharing market if \( p > L_E(0) \).

If \( p \leq L_E(0) \), due to the strict convexity of \( c_E \), this agent will not take any risks. On the other hand, if \( p \geq L_I^-(0) \), which means that it is expensive to transfer the loss externally, then the internal agent has no incentive to transfer. For a small risk exchange to benefit both parties, we need \( L_E(0) < p < L_I^-(0) \). This implies, in particular,

\[
\rho_E(X) \leq L_E(0) < p < L_I^-(0) \leq \rho_I(X),
\]

which means that the risk is more acceptable to the external agents than to the internal agents, and the price is somewhere between the two risk assessments. The above intuition is helpful to understand the conditions in the following theorem. Denote by \( 0_n = (0, \ldots, 0) \in \mathbb{R}^n \).

**Theorem A.1.** Consider the super-Pareto risk sharing market of \( n \) internal and \( m = kn \) external agents. Let \( E = (p, w^1, \ldots, w^n, u^1, \ldots, u^m) \).

(i) Suppose that \( L_E(a/k) < L_I(-a) \). The tuple \( E \) is an equilibrium if and only if \( p = (p, \ldots, p) \), \( p = L_E(a/k) \), \( (u^1, \ldots, u^m) \) is a permutation of \( u^*(e_{[1/k], n}, \ldots, e_{[m/k], n}) \), \( u^* = a/k \), and \( (w^1, \ldots, w^n) = (0_n, \ldots, 0_n) \).

(ii) Suppose that \( L_E(a/k) \geq L_I(-a) \) and \( L_E(0) < L_I^-(0) \). Let \( u^* \) be the unique solution to

\[
L_E(u) = L_I(-ku), \quad u \in (0, a/k).
\] (A.6)

The tuple \( E \) is an equilibrium if and only if \( p = (p, \ldots, p) \), \( p = L_E(u^*) \), \( (u^1, \ldots, u^m) = u^*(e_{k_1, n}, \ldots, e_{k_m, n}) \), and \( (w^1, \ldots, w^n) = (a - ku^*) (e_{\ell_1, n}, \ldots, e_{\ell_n, n}) \), where \( k_1, \ldots, k_m \in [n] \) and \( \ell_1, \ldots, \ell_n \in [n] \) such that \( u^* \sum_{j=1}^m \mathbb{I}_{\{k_j = s\}} + (a - ku^*) \sum_{i=1}^n \mathbb{I}_{\{\ell_i = s\}} = a \) for each \( s \in [n] \).

Moreover, if \( u^* < a/(2k) \), then the tuple \( E \) is an equilibrium if and only if \( p = (p, \ldots, p) \), \( p = L_E(u^*) \), \( (u^1, \ldots, u^m) \) is a permutation of \( u^*(e_{[1/k], n}, \ldots, e_{[m/k], n}) \), and \( (w^1, \ldots, w^n) \) is a permutation of \( (a - ku^*)(e_{1, n}, \ldots, e_{n, n}) \).

(iii) Suppose that \( L_E(0) \geq L_I^-(0) \). The tuple \( E \) is an equilibrium if and only if \( p = (p, \ldots, p) \), \( p \in [L_I^-(0), L_E(0)\wedge L_I^+(0)] \), \( (u^1, \ldots, u^m) = (0_n, \ldots, 0_n) \), and \( (w^1, \ldots, w^n) \) is a permutation of \( a(e_{1, n}, \ldots, e_{n, n}) \).

Compared with Theorem 2, where no benefits exist from risk sharing among the internal agents, Theorem A.1 (ii) implies that in the presence of external agents, every party in the market may get better from risk sharing. More specifically, if \( L_E(0) < L_I^-(0) \), (i.e., the marginal cost of increasing
an external agent’s position from 0 is smaller than the marginal benefit of decreasing an internal agent’s position from $a$), there exists an equilibrium price $p \in [L_E(0), L_I^-(0)]$ such that both internal and external agents in the market can improve their objectives. The condition $L_E(0) < L_I^-(0)$ is crucial to such a win-win situation, as a price less than $L_I^-(0)$ will motivate the internal agents to transfer risk, and a price greater than $L_E(0)$ will motivate the external agents to receive risks. Theorem A.1 (i) shows that if $L_E(a/k) < L_I^-(0)$, all the losses will be transferred to the external agents.

As shown by Theorem A.1 (iii), if $L_E(0) \geq L_I^-(0)$, there are no incentives for the internal and external agents to participate in the risk sharing market, and their positions remain the same. Moreover, if $u^* < a/2$, i.e., the optimal position of each external agent is very small compared with the total position of each loss in the market, the loss $X_i$ for each $i \in [n]$, has to be shared by one internal agent and $k$ external agents in order to achieve an equilibrium.

We make further observations on Theorem A.1 (ii). From (A.6), it is straightforward to see that if $k$ gets larger (more external agents are in the market), the equilibrium price $p$ gets smaller. Intuitively, as more external agents are willing to take risks, they have to make some compromise on the received compensation to get the amount of risks they want. The lower price further motivates the internal agents to transfer more risks to the external agents. Indeed, by (A.6), $ku^*$ gets larger as $k$ increases. On the other hand, $u^*$ gets smaller as $k$ increases. In the equilibrium model, each external agent will take less risk if more external agents are in the market. These observations can be seen more clearly in the example below.

**Example A.1 (Quadratic cost).** Suppose that the conditions in Theorem A.1 (ii) are satisfied (this implies $\rho_E(X) < \rho_I(X)$ in particular), $c_I(x) = \lambda_I x^2$, and $c_E(x) = \lambda_E x^2$, $x \in \mathbb{R}$, where $\lambda_I, \lambda_E > 0$. We can compute the equilibrium price

$$p = \frac{k\lambda_I}{k\lambda_I + \lambda_E} \rho_E(X) + \frac{\lambda_E}{k\lambda_I + \lambda_E} \rho_I(X).$$

Therefore, the equilibrium price is a weighted average of $\rho_E(X)$ and $\rho_I(X)$, where the weights depend on $k$, $\lambda_I$, and $\lambda_E$. We also have the equilibrium allocations $u^* = (u, \ldots, u)$ and $w^* = (w, \ldots, w)$ where

$$u = \frac{\rho_I(X) - \rho_E(X)}{2(k\lambda_I + \lambda_E)} \quad \text{and} \quad w = \frac{k(\rho_E(X) - \rho_I(X))}{2(k\lambda_I + \lambda_E)} + a.$$

It is clear that $p$ moves in the opposite direction of $k$. Moreover, if more external agents are in the market, each external agent will take fewer losses, while each internal agent will transfer more
losses to the external agents. If \( \lambda_I \) increases, the internal agents will be less motivated to transfer their losses. To compensate for the increased penalty, the price paid by the internal agents will decrease so that they are still willing to share risks to some extent. The interpretation is similar if \( \lambda_E \) changes. Although the increase of different penalties (\( \lambda_E \) or \( \lambda_I \)) have different impacts on the price, the increase of either \( \lambda_E \) or \( \lambda_I \) leads to less incentives for the internal and external agents to participate in the risk sharing market.

D.3 Risk exchange for losses with finite mean

In contrast to the settings in Sections 4 and D.2, we study losses with finite mean below. Consider a market which is the same as the super-Pareto risk sharing market except that the losses are iid with finite mean. This market is called a risk sharing market with finite mean. The following proposition shows that agents prefer to share finite-mean losses among themselves if they are equipped with ES.

**Proposition A.4.** In a risk sharing market with finite mean, suppose that \( \rho_1 = \cdots = \rho_n = \text{ES}_q \) for some \( q \in (0, 1) \). Let

\[
\mathbf{w}^{i*} = \frac{a_i}{\sum_{j=1}^{n} a_j} \sum_{j=1}^{n} a^j \quad \text{for } i \in [n] \quad \text{and} \quad \mathbf{p}^* = (\mathbb{E}[X_1|A], \ldots, \mathbb{E}[X_n|A]),
\]

where \( A = \{ \sum_{i=1}^{n} a_i X_i \geq \text{VaR}_q \left( \sum_{i=1}^{n} a_i X_i \right) \} \). Then the tuple \((\mathbf{p}^*, \mathbf{w}^{1*}, \ldots, \mathbf{w}^{n*})\) is an equilibrium.

A sharp contrast is visible between the equilibrium in Theorem 2 and that in Proposition A.4. For WNAID super-Pareto losses, which do not have finite mean, the equilibrium price is the same across individual losses, and agents do not share losses at all. For iid finite-mean losses and ES agents, each individual loss has a different equilibrium price, and agents share all losses proportionally.

We choose the risk measure ES here because it leads to an explicit expression of the equilibrium. Although ES is not finite for super-Pareto losses (thus, it does not fit Theorem 2), it can be approximated arbitrarily closely by RVaR (e.g., Embrechts et al. (2018)) which fits the condition of Theorem 2. By this approximation, we expect a similar situation if ES in Proposition A.4 is replaced by RVaR, although we do not have an explicit result.

**Remark A.1.** Proposition A.4, in the case of iid losses with finite mean, works for all convex risk measures. The intuition is that the value of convex risk measures can be reduced by diversification, i.e., \( \rho(\lambda X + (1 - \lambda) Y) \leq \lambda \rho(X) + (1 - \lambda) \rho(Y) \) where \( \rho \) is a convex risk measure, \( X \) and \( Y \) are
two random variables with finite mean, and \( \lambda \in (0, 1) \). Convex risk measures are not suitable for the case of super-Pareto risks as they will always be infinite for risks without finite mean (see e.g., Filipović and Svindland (2012)).

D.4 Proofs of Theorem A.1 and Proposition A.4

Proof of Theorem A.1. As in Section 4, an optimal position for either the internal or the external agents is to concentrate on one of the losses \( X_i, i \in [n] \). By the same arguments as in Theorem 2 (i), the equilibrium price, if it exists, must be of the form \( \mathbf{p} = (p, \ldots, p) \). For such a given \( \mathbf{p} \), using the assumption that \( \rho_E \) and \( \rho_I \) are mildly monotone and Proposition 4, we can rewrite the optimization problems in (A.3) and (A.4) as

\[
\min_{u^j \in \mathbb{R}^n_+} \{ \rho_E (L_E(u^j, \mathbf{p})) + c_E(\|u^j\|) \} = \min_{u \in \mathbb{R}^n_+} \{ u(\rho_E(X) - p) + c_E(u) \}, \tag{A.7}
\]

and

\[
\min_{w^i \in \mathbb{R}^n_+} \{ \rho_I (L_I(w^i, \mathbf{p})) + c_I(\|w^i\| - \|a^i\|) \} = \min_{w \in \mathbb{R}^n_+} \{ w(\rho_I(X) - p) + ap + c_I(w - a) \}, \tag{A.8}
\]

for \( j \in [m] \) and \( i \in [n] \). Note that the derivative of the function inside the minimum of the right-hand side of (A.7) with respect to \( u \) is \( L_E(u) - p \), and similarly, \( L_I(w - a) - p \) is the derivative of the function inside the minimum of the right-hand side of (A.8). Using strict convexity of \( c_E \) and \( c_I \), we get the following facts.

1. The optimizer \( u \) to (A.7) has two cases:
   
   (a) If \( L_E(0) \geq p \), then \( u = 0 \).
   
   (b) If \( L_E(0) < p \), then \( u > 0 \) and \( L_E(u) = p \).

2. The optimizer \( w \) to (A.8) has four cases:
   
   (a) If \( L_I^+(0) < p \), then \( w > a \). This is not possible in an equilibrium.
   
   (b) If \( L_I^+(0) \geq p \geq L_I^-(0) \), then \( w = a \).
   
   (c) If \( L_I^-(0) > p > L_I^(-(a)) \), then \( 0 < w < a \) and \( L_I(w - a) = p \).
   
   (d) If \( L_I^-(a) \geq p \), then \( w = 0 \).

From the above analysis, we see that the optimal positions for each different external agent are either all 0 or all positive, and they are identical due to the strict monotonicity of \( L_E \). We can
say the same for the internal agents. Suppose that there is an equilibrium. Let $u$ be the external agent’s common exposure, and $w$ be the internal agent’s exposure. By the clearance condition (A.5) we have $w + ku = a$. If $0 < ku < a$, then from (1.b) and (2.c) above, we have $L_E(u) = L_I(-ku)$.

Below we show the three statements.

(i) The “if” statement is clear by (1.b) and (2.d). We show the “only if” statement. We first assume that $p \in (L_I(-a), L_I(0))$. Since we also have $p > L_E(a/k) > L_E(0)$, from (1.b) and (2.c), $u$ should satisfy $L_E(u) = L_I(-ku)$. However, by strict monotonicity of $L_E$ and $L_I$, there is no $u \in (0, a/k]$ such that $L_E(u) = L_I(-ku)$. Moreover, if $p \leq L_E(0)$ or $p \geq L_I(0)$, the clearance condition (A.5) cannot be satisfied. Therefore, we must have $L_E(0) < p \leq L_I(-a)$. In this case, $w = 0$ by (2.d). Consequently, $u = a/k$ by the clearance condition (A.5) and (1.b) gives $p = L_E(a/k)$.

(ii) In this case, there exists a unique $u^* \in (0, a/k]$ such that $L_E(u^*) = L_I(-ku^*)$. It follows that $u = u^*$ optimizes (A.7) and $w = a - ku^*$ optimizes (A.8). It is straightforward to verify that $\mathcal{E}$ is an equilibrium, and thus the “if” statement holds. To show the “only if” statement, it suffices to notice that $L_E(u) = L_I(-ku) = p$ has to hold, where $p$ is an equilibrium price and $u$ is the optimizer to (A.7), and such $u$ and $p$ are unique. Next, we show the “only if” statement for $u^* < a/2k$. As the optimal position for each external agent is $a - ku^* > a/2$, if more than two of the internal agents take the same loss, then the clearance condition (A.5) does not hold. Hence, the internal agents have to take different losses. Moreover, as the optimal position for the internal agents are the same, the loss $X_i$ for each $i \in [n]$, must be shared by one internal and $k$ external agents. The equilibrium is preserved under the permutation of allocations. Thus, we have the “only if” statement for $u^* < a/2k$. The “if” statement is obvious.

(iii) The “if” statement can be verified directly by using Theorem 2 (ii). Next, we show the “only if” statement. By (2.a), it is clear that the equilibrium price $p$ satisfies $p \leq L_I^+(0)$. If $p < L_I^-(0)$, by (1.a), (2.c), and (2.d), the clearance condition (A.5) cannot be satisfied. Thus, $p \geq L_I^-(0)$. By a similar argument, we have $p \leq L_E(0)$. Hence, we get $p \in [L_I^-(0), L_E(0) \land L_I^+(0)]$. From (1.a) and (2.b), we have $u = 0$ and $w = a$ and thus the desired result.

Proof of Proposition A.4. The clearance condition (21) is clearly satisfied. As ES is translation invariant, it suffices to show that $w^i \cdot X - w^i \cdot p^*$ minimizes $\text{ES}_q(w^i \cdot X - w^i \cdot p^*) + c_i(\|w^i\| - \|a^i\|)$ for $i \in [n]$. Write $r : w \mapsto \text{ES}_q(w \cdot X)$ for $w = (w_1, \ldots, w_n) \in [0, 1]^n$. By Corollary 4.2 of Tasche (2000),

$$\frac{\partial r}{\partial w_i}(w) = \mathbb{E}[X_i | A_w], \quad i \in [n],$$

Proof of Proposition A.4. The clearance condition (21) is clearly satisfied. As ES is translation invariant, it suffices to show that $w^i \cdot X - w^i \cdot p^*$ minimizes $\text{ES}_q(w^i \cdot X - w^i \cdot p^*) + c_i(\|w^i\| - \|a^i\|)$ for $i \in [n]$. Write $r : w \mapsto \text{ES}_q(w \cdot X)$ for $w = (w_1, \ldots, w_n) \in [0, 1]^n$. By Corollary 4.2 of Tasche (2000),

$$\frac{\partial r}{\partial w_i}(w) = \mathbb{E}[X_i | A_w], \quad i \in [n],$$
where $A_w = \{ \sum_{i=1}^n w_i X_i \geq \text{VaR}_q (\sum_{i=1}^n w_i X_i) \}$. Moreover, using convexity of $r$, we have (see McNeil et al. (2015, p. 321))

$$r(w) - w \cdot p^* \geq \sum_{i=1}^n w_i \frac{\partial r}{\partial w_i}(a_1, \ldots, a_n) - w \cdot p^* = 0.$$ 

By Euler’s rule (see McNeil et al. (2015, (8.61))), the equality holds if $w = \lambda (a_1, \ldots, a_n)$ for any $\lambda > 0$. By taking $\lambda = a_i / \sum_{j=1}^n a_j$, we get $\|w\| = a_i = \|a^i\|$, and hence $c_i(\|w\| - \|a^i\|)$ is minimized by $w = \lambda (a_1, \ldots, a_n)$. Therefore, $w^*$ is an optimizer for each $i \in [n]$.

## E Some details of the numerical results in Section 6.2

We first provide the Hill plots of the marine losses and wildfire suppression costs. The Hill estimates of the tail indices $\alpha$ are presented in Figure A.1, where the black curves represent the point estimates and the red curves represent the 95% confidence intervals with varying thresholds; see McNeil et al. (2015) for more details on the Hill estimator. As suggested by McNeil et al. (2015), one may roughly chose a threshold around the top 5% order statistics of the data. Following this suggestion, the tail indices $\alpha$ for the marine losses and wildfire suppression costs are estimated as 0.916 and 0.847 with 95% confidence intervals being (0.674, 1.158) and (0.776, 0.918), respectively.

Figure A.1: Hill plots for the marine losses and wildfire suppression costs: For each risk, the Hill estimates are plotted as black curve with the 95% confidence intervals being red curves.
Next, we will verify (27). Since the marine losses data were scaled to mask the actual losses, we renormalize it by multiplying the data by 500 to make it roughly on the same scale as that of the wildfire suppression costs;\(^5\) this normalization is made only for better visualization. Recall that \(\widehat{F}_1\) is the empirical distribution of the marine losses (renormalized) and \(\widehat{F}_2\) is the empirical distribution of the wildfire suppression costs. Take independent random variables \(\widehat{Y}_1 \sim \widehat{F}_1\) and \(\widehat{Y}_2 \sim \widehat{F}_2\). Let \(\widehat{F}_1 \oplus \widehat{F}_2\) be the distribution with quantile function \(p \mapsto \text{VaR}_p(\widehat{Y}_1) + \text{VaR}_p(\widehat{Y}_2)\), i.e., the comonotonic sum, and \(\widehat{F}_1 \ast \widehat{F}_2\) be the distribution of \(\widehat{Y}_1 + \widehat{Y}_2\), i.e., the independent sum.

The differences between the distributions \(\widehat{F}_1 \oplus \widehat{F}_2\) and \(\widehat{F}_1 \ast \widehat{F}_2\) can be seen in Figure A.2a. We observe that \(\widehat{F}_1 \ast \widehat{F}_2\) is less than \(\widehat{F}_1 \oplus \widehat{F}_2\) over a wide range of loss values. In particular, the relation holds for all losses less than 267,659.5 (marked by the vertical line in Figure A.2a). Equivalently, we can see from Figure A.2b that

\[
\text{VaR}_p(\widehat{Y}_1 + \widehat{Y}_2) > \text{VaR}_p(\widehat{Y}_1) + \text{VaR}_p(\widehat{Y}_2)
\]

holds unless \(p\) is greater than 0.9847 (marked by the vertical line in Figure A.2b).

Let \(F_1\) and \(F_2\) be the true distributions (unknown) of the marine losses (renormalized) and wildfire suppression costs, respectively. We are interested in whether the first-order stochastic dominance relation \(F_1 \ast F_2 \leq F_1 \oplus F_2\) holds. Since we do not have access to the true distributions, we generate two independent random samples of size \(10^4\) (roughly equal to the sum of the sizes of the datasets, thus with a similar magnitude of randomness) from the distributions \(\widehat{F}_1 \oplus \widehat{F}_2\) and \(\widehat{F}_1 \ast \widehat{F}_2\). We treat these samples as independent random samples from \(F_1 \oplus F_2\) and \(F_1 \ast F_2\) and test the hypothesis using Proposition 1 of Barrett and Donald (2003). The p-value of the test is greater than 0.5 and we are not able to reject the hypothesis \(F_1 \ast F_2 \leq F_1 \oplus F_2\).

\(^5\)The average marine losses (renormalized) and the average wildfire suppression costs are 12400 and 12899.
(a) Differences of the distributions: $\hat{F}_1 \oplus \hat{F}_2 - \hat{F}_1 \ast \hat{F}_2$

(b) Sample quantiles for $p \in (0.8, 0.99)$

Figure A.2: Plots for $\hat{F}_1 \oplus \hat{F}_2 - \hat{F}_1 \ast \hat{F}_2$ and sample quantiles