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

Yuyu Chen* Paul Embrechts† Ruodu Wang‡

September 10, 2022

Abstract

We show the perhaps surprising inequality that the weighted average of iid ultra heavy-tailed (i.e., infinite mean) Pareto losses is larger than a standalone loss in the sense of first-order stochastic dominance. This result is further generalized to allow for random total number and weights of Pareto losses and for the losses to be triggered by catastrophic events. We discuss several implications of these results via an equilibrium analysis in a risk exchange market. First, diversification of ultra heavy-tailed Pareto losses increases portfolio risk, and thus a diversification penalty exists. Second, agents with ultra heavy-tailed Pareto losses will not share risks in a market equilibrium. Third, transferring losses from agents bearing 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 ultra 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 Statistics and Actuarial Science, University of Waterloo, Canada. ✉ yuyu.chen@uwaterloo.ca

†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)). 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 life time of random variables exceeding a high level.

Stochastic dominance relations are an important tool in economic decision theory which allow 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.

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 $X \le_{\mathrm{st}} Y$, if $\mathbb{P}(X \le x) \ge \mathbb{P}(Y \le x)$ for all $x \in \mathbb{R}$. Write $X \simeq_{\mathrm{st}} Y$ if X and Y have the same distribution. The relation $X \le_{\mathrm{st}} Y$ means that all decision makers with an increasing[1] utility function will prefer the loss X to the loss Y, as studied by Quirk and Saposnik (1962) and Hadar and Russell (1969, 1971), and Müller and Stoyan (2002) and Shaked and Shanthikumar (2007) for the mathematics of stochastic dominance.

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

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

and the inequality (1) is strict in a natural sense. As far as we are aware, the inequality (1) is not known in the literature, even in the case that $\theta_1, \dots, \theta_n$ are equal (i.e., they are $1/n$). It is somewhat surprising that, for infinite-mean losses, the inequality (1) holds for the strongest form of risk comparison: for every monotone decision maker (with precise definition in Section 4), a diversified portfolio of such iid Pareto losses is less preferred to a non-diversified one. We call such

[1]In this paper, all terms like "increasing" and "decreasing" are in the non-strict sense.

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 $X_1, \dots, X_n$ with finite mean, regardless of their distribution or dependence structure, for $\theta_1, \dots, \theta_n > 0$ with $\sum_{i=1}^n \theta_n = 1$, (1) can only hold if $X_1 = \cdots = X_n$ (almost surely), in which case we have the trivial equality $X_1 = \theta_1 X_1 + \cdots + \theta_n X_n$; see Proposition 1. 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 p. 271 in Fama and Miller (1972) and Malinvaud (1972). The 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). A relevant result of Ibragimov (2009) is that for iid random variables $Z_1, \dots, Z_n$ which follow a convolution of symmetric stable distributions without finite mean, $\mathbb{P}(\theta_1 Z_1 + \cdots + \theta_n Z_n \le x) \le \mathbb{P}(Z_1 \le x)$ for $x > 0$ but the opposite holds for $x < 0$ (and hence first-order stochastic dominance does not hold[2]). The symmetry of distributions is essential for this inequality, and $Z_1, \dots, Z_n$ can take negative values, unlike Pareto losses, which are positive, skewed and more suitable for the modeling of extreme losses.

In the realm of banking and insurance, Pareto 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, the 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 not just in some asymptotic sense.

We obtain several generalizations of the inequality (1) for other models in Sections 2 and 3.

[2]This means that $\theta_1 Z_1 + \cdots + \theta_n Z_n$ is "more spread out" than Z_1. This notion is closer to second-order stochastic dominance, which captures mean-preserving spreads (although here the mean does not exist); see Ibragimov and Walden (2007).

In particular, Proposition 3 in Section 2 deals with losses that are Pareto only in the tail region, and Theorem 2 in Section 3 addresses losses triggered by catastrophic events, a setting where ultra heavy-tailed Pareto losses (hence infinite mean) are relevant.

We discuss in Section 4 the implications of (1) and related inequalities on the risk management decision of a single agent. It follows from (1) that the action of diversification increases the risk of ultra heavy-tailed Pareto losses *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 iid Pareto losses without finite mean and aims to minimize their risk by choosing a position across these losses, the optimal decision is to take only one of the Pareto losses (i.e., no diversification).

We proceed to study equilibria of a risk exchange market for Pareto losses under a few different settings in Section 5. As individual agents do not benefit from diversification in a risk exchange market where iid Pareto losses without finite mean 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 Pareto losses, the agents will merely exchange the entire loss position instead of risk sharing in an equilibrium model (Theorem 3 (i)). The situation becomes quite different if the agents with initial losses are allowed to transfer their losses to external parties. If the external agents have a stronger risk tolerance, then it is possible that both the internal and external agents can benefit by transferring losses from the internal to the external agents (Theorem 4 (ii)). In Proposition 7, we show that agents prefer to share Pareto losses with finite mean among themselves; this is in sharp contrast to the case of Pareto losses without finite mean. The above results are consistent with the observations made in Ibragimov et al. (2011) based on a different model.

In Section 6, numerical and real data examples are presented to illustrate the presence of ultra heavy tails in two real datasets in which the phenomenon of the inequality (1) can be empirically observe. We proceed to study the diversification effects of ultra heavy-tailed Pareto losses with different tail indices. Section 7 concludes the paper. Some background on risk measures is put in Appendix A, and proofs of all technical results are put in Appendix B.

We fix some notations. Throughout, random variables are defined on an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $\mathbb{N}$ the set of all positive integers and $\mathbb{R}_+$ the set of non-negative real numbers. For $n \in \mathbb{N}$, let $[n] = \{1, \dots, n\}$. Denote by Δ_n the standard simplex, that is, $\Delta_n = \{(\theta_1, \dots, \theta_n) \in [0,1]^n : \sum_{i=1}^n \theta_i = 1\}$. For $x, y \in \mathbb{R}$, write $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$, and

$x_+ = \max\{x, 0\}$. Recall that $X \simeq_{\mathrm{st}} Y$ means equality in distribution. We always assume $n \geq 2$.

1.1 Infinite-mean Pareto models

The key assumption of our paper is that Pareto losses have infinite mean, hence are so-called ultra heavy-tailed. 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). As we will discuss in Proposition 3 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 ultra heavy-tailed Pareto models; extra data examples are provided in Section 6.2.[3]

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 α of earthquake losses lie in the range $[0.6, 1.5]$. Estimated by Rizzo (2009), the tail indices α for some wind catastrophic losses are around 0.7. Hofert and Wüthrich (2012) showed that the tail indices α of losses caused by nuclear power accidents are around $[0.6, 0.7]$; similar observations 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 α 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 Nešlehová 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 α 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 α around 1. Biffis and Chavez (2014) showed that a number of large commercial property losses collected from two Lloyd's syndicates have tail indices α considerably less than 1. Silverberg and Verspagen (2007) concluded that the tail indices α are less than 1 for financial returns from some technological innovations. 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

[3] 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.

(2013) and Cirillo and Taleb (2020). Heavy-tailed to ultra heavy-tailed models also occur in the realm of climate change and environmental economics. Weitzman 's Dismal Theorem (see Weitzman (2009)) discusses the break-down 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 ultra heavy-tailed models, like cutting off the loss distribution model at some specific level, or tapering (concatinating a light-tailed distribution far in the tail of the loss distribution). Our experience shows that 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 Diversification of Pareto losses without finite mean

2.1 An unexpected stochastic dominance

A common parameterization of Pareto distributions is given by, for $\theta, \alpha > 0$,

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

Note that if $X \sim P_{\alpha,1}$, then $\theta X \sim P_{\alpha,\theta}$, and thus θ is a scale parameter. For $X \sim P_{\alpha,1}$, we write $X \sim \mathrm{Pareto}(\alpha)$. Moreover, the mean of $\mathrm{Pareto}(\alpha)$ is infinite if and only if the tail parameter $\alpha \in (0, 1]$. We say that the $\mathrm{Pareto}(\alpha)$ distribution is *ultra heavy-tailed* if $\alpha \leq 1$, and it is *moderately heavy-tailed* if $\alpha > 1$.

Theorem 1. *Let $X, X_1, \dots, X_n$ be iid $\mathrm{Pareto}(\alpha)$ random variables, $\alpha \in (0, 1]$. For $(\theta_1, \dots, \theta_n) \in \Delta_n$, we have*

$$X \leq_{\mathrm{st}} \sum_{i=1}^{n} \theta_i X_i. \tag{2}$$

Moreover, for $t > 1$, $\mathbb{P}\left(\sum_{i=1}^{n} \theta_i X_i > t\right) > \mathbb{P}\left(X > t\right)$ if $\theta_i > 0$ for at least two $i \in [n]$.

Remark 1 (Generalized Pareto distributions). The inequality (2) can be stated equivalently for other parameterizations of Pareto distributions without finite mean. For instance, it is often useful to consider generalized Pareto distributions, which provide an approximation for the excess losses beyond some high threshold. The generalized Pareto distribution for $\xi \geq 0$ is parametrized by

$$G_{\xi,\beta}(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x \geq 0, \tag{3}$$

where $\xi \geq 0$ ($\xi = 0$ corresponds to an exponential distribution) and $\beta > 0$; see Embrechts et al. (1997). If $\xi \geq 1$, then $G_{\xi,\beta}$ does not have finite mean. For $\xi > 0$, a generalized Pareto distribution in (3) can be converted to $P_{1/\xi,1}$ through a location-scale transform. Therefore, (2) implies that for $\xi \geq 1$, $(\beta_1, \ldots, \beta_n) \in (0, \infty)^n$ and independent random variables $Y_i \sim G_{\xi,\beta_i}$, $i \in [n]$, we have $Y \leq_{\text{st}} \sum_{i=1}^n Y_i$, where $Y \sim G_{\xi,\beta}$ with $\beta = \sum_{i=1}^n \beta_i$.

We will say that a *diversification penalty* exists if (2) holds, which is naturally interpreted as that having exposures in multiple iid ultra heavy-tailed Pareto losses is worse than having just one Pareto loss of the same total exposure. This observation will be generalized to a few other models later.

To better understand the result in Theorem 1, we stress that (2) cannot be expected if $X_1, \ldots, X_n$ have finite mean, regardless of their dependence structure, as summarized in the following proposition.

Proposition 1. *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,* (2) *holds if and only if $X_1 = \cdots = X_n$ almost surely.*

Proposition 1 implies, in particular, that (2) never holds for iid non-degenerate random variables $X, X_1, \ldots, X_n$ with finite mean. As such, it seems that 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 in the literature for which (2) holds other than the ones in this paper, all built on the basis of Theorem 1. We discuss the relation of Theorem 1 to the literature and some immediate relaxations in the next few remarks.

Remark 2. In the literature of EVT, it has been observed that, for iid ultra heavy-tailed Pareto risks $X_1, \ldots, X_n$,

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

holds true asymptotically as $t \to \infty$; see, e.g., Kaas et al. (2004), Albrecher et al. (2006), and Embrechts et al. (2009).[4] 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 4 and 5.

Remark 3. The inequality (2) also holds for some correlated ultra heavy-tailed Pareto risks. First, the inequality (2) simply holds for perfectly positively dependent ultra heavy-tailed Pareto risks (i.e., $X_1 = \cdots = X_n$ almost surely). Therefore, (2) remains true if the dependence structure (i.e., copula) of risks $X_1, \ldots, X_n$ is a mixture of independence and perfectly positive dependence; see Nelsen (2006) for an introduction to copulas. Besides this specific type of positive dependence structure, the inequality (2) may also hold for other dependence structures, but a rigorous analysis is beyond the scope of this paper.

Remark 4. An *ultra heavy-tailed Pareto sum* is a random variable $\sum_{j\in\mathbb{N}} \lambda_j Y_j$ where $Y_j \sim \mathrm{Pareto}(\alpha_j)$, $j \in \mathbb{N}$, are independent, $\alpha_j \in (0,1]$, $\lambda_j \in \mathbb{R}_+$, and $\sum_{j\in\mathbb{N}} \lambda_j < \infty$. The inequality (2) in Theorem 1 holds also for iid ultra heavy-tailed Pareto sums $X, X_1, \ldots, X_n$, and this can be shown by applying Theorem 1 to iid copies of each Y_j.

For an equally weighted pool of k iid Pareto losses, it is interesting to see whether enlarging k increases the risk in first-order stochastic dominance, i.e., for iid $\mathrm{Pareto}(\alpha)$ random variables $X_1, \ldots, X_\ell$, $\alpha \in (0,1]$, whether it holds that

$$\frac{1}{k}\sum_{i=1}^{k} X_i \le_{\mathrm{st}} \frac{1}{\ell}\sum_{i=1}^{\ell} X_i \quad \text{for } k, \ell \in \mathbb{N} \text{ and } k \le \ell. \tag{4}$$

The case of $k = 1$ in (4) corresponds to (2) with equal weights $\theta_1, \ldots, \theta_n$. The inequality (4) means that the more we diversify ultra heavy-tailed Pareto losses, the higher the penalty. In the next result, we show this inequality for the case that ℓ is a multiple of k.

Proposition 2. *For $m, n \in \mathbb{N}$, let $X_1, \ldots, X_{mn}$ be iid* $\mathrm{Pareto}(\alpha)$ *random variables, $\alpha \in (0,1]$. We have*

$$\frac{1}{m}\sum_{i=1}^{m} X_i \le_{\mathrm{st}} \frac{1}{mn}\sum_{i=1}^{mn} X_i.$$

Based on our numerical results in Section 6.1, we conjecture that the inequality (4) is true also for the general case that ℓ is not a multiple of k.

[4] Such an asymptotic result also holds for dependent ultra heavy-tailed Pareto risks (e.g., Embrechts et al. (2009) and Kley et al. (2016)).

2.2 Tail 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. For $\alpha > 0$, we say that Y has a Pareto(α) distribution beyond $x \geq 1$ if $\mathbb{P}(Y > t) = t^{-\alpha}$ 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 3. *Let $Y, Y_1, \dots, Y_n$ be iid random variables distributed as* Pareto(α) *beyond $x \geq 1$ and $\alpha \in (0, 1]$. Assume that $Y \geq_{\text{st}} X \sim$* Pareto($\alpha$)*. For $(\theta_1, \dots, \theta_n) \in \Delta_n$ and $t \geq x$, we have $\mathbb{P}\left(\sum_{i=1}^n \theta_i Y_i > t\right) \geq \mathbb{P}(Y > t)$, and the inequality is strict if $t > 1$ and $\theta_i > 0$ for at least two $i \in [n]$.*

In Proposition 3, the assumption $Y \geq_{\text{st}} X \sim$ Pareto(α), that is, $\mathbb{P}(Y > t) \leq t^{-\alpha}$ for $t \in [1, x]$, is not dispensable. Here we cannot allow the distribution of Y on $[1, x]$ to be arbitrary; the entire distribution is relevant in order to establish the inequality $\mathbb{P}\left(\sum_{i=1}^n \theta_i Y_i > t\right) \geq \mathbb{P}(Y > t)$, even for t in the tail region.

Let $X, X_1, \dots, X_n$ be iid Pareto(α) random variables with $\alpha \in (0, 1]$. As a particular application of Proposition 3, it holds that, for any $m \geq 1$,

$$X \vee m \leq_{\text{st}} \sum_{i=1}^n \theta_i (X_i \vee m). \tag{5}$$

This inequality follows by noting that $X \vee m$ has a Pareto distribution beyond m and applying Proposition 3 to $t \geq m$. A location shift of (5) also gives

$$(X - m)_+ \leq_{\text{st}} \sum_{i=1}^n \theta_i (X_i - m)_+. \tag{6}$$

For (5) and (6) to hold, it suffices to assume that $X_1, \dots, X_n$ are Pareto(α) beyond m, as their distribution on $(-\infty, m]$ does not matter.

2.3 A classic model in insurance

Theorem 1 can be easily generalized to include random weights and a random number of risks, which 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 $\{0, 1, 2, \dots\}$), 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_iX_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 $X_1, X_2, \dots$ are iid, then this model recovers the classic collective risk model. The total loss of a portfolio of insurance policies is given by $\sum_{i=1}^N W_iX_i$, and its average loss across claims is $(\sum_{i=1}^N W_iX_i)/(\sum_{i=1}^N W_i)$ where both terms are 0 if $N = 0$.

Proposition 4. *Let $X, X_1, X_2, \dots$ be iid* Pareto(α) *random variables, $\alpha \in (0, 1]$, $W_1, W_2, \dots$ 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*

$$X\mathbb{1}_{\{N\geq 1\}} \leq_{\rm st} \frac{\sum_{i=1}^N W_iX_i}{\sum_{i=1}^N W_i} \quad \textit{and} \quad \sum_{i=1}^N W_iX \leq_{\rm st} \sum_{i=1}^N W_iX_i. \tag{7}$$

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

If $W_1 = W_2 = \cdots = 1$ as in the classic collective risk model, then, under the assumptions of Proposition 4, we have

$$X_1\mathbb{1}_{\{N\geq 1\}} \leq_{\rm st} \frac{1}{N}\sum_{i=1}^N X_i \quad \text{and} \quad NX_1 \leq_{\rm st} \sum_{i=1}^N X_i.$$

To interpret the above inequalities, the average of a randomly counted sequence of iid Pareto(α) losses is stochastically larger than one member of the sequence if $\alpha \leq 1$. Therefore, building an insurance portfolio for iid ultra heavy-tailed Pareto claims does not reduce the total risk on average. In this setting, it is less risky to insure one large policy than to insure many independent policies of the same type of ultra heavy-tailed Pareto loss and thus the basic principle of insurance does not apply to ultra heavy-tailed Pareto losses.

3 A model for catastrophic losses

Catastrophic losses are large losses that usually occur with very small probabilities. It is practical to model an individual catastrophic loss as $X\mathbb{1}_A$, where A is the triggering event of the loss such that $X\mathbb{1}_A$ is Pareto distributed conditional on A (hence, we can assume that X is Pareto distributed and independent of A). Let $A_1, \dots, A_n$ be the triggering events of independent Pareto losses $X_1, \dots, X_n \sim$ Pareto(α), $\alpha \in (0, 1]$, such that $A_1, \dots, A_n$ are independent of the loss portfolio $(X_1, \dots, X_n)$. Let $(\theta_1, \dots, \theta_n) \in \mathbb{R}_+^n$ be the exposure vector. The total loss can then be written as $\theta_1X_1\mathbb{1}_{A_1} + \cdots + \theta_nX_n\mathbb{1}_{A_n}$. If $A_1 = \cdots = A_n$, meaning that $X_1, \dots, X_n$ represent different losses caused by the same catastrophic event, then, by Theorem 1, for $\lambda = \sum_{i=1}^n \theta_i > 0$,

$$X_1 \mathbb{1}_{A_1} \le_{\mathrm{st}} \frac{1}{\lambda} \sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i}. \tag{8}$$

Hence, diversification of losses from the same catastrophe *increases* the portfolio risk, and thus there is a diversification penalty. It remains to investigate whether a diversification penalty exists in this model (i.e., (8) holds) if $A_1, \dots, A_n$ are different, meaning that $X_1, \dots, X_n$ may represent losses caused by different catastrophic events. Diversification has two competing effects on the loss portfolio: It increases the frequency of having losses and decreases the sizes of the individual losses.

To illustrate the above trade-off, we first look at the diversification of two ultra heavy-tailed Pareto losses. Let X_1, X_2 be iid Pareto(α) random variables, $\alpha \in (0, 1]$, and A_1, A_2 be any events independent of (X_1, X_2). For simplicity, we assume that $(\theta_1, \theta_2) = (1/2, 1/2)$, and $\mathbb{P}(A_1) = \mathbb{P}(A_2)$. We have

$$\begin{aligned} \frac{1}{2} X_1 \mathbb{1}_{A_1} + \frac{1}{2} X_2 \mathbb{1}_{A_2} &= \frac{1}{2}(X_1 + X_2) \mathbb{1}_{A_1 \cap A_2} + \frac{1}{2} X_1 \mathbb{1}_{A_1 \cap A_2^c} + \frac{1}{2} X_2 \mathbb{1}_{A_1^c \cap A_2} \\ &\simeq_{\mathrm{st}} \frac{1}{2}(X_1 + X_2) \mathbb{1}_{A_1 \cap A_2} + \frac{1}{2} X_1 \mathbb{1}_{(A_1 \cap A_2^c) \cup (A_1^c \cap A_2)} \\ &\ge_{\mathrm{st}} X_1 \mathbb{1}_{A_1 \cap A_2} + \frac{1}{2} X_1 \mathbb{1}_{(A_1 \cap A_2^c) \cup (A_1^c \cap A_2)}, \end{aligned}$$

where the second-last equality holds as $A_1 \cap A_2^c$ and $A_1^c \cap A_2$ are mutually exclusive and $X_1 \simeq_{\mathrm{st}} X_2$, and the last inequality uses $\frac{1}{2}(X_1 + X_2) \mathbb{1}_{A_1 \cap A_2} \ge_{\mathrm{st}} X_1 \mathbb{1}_{A_1 \cap A_2}$ which follows by combining Theorem 1 and Theorem 1.A.14 of Shaked and Shanthikumar (2007). Write

$$X_1 \mathbb{1}_{A_1} = X_1 \mathbb{1}_{A_1 \cap A_2} + X_1 \mathbb{1}_{A_1 \cap A_2^c}.$$

Therefore, whether (8) holds in this setting boils down to whether

$$X_1 \mathbb{1}_{A_1 \cap A_2^c} \le_{\mathrm{st}} \frac{1}{2} X_1 \mathbb{1}_{(A_1 \cap A_2^c) \cup (A_1^c \cap A_2)} \tag{9}$$

holds. As $\mathbb{P}(A_1) = \mathbb{P}(A_2)$, $\mathbb{P}((A_1 \cap A_2^c) \cup (A_1^c \cap A_2)) = 2\mathbb{P}(A_1 \cap A_2^c)$. We write $p = \mathbb{P}(A_1 \cap A_2^c)$. We can directly compute, for $t \ge 0$,

$$\mathbb{P}(X_1 \mathbb{1}_{A_1 \cap A_2^c} > t) = p(t^{-\alpha} \wedge 1) \quad \text{and} \quad \mathbb{P}\left(\frac{1}{2} X_1 \mathbb{1}_{(A_1 \cap A_2^c) \cup (A_1^c \cap A_2)} > t\right) = (2p)((2t)^{-\alpha} \wedge 1).$$

Since $2((2t)^{-\alpha} \wedge 1) = 2^{1-\alpha}(t^{-\alpha} \wedge 2^{\alpha}) \ge (t^{-\alpha} \wedge 1)$, we obtain (9). Hence, diversification of two ultra heavy-tailed Pareto losses increases the portfolio risk if the two losses are triggered with the same probability. Theorem 2 provides a general result for diversifying any number of ultra heavy-tailed Pareto losses triggered with (possibly) different probabilities. To establish Theorem 2, we need the following lemma, which itself has a nice interpretation.

Lemma 1. *Let $X \sim \text{Pareto}(\alpha)$, $\alpha \in (0,1]$, and $B_1, \dots, B_n$ be mutually exclusive events independent of X. For $(c_1, \dots, c_n) \in [0,1]^n$, we have*

$$X\mathbb{1}_A \le_{\text{st}} \sum_{i=1}^{n} c_i X \mathbb{1}_{B_i},$$

where A is an event independent of X satisfying $\mathbb{P}(A) = \sum_{i=1}^{n} c_i \mathbb{P}(B_i)$.

Lemma 1 implies $X\mathbb{1}_A \le_{\text{st}} cX\mathbb{1}_B$, where $\mathbb{P}(A) = c\mathbb{P}(B)$ and $c \in (0,1]$. This implies that if we decrease the size of an ultra heavy-tailed Pareto loss (i.e., multiply X by c) and increase the probability of having the loss (i.e., divide $\mathbb{P}(A)$ by c), the loss becomes larger in first-order stochastic dominance. In general, the stochastic dominance cannot hold if X is a moderately heavy-tailed Pareto loss (i.e., X has a finite mean). For a moderately heavy-tailed Pareto loss X, $\mathbb{E}[cX\mathbb{1}_B] = \mathbb{E}[X\mathbb{1}_A]$. If, in addition, $X\mathbb{1}_A \le_{\text{st}} cX\mathbb{1}_B$ holds, then one has $X\mathbb{1}_A \simeq_{\text{st}} cX\mathbb{1}_B$ (Theorem 1.A.8 of Shaked and Shanthikumar (2007)), which does not hold unless $c = 1$. The above observation of ultra heavy-tailed Pareto losses consequently leads to Theorem 2.

Theorem 2. *Let $X_1, \dots, X_n$ be iid $\text{Pareto}(\alpha)$ random variables, $\alpha \in (0,1]$, and $A_1, \dots, A_n$ be any events independent of $(X_1, ..., X_n)$. For $(\theta_1, \dots, \theta_n) \in \mathbb{R}_+^n$, we have*

$$\lambda X\mathbb{1}_A \le_{\text{st}} \sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i}, \tag{10}$$

where $\lambda \ge \sum_{i=1}^{n} \theta_i$, $X \sim \text{Pareto}(\alpha)$, and A is independent of X satisfying $\lambda\mathbb{P}(A) = \sum_{i=1}^{n} \theta_i \mathbb{P}(A_i)$.

Remark 5. By setting $\mathbb{P}(A_1) = \dots = \mathbb{P}(A_n) = 1$, $(\theta_1, \dots, \theta_n) \in \Delta_n$ and $\lambda = 1$, Theorem 2 recovers the inequality (2) in Theorem 1. Moreover, a strict inequality

$$\mathbb{P}\left(\sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i} > t\right) > \mathbb{P}\left(\lambda X\mathbb{1}_A > t\right) \tag{11}$$

similar to Theorem 1 can be expected. A sufficient condition can be obtained using the strict inequality in Theorem 1: If there exists $S \subseteq [n]$ with at least two elements such that $\theta_i > 0$ for $i \in S$ and $\mathbb{P}(B_S) > 0$ where $B_S = \left(\bigcap_{i\in S} A_i\right) \cap \left(\bigcap_{i\in S^c} A_i^c\right)$, then (11) holds for $t > \sum_{i\in S} \theta_i$.

We discuss a special case of Theorem 2, which has practical relevance in risk sharing. Let $\mathbb{P}(A_1) = \cdots = \mathbb{P}(A_n)$, $X = X_1$, $A = A_1$, and $(\theta_1, \ldots, \theta_n) \in \Delta_n$. The inequality (10) can be rewritten as

$$X_1 \mathbb{1}_{A_1} \le_{\mathrm{st}} \sum_{i=1}^{n} \theta_i X_i \mathbb{1}_{A_i}. \tag{12}$$

The left-hand side of (12) can be regarded as the loss of an agent who keeps their own risk, and the right-hand side of (12) is the loss of an agent who shares risks with other agents. By pooling among ultra heavy-tailed Pareto losses, triggered by (possibly) different catastrophes, agents expect to suffer less loss when their own catastrophic loss occurs. However, every agent in the pool will have a higher frequency of bearing losses. Theorem 2 shows that the combined effects of diversification of ultra heavy-tailed Pareto losses lead to a higher probability of default at any capital reserve level, i.e., $\mathbb{P}(\sum_{i=1}^n \theta_i X_i \mathbb{1}_{A_i} > t) \ge \mathbb{P}(X_1 \mathbb{1}_{A_1} > t)$ for all $t > 0$.

4 Risk management decisions of a single agent

4.1 No diversification for a monotone agent

As hinted by (12) in Section 3, in a model of catastrophic losses $(X_1, \ldots, X_n)$ and triggering events $(A_1, \ldots, A_n)$, an agent who can choose between keeping their own risk or sharing risk 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 $\mathcal{X}$ be the set of all random variables, and let $L^1 \subseteq \mathcal{X}$ 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,

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

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

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 $\rho : \mathcal{X}_\rho \to \overline{\mathbb{R}} := [-\infty, \infty]$, 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 ρ for their preference, in the sense that the agent prefers a smaller value of ρ. 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 the ultra heavy-tailed 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 ρ, in the following two forms.

(a) Weak monotonicity: $\rho(X) \le \rho(Y)$ for $X, Y \in \mathcal{X}_\rho$ if $X \le_{\rm st} Y$.

(b) Mild monotonicity: ρ is weakly monotone and $\rho(X) < \rho(Y)$ if $F_X^{-1} < F_Y^{-1}$ on $(0,1)$.

Each of weak and mild monotonicity implies that $\rho(X) = \rho(Y)$ holds for $X \simeq_{\rm st} Y$. Common examples of preference models are all 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

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

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

$$\mathrm{VaR}_p(X) = F_X^{-1}(p) = \inf\{t \in \mathbb{R} : F_X(t) \ge p\}, \tag{13}$$

and ES is defined as

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

For $X \notin L^1$, such as the ultra heavy-tailed Pareto losses, $\mathrm{ES}_p(X)$ can be ∞, whereas $\mathrm{VaR}_p(X)$ is always finite. VaR is mildly monotone on $\mathcal{X}$, whereas ES is mildly monotone only on L^1.

In Theorems 1 and 2, we have established a diversification penalty for two models, which we will denote by $\mathbf{Y} = (Y_1, \dots, Y_n)$. In both models A and B below, let $X, X_1, \dots, X_n$ be iid Pareto(α) random variables, $\alpha \in (0,1]$, and $(\theta_1, \dots, \theta_n) \in \Delta_n$.

A. $Y_i = X_i$, $i \in [n]$ and $Y = X$.

B. $Y_i = X_i \mathbb{1}_{A_i}$, $i \in [n]$ and $Y = X\mathbb{1}_A$, where $A_1, \dots, A_n$ are any events independent of $(X_1, ..., X_n)$, and A is independent of X and satisfies $\mathbb{P}(A) = \sum_{i=1}^n \theta_i \mathbb{P}(A_i)$.

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 $\mathbb{E}[X] = \infty$). The following result on the diversification penalty of ultra heavy-tailed Pareto losses for a monotone agent follows directly from Theorems 1 and 2.

Proposition 5. *For $(\theta_1, \dots, \theta_n) \in \Delta_n$ and a weakly monotone risk measure $\rho : \mathcal{X}_\rho \to \overline{\mathbb{R}}$, for both models A and B, we have*

$$\rho\left(\sum_{i=1}^{n} \theta_i Y_i\right) \geq \rho(Y). \tag{14}$$

The inequality in (14) is strict for model A if ρ is mildly monotone and $\theta_i > 0$ for at least two $i \in [n]$.

We distinguish strict and non-strict inequalities in (14) because a strict inequality has stronger implications on the optimal decision of an agent. As an important consequence of Proposition 5, for $p \in (0,1)$ and $(\theta_1, \dots, \theta_n) \in \Delta^n$, in models A and B,

$$\mathrm{VaR}_p\left(\sum_{i=1}^{n} \theta_i Y_i\right) \geq \mathrm{VaR}_p(Y), \tag{15}$$

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

$$\mathrm{VaR}_p\left(\sum_{i=1}^{n} \theta_i Y_i\right) > \sum_{i=1}^{n} \theta_i \mathrm{VaR}_p(Y_i). \tag{16}$$

The inequality (16) and its non-strict version will be referred to as diversification penalty for VaR_p.

Remark 6. Diversification penalty for VaR_p also holds for other models that we consider. For instance, by Proposition 3, if $Y, Y_1, \dots, Y_n$ are iid Pareto(α) beyond $x \geq 1$ and $Y \geq_{\mathrm{st}} X \sim$ Pareto(α), then inequalities (15) and (16) hold for $p \geq 1 - x^{-\alpha}$.

From now on, we will focus on model A as it allows us to have a simple interpretation of the diversification penalty as in (16). Since all commonly used preference models are mildly monotone, Proposition 5 suggests that diversification of ultra heavy-tailed Pareto losses is detrimental for the agent.

Proposition 5 implies the following optimal decision for an agent in a market where several iid ultra heavy-tailed Pareto losses are present. Suppose that the agent needs to decide on a position $\mathbf{w} \in \mathbb{R}_+^n$ across these losses to minimize the total risk. The agent faces a total loss $\mathbf{w} \cdot \mathbf{Y} - g(\|\mathbf{w}\|)$ where the function g represents a compensation that depends on $\mathbf{w}$ through $\|\mathbf{w}\|$, and $\mathbf{Y}$ is as in model A or B. The agent's optimization problem then becomes

$$\text{to minimize} \quad \rho\left(\mathbf{w}\cdot\mathbf{Y}-g(\|\mathbf{w}\|)\right) \quad \text{subject to } \mathbf{w}\in\mathbb{R}_+^n \text{ and } \|\mathbf{w}\|=w \text{ with given } w>0, \tag{17}$$

or

$$\text{to minimize} \quad \rho\left(\mathbf{w}\cdot\mathbf{Y}-g(\|\mathbf{w}\|)\right) \quad \text{subject to } \mathbf{w}\in\mathbb{R}_+^n. \tag{18}$$

For $i\in[n]$, let $\mathbf{e}_{i,n}$ be the ith column vector of the $n\times n$ identity matrix, and $E_w=\{w\mathbf{e}_{i,n}: i\in[n]\}$ for $w\geq 0$, which represents the positions of only taking one loss with exposure w.

Proposition 6. *Let $\rho:\mathcal{X}_\rho\to\overline{\mathbb{R}}$ be weakly monotone and $g:\mathbb{R}\to\mathbb{R}$.*

(i) For model A, if ρ is mildly monotone, then the set of minimizers of (17) is E_w, and that of of (18) is contained in $\bigcup_{w\in\mathbb{R}_+}E_w$.

(ii) For models A and B, if (17) has an optimizer, then it has an optimizer in E_w; if (18) has an optimizer, then it has an optimizer in $\bigcup_{w\in\mathbb{R}_+}E_w$.

Remarkably, there are almost no restrictions on ρ and g in Proposition 6 other than monotonicity of ρ, and hence this result can be applied to many economic decision models.

Remark 7. Since ES_p is ∞ for the losses in models A and B, Proposition 6 applied to ES gives the trivial statement that every position has infinite risk. The main context of application for Proposition 6 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.

4.2 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 interesting 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 5 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 $X_1,\dots,X_n\sim$ Pareto(α), $\alpha\in(0,1]$, take $Y_i=X_i\wedge c_i$, where $c_i>1$ is the threshold, $i=1,\dots,n$. Note that each Y_i is bounded. Since Y_i has a finite mean, we cannot expect (15) or (16) to hold for all $p\in(0,1)$.

Nevertheless, we will see below that for a given p and large $c_1, \dots, c_n$, (16) holds, and thus there exists a diversification penalty for VaR_p.

For $p \in (0,1)$ and $(\theta_1, \dots, \theta_n) \in \Delta_n$, take $c_i \geq \mathrm{VaR}_p(\sum_{i=1}^n \theta_i X_i)/\theta_i$ for $i \in [n]$. Given that $X_i \geq c_i$ for $i \in [n]$, the distribution of X_i does not contribute to the calculation of $\mathrm{VaR}_p(\sum_{i=1}^n \theta_i X_i)$, and we have $\mathrm{VaR}_p(\sum_{i=1}^n \theta_i Y_i) = \mathrm{VaR}_p(\sum_{i=1}^n \theta_i X_i)$. Therefore, (16) holds for this choice of p and $(c_1, \dots, c_n)$. Hence, a diversification penalty for VaR_p exists for a fixed p if the thresholds $c_1, \dots, c_n$ are high enough.

If the excess-of-loss coverage is provided on an aggregate basis, then stochastic dominance holds as $X_1 \wedge c \leq_{\mathrm{st}} (\sum_{i=1}^n \theta_i X_i) \wedge c$ where $c > 1$ 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 $\rho(X_1 \wedge c) \leq \rho((\sum_{i=1}^n \theta_i X_i) \wedge c)$, and a diversification penalty exists for ρ. Unlike the situation of model A in Proposition 5, strict inequality may not hold for $\rho = \mathrm{VaR}_p$ because $X_1 \wedge c$ and $(\sum_{i=1}^n \theta_i X_i) \wedge c$ have the same p-quantile c for large p. Nevertheless, for the expected utility preference E_v, we have

$$\mathbb{E}[v(X_1 \wedge c)] < \mathbb{E}[v((\theta_1 X_1 + \cdots + \theta_n X_n) \wedge c)],$$

for $c > 1$ and v strictly increasing on $[1, c]$. This is because E_v is strictly monotone in the sense that for $X \leq_{\mathrm{st}} Y$ taking values in $[1, c]$ and $X \not\simeq_{\mathrm{st}} Y$, we have $E_v(X) < E_v(Y)$.

Remark 8. If the minimum in the above discussion is replaced by a maximum, then stochastic dominance holds, as discussed in (5) and (6).

5 Equilibrium analysis in a risk exchange economy

5.1 The Pareto risk sharing market model

Suppose that there are $n \geq 2$ agents in a risk exchange market. Let $\mathbf{X} = (X_1, \dots, X_n)$, where $X_1, \dots, X_n$ are iid Pareto(α) random variables with $\alpha > 0$. The ith 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 $\mathbf{a}^i \cdot \mathbf{X} = a_i X_i$.

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 bp_i, which is linear in b. Denote by $\mathbf{p} = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ the (endogenously generated) premium vector. Let $\mathbf{w}^i \in \mathbb{R}_+^n$ be the exposure vector of the ith agent from $\mathbf{X}$ after risk sharing. Then the loss of agent $i \in [n]$ after risk sharing is

$$L_i(\mathbf{w}^i, \mathbf{p}) = \mathbf{w}^i \cdot \mathbf{X} - (\mathbf{w}^i - \mathbf{a}^i) \cdot \mathbf{p}.$$

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 $\{\mathbf{X}\} \cup \mathbb{R}^n$. 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 *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 $\rho_i(L_i(\mathbf{w}^i, \mathbf{p})) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)$ over $\mathbf{w}^i \in \mathbb{R}^n_+$, $i \in [n]$. An *equilibrium* of the market is a tuple $\left(\mathbf{p}^*, \mathbf{w}^{1*}, \dots, \mathbf{w}^{n*}\right) \in (\mathbb{R}^n_+)^{n+1}$ if the following two conditions are satisfied.

(a) Individual optimality:

$$\mathbf{w}^{i*} \in \operatorname*{arg\,min}_{\mathbf{w}^i \in \mathbb{R}^n_+} \left\{\rho_i\left(L_i(\mathbf{w}^i, \mathbf{p}^*)\right) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)\right\}, \quad \text{for each } i \in [n]. \tag{19}$$

(b) Market clearance:

$$\sum_{i=1}^n \mathbf{w}^{i*} = \sum_{i=1}^n \mathbf{a}^i. \tag{20}$$

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

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

$$\rho(Y) = \int_{-\infty}^0 (h(\mathbb{P}(Y > x)) - 1)\mathrm{d}x + \int_0^\infty h(\mathbb{P}(Y > x))\mathrm{d}x, \tag{21}$$

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 ρ, 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 $\{\mathbf{X}\} \cup \mathbb{R}^n$; this always holds in case ρ is VaR or RVaR, and it holds for ρ being ES if $\alpha > 1$.

5.2 No risk exchange for ultra heavy-tailed Pareto losses

As anticipated from Proposition 6, each agent's optimal strategy is not to share with the other agents if their risk measure is mildly monotone and the Pareto losses are ultra heavy-tailed. 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 be a generic Pareto(α) random variable.

Theorem 3. *In a Pareto risk sharing market, suppose that $\alpha \in (0,1]$, and $\rho_1, \dots, \rho_n$ are mildly monotone.*

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

(ii) Suppose that $\rho_1, \dots, \rho_n$ are distortion risk measures on $\mathcal{X}$. The tuple $\left((p, \dots, p), \mathbf{a}^1, \dots, \mathbf{a}^n\right)$ is an equilibrium if p satisfies

$$c'_{i+}(0) \geq p - \rho_i(X) \geq c'_{i-}(0) \quad \text{for } i \in [n]. \tag{22}$$

(iii) Suppose that $\rho_1, \dots, \rho_n$ are distortion risk measures on $\mathcal{X}$. If $(p, \dots, p)$ is an equilibrium price, then

$$\max_{j \in [n]} c'_{i+}(a_j - a_i) \geq p - \rho_i(X) \geq \min_{j \in [n]} c'_{i-}(a_j - a_i) \quad \text{for } i \in [n]. \tag{23}$$

Theorem 3 (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 says that diversification among multiple ultra heavy-tailed Pareto losses increases risk in a uniform sense. As there is no diversification in the optimal allocation for each agent, taking any of these iid 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 (22). 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 (23). 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.

As a general message, 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. We will see later in Section 5.3 that in the presence of an external market, the picture is drastically different: the agents will benefit from transferring some losses to an external market.

In general, (22) and (23) 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 (22) and (23) are a necessary and sufficient condition for $(p, \dots, p)$ to be an equilibrium price. Hence, the tuple $\left(\mathbf{p}^*, \mathbf{w}^{1*}, \dots, \mathbf{w}^{n*}\right)$ is an equilibrium if and only if (22) holds and $(\mathbf{w}^{1*}, \dots, \mathbf{w}^{n*})$ is an n-permutation of $(\mathbf{a}^1, \dots, \mathbf{a}^n)$, which can be checked by Theorem 3 (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.

Although the agents will not benefit from sharing ultra heavy-tailed Pareto losses, the situation becomes different if these Pareto losses are moderately heavy-tailed, which will be discussed in Section 5.4.

Example 1 (Equilibrium for VaR agents with no costs)**.** Suppose that $c_i = 0$ for $i \in [n]$. Let $\rho_i = \mathrm{VaR}_q$, $q \in (0,1)$, $i \in [n]$. The tuple $\left(\mathbf{p}^*, \mathbf{w}^{1*}, \dots, \mathbf{w}^{n*}\right)$ is an equilibrium where $\mathbf{p}^* = ((1-q)^{-1/\alpha}, \dots, (1-q)^{-1/\alpha})$, and $(\mathbf{w}^{1*}, \dots, \mathbf{w}^{n*})$ is an n-permutation of $(\mathbf{a}^1, \dots, \mathbf{a}^n)$. For $i \in [n]$, $\rho_i\left(L_i(\mathbf{w}^{i*}, \mathbf{p}^*)\right) = \mathrm{VaR}_q(a_i X) = a_i(1-q)^{-1/\alpha}$.

Remark 9. We offer a few further technical remarks on Theorem 3.

1. Theorem 3 (ii) and (iii) remain valid for all mildly monotone, translation invariant, and positively homogeneous risk measures.

2. If the range of $\mathbf{w}^i = (w^i_1, \dots, w^i_n)$ in (19) is constrained to $0 \le w^i_j \le a_j$ for $j \in [n]$, then $\left((p, \dots, p), \mathbf{a}^1, \dots, \mathbf{a}^n\right)$ in Theorem 3 (ii) is still an equilibrium under the condition (22). However, the characterization statement in (i) is no longer guaranteed, which can be seen from the proof of Theorem 3 in Appendix B. As a result, (iii) cannot be obtained either.

3. The Pareto risk sharing market is closely related to model A in Section 4. Since model B has similar properties to model A in Proposition 6, we can check that the equilibrium in Theorem 3 (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 ultra heavy-tailed Pareto risks may not be the only optimal strategy for agents in model B; see Proposition 6.

5.3 A market with external risk transfer

In the setting of Section 5.2, we have considered a 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 3, agents cannot reduce their risks by sharing ultra 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 ultra heavy-tailed Pareto losses.

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

$$L_E(\mathbf{u}^j, \mathbf{p}) = \mathbf{u}^j \cdot \mathbf{X} - \mathbf{u}^j \cdot \mathbf{p},$$

where $\mathbf{p} = (p_1, \dots, p_n)$ 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 ρ_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 ρ_I and ρ_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 $(\mathbf{p}^*, \mathbf{w}^{1*}, \dots, \mathbf{w}^{n*}, \mathbf{u}^{1*}, \dots, \mathbf{u}^{m*}) \in (\mathbb{R}_+^n)^{n+m+1}$ if the following two conditions are satisfied.

(a) Individual optimality:

$$\mathbf{w}^{i*} \in \operatorname*{arg\,min}_{\mathbf{w}^i \in \mathbb{R}_+^n} \left\{ \rho_I \left(L_i(\mathbf{w}^i, \mathbf{p}^*) \right) + c_I(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|) \right\}, \quad \text{for each } i \in [n]; \tag{24}$$

$$\mathbf{u}^{j*} \in \operatorname*{arg\,min}_{\mathbf{u}^j \in \mathbb{R}_+^n} \left\{ \rho_E \left(L_E(\mathbf{u}^j, \mathbf{p}^*) \right) + c_E(\|\mathbf{u}^j\|) \right\}, \quad \text{for each } j \in [m]. \tag{25}$$

(b) Market clearance:

$$\sum_{i=1}^n \mathbf{w}^{i*} + \sum_{j=1}^m \mathbf{u}^{j*} = \sum_{i=1}^n \mathbf{a}^i. \tag{26}$$

The vector $\mathbf{p}^*$ is an *equilibrium price*, and $(\mathbf{w}^{1*}, \dots, \mathbf{w}^{n*})$ and $(\mathbf{u}^{1*}, \dots, \mathbf{u}^{m*})$ are *equilibrium allocations* for the internal and external agents, respectively. Before identifying the equilibria in this market, we first make some simple observations. Let

$$L_E(b) = c'_E(b) + \rho_E(X) \quad \text{and} \quad L_I(b) = c'_I(b) + \rho_I(X), \qquad b \in \mathbb{R}.$$

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 3 (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 6) among $X_1, \dots, X_n$ (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 $p \le L_E(0)$, due to the strict convexity of c_E, this agent will not take any risks. On the other hand, if $p \ge L_I^-(0)$, 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,

$$\rho_E(X) \le L_E(0) < p < L_I^-(0) \le \rho_I(X),$$

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.

Theorem 4. *In the Pareto risk sharing market of n internal and $m = kn$ external agents, suppose* *that* $\alpha \in (0,1]$. *Let* $\mathcal{E} = (\mathbf{p}, \mathbf{w}^{1*}, \ldots, \mathbf{w}^{n*}, \mathbf{u}^{1*}, \ldots, \mathbf{u}^{m*})$.

(i) If $L_E(a/k) < L_I(-a)$, *then there is no equilibrium.*

(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*

$$L_E(u) = L_I(-ku), \qquad u \in (0, a/k]. \tag{27}$$

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

Moreover, if $u^* < a/(2k)$, *then the tuple* $\mathcal{E}$ *is an equilibrium if and only if* $\mathbf{p} = (p, \ldots, p)$, $p = L_E(u^*)$, $(\mathbf{u}^{1*}, \ldots, \mathbf{u}^{m*})$ *is a permutation of* $u^*(\mathbf{e}_{\lceil 1/k \rceil,n}, \ldots, \mathbf{e}_{\lceil m/k \rceil,n})$, *and* $(\mathbf{w}^{1*}, \ldots, \mathbf{w}^{n*})$ *is a permutation of* $(a - ku^*)(\mathbf{e}_{1,n}, \ldots, \mathbf{e}_{n,n})$.

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

To interpret Theorem 4 (i), note that $L_E(a/k) < L_I(-a)$ implies $L_E(u) < L_I(w - a)$ for all $u \in [0, a/k]$ and $w \in [0, a]$. It means that if the price of transferring a unit of risk is in $[L_E(a/k), L_I(-a)]$, the optimal position for each internal agent will be 0, and the external agents will have the incentives to increase their exposures from 0 to more than a/k. In this case, the individual optimality conditions (24) and (25) and the clearance condition (26) cannot be satisfied at the same time. Therefore, there is no equilibrium.

Compared with Theorem 3, where no benefits exist from risk sharing among the internal agents, Theorem 4 (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. As shown by Theorem 4 (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 $u^* < a/2k$, 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 4 (ii). From (27), 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 (27), 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 2 (Quadratic cost)**.** Suppose that the conditions in Theorem 4 (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, λ_I, and λ_E. We also have the equlibrium allocations $\mathbf{u}^* = (u, \dots, u)$ and $\mathbf{w}^* = (w, \dots, w)$ 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 λ_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 λ_E changes. Although the increase of different penalties (λ_E or λ_I) have different impacts on the price, the increase of either λ_E or λ_I leads to less incentives for the internal and external agents to participate in the risk sharing market.

5.4 Risk exchange for moderately heavy-tailed Pareto losses

In contrast to the settings in Sections 5.2 and 5.3, we consider moderately heavy-tailed Pareto losses below. The following proposition shows that agents prefer to share moderately heavy-tailed

Pareto losses among themselves if they are equipped with ES.

Proposition 7. *In the Pareto risk sharing market, suppose that $\alpha \in (1,\infty)$, and $\rho_1 = \cdots = \rho_n = \mathrm{ES}_q$ for some $q \in (0,1)$. Let*

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

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

A sharp contrast is visible between the equilibrium in Theorem 3 and that in Proposition 7. For $\alpha \in (0,1]$, the equilibrium price is the same across individual losses, and agents do not share losses at all. For $\alpha \in (1,\infty)$ 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 ultra heavy-tailed Pareto losses (thus, it does not fit Theorem 3), it can be approximated arbitrarily closely by RVaR (e.g., Embrechts et al. (2018)) which fits the condition of Theorem 3. By this approximation, we expect a similar situation if ES in Proposition 7 is replaced by RVaR, although we do not have an explicit result.

Remark 10. Proposition 7, in the case of Pareto(α), $\alpha > 1$, 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 ρ 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 ultra heavy-tailed Pareto risks as they will always be infinite for risks without finite mean (see e.g., Filipović and Svindland (2012)).

6 Numerical examples

6.1 Diversification effects as n increases

For $\alpha \in (0,1]$, $p \in (0,1)$, and iid Pareto(α) random variables $X_1, \dots, X_n$, we compute $\mathrm{VaR}_p(\sum_{i=1}^n X_i/n)$ for $n = 2, \dots, 6$. From Figure 1, we observe that $\mathrm{VaR}_p(\sum_{i=1}^n X_i/n)$ increases as n increases. The difference between the curves for different n becomes more pronounced as α becomes smaller, i.e., the tail of the Pareto losses becomes heavier. From these numerical results, we may expect that

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

where $k, \ell \in \mathbb{N}$ and $k \leq \ell$. We were only able to show the case where ℓ is a multiple of k in Proposition 2.

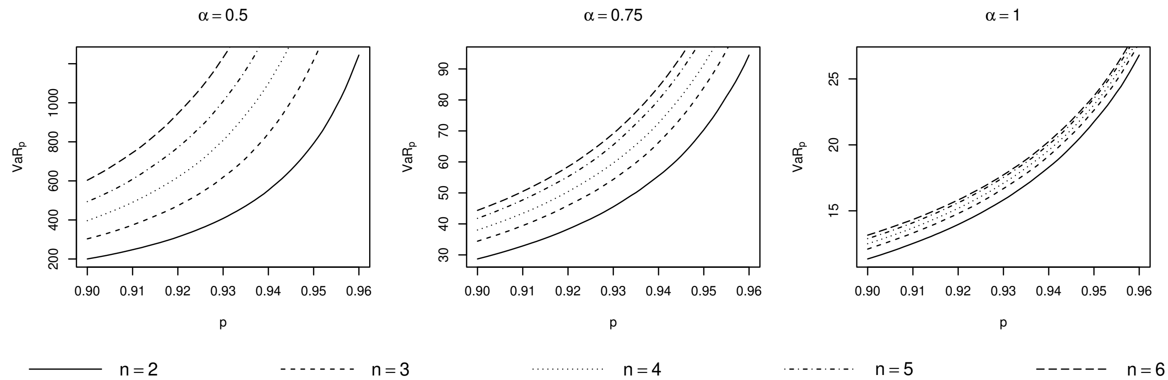


Figure 1: $\mathrm{VaR}_p((X_1 + \cdots + X_n)/n)$ for $n = 2, \ldots, 6$ and $p \in (0.9, 0.96)$.

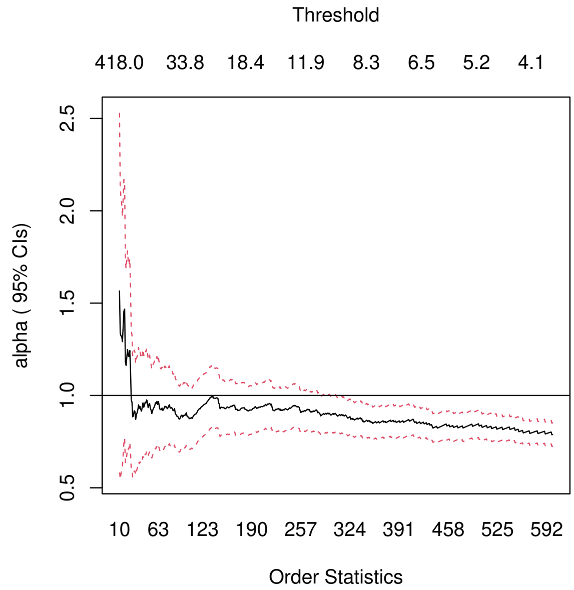
6.2 Examples of ultra heavy-tailed Pareto losses

In addition to the many examples mentioned in Section 1.1, we provide two further data examples: a first one on marine losses, and a second one on suppression costs of wildfires. Using EVT, we will show that both examples exhibit infinite mean behavior. The marine losses dataset, from the insurance data repository `CASdatasets`,[5] was originally collected by a French private insurer and comprises 1,274 marine losses (paid) between January 2003 and June 2006. The wildfire dataset[6] 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 Hill estimates of the tail indices α are presented in Figure 2, 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 α 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; thus, these losses/costs have infinite mean if they follow Pareto distributions in their tails regions.

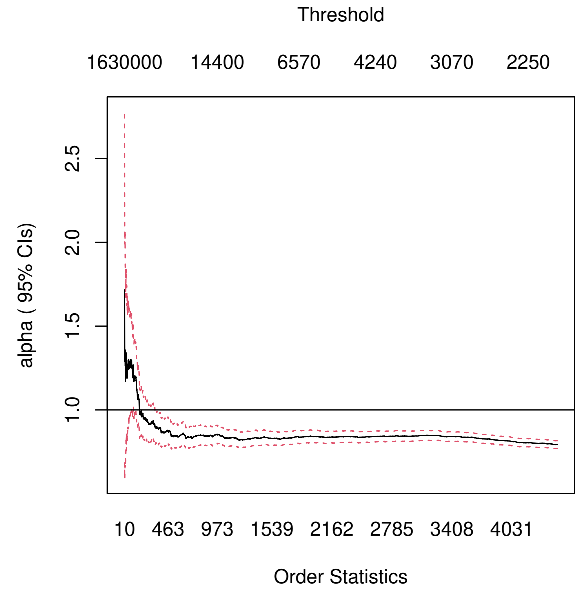
The observations in Figure 2 suggest that the two loss datasets may have similar tail parameters. As discussed in Remark 1, Theorem 1 can be applied to generalized Pareto distributions. If two loss random variables X_1 and X_2 are independent and follow generalized Pareto distributions

[5]Available at http://cas.uqam.ca/.

[6]Available at https://wildfire.alberta.ca/resources/historical-data/historical-wildfire-database.aspx.



(a) Marine losses



(b) Wildfire suppression costs

Figure 2: 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.

with the same tail parameter $\alpha = 1/\xi < 1$ (see (3)), then, for all $p \in (0,1)$,

$$\mathrm{VaR}_p(X_1 + X_2) > \mathrm{VaR}_p(X_1) + \mathrm{VaR}_p(X_2). \tag{28}$$

Even if X_1 and X_2 are not Pareto distributed, as long as their tails are Pareto, (28) may hold for p relatively large, as suggested by Proposition 3 and Remark 6.

We will verify (28) on our datasets to show how the implication of our main results holds for real data. 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;[7] this normalization does not matter for (28) and is made only for better visualization. Let $\widehat{F}_1$ be the empirical distribution of the marine losses (renormalized) and $\widehat{F}_2$ be 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 $p \mapsto \mathrm{VaR}_p(\widehat{Y}_1) + \mathrm{VaR}_p(\widehat{Y}_2)$, i.e., the comonotonic sum, and $\widehat{F}_1 * \widehat{F}_2$ 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 $\widehat{F}_1 * \widehat{F}_2$ can be seen in Figure 3a. We observe that $\widehat{F}_1 * \widehat{F}_2$ 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 3a). Equivalently, we

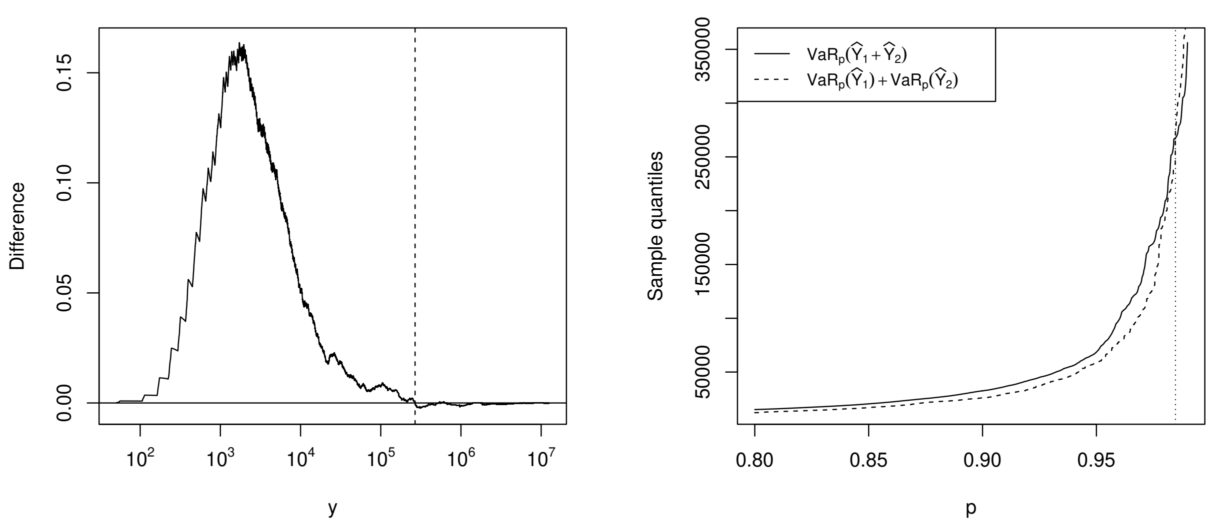
[7] The average marine losses (renormalized) and the average wildfire suppression costs are 12400 and 12899.

can see from Figure 3b that

$$\mathrm{VaR}_p(\widehat{Y}_1 + \widehat{Y}_2) > \mathrm{VaR}_p(\widehat{Y}_1) + \mathrm{VaR}_p(\widehat{Y}_2) \tag{29}$$

holds unless p is greater than 0.9847 (marked by the vertical line in Figure 3b). Recall that $\widehat{F}_1 * \widehat{F}_2 \le \widehat{F}_1 \oplus \widehat{F}_2$ is equivalent to (29) holding for all $p \in (0,1)$. Since the quantiles are directly computed from data, thus from distributions with bounded supports, for p close enough to 1 it must hold that $\mathrm{VaR}_p(\widehat{Y}_1 + \widehat{Y}_2) \le \mathrm{VaR}_p(\widehat{Y}_1) + \mathrm{VaR}_p(\widehat{Y}_2)$; see the similar observation made in Proposition 1. Nevertheless, we observe (29) for most values of $p \in (0,1)$. Note that the observation of (29) is entirely empirical and it does not use the fitted models.

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 $F_1 * F_2 \le F_1 \oplus F_2$ 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 $\widehat{F}_1 * \widehat{F}_2$. We treat these samples as independent random samples from $F_1 \oplus F_2$ and $F_1 * F_2$ 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 $F_1 * F_2 \le F_1 \oplus F_2$.



(a) Differences of the distributions: $\widehat{F}_1 \oplus \widehat{F}_2 - \widehat{F}_1 * \widehat{F}_2$

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

Figure 3: Plots for $\widehat{F}_1 \oplus \widehat{F}_2 - \widehat{F}_1 * \widehat{F}_2$ and sample quantiles

6.3 Aggregation of Pareto risks with different parameters

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

$$\sum_{i=1}^{n} \mathrm{VaR}_p(Y_i) \leq \mathrm{VaR}_p\left(\sum_{i=1}^{n} Y_i\right), \text{ usually with strict inequality.} \tag{30}$$

Inspired by the results in Section 6.2, we are interested in whether (30) 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 $Y_1, \dots, Y_6$ 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
ξ_i	1.19	1.17	1.01	1.39	1.23	1.22
β_i	774	254	233	412	107	243

Table 1: The estimated parameters ξ_i and β_i, $i \in [6]$.

For the purpose of this numerical example, we assume that $Y_1, \dots, Y_6$ are independent and plot $\sum_{i=1}^{6} \mathrm{VaR}_p(Y_i)$ and $\mathrm{VaR}_p(\sum_{i=1}^{6} Y_i)$ for $p \in (0.95, 0.99)$ in Figure 4. We can see that $\mathrm{VaR}_p(\sum_{i=1}^{6} Y_i)$ is larger than $\sum_{i=1}^{6} \mathrm{VaR}_p(Y_i)$, and the gap between the two values gets larger as the level p approaches 1. This observation further suggests that, even if the ultra 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 $Y_1, \dots, Y_n$ with shape parameters $\xi_1, \dots, \xi_n \in [1, \infty)$, although we do not have a proof. Similarly, we may expect that $\sum_{i=1}^{n} \theta_i \mathrm{VaR}_p(X_i) \leq \mathrm{VaR}_p(\sum_{i=1}^{n} \theta_i X_i)$ holds for any Pareto losses $X_1, \dots, X_n$ with tail parameters $\alpha_1, \dots, \alpha_n \in (0, 1]$,

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 equal parameters, 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 ultra-heavy tailed 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

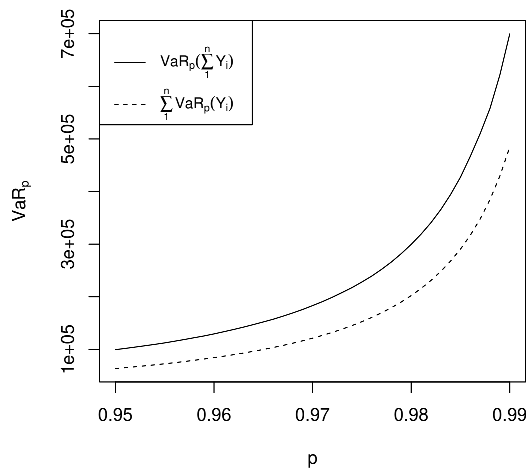


Figure 4: Curves of $\mathrm{VaR}_p(\sum_{i=1}^n Y_i)$ and $\sum_{i=1}^n \mathrm{VaR}_p(Y_i)$ for $n = 6$ generalized Pareto losses with parameters in Table 1 and $p \in (0.95, 0.99)$.

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 4 and 5 carry over to the above heterogeneous setting.

7 Concluding remarks

We establish in Theorem 1 the inequality that the diversification of iid Pareto losses without finite mean is greater than an individual Pareto loss in the sense of first-order stochastic dominance, which is a very strong dominance relation. The result of stochastic dominance is further generalized to three cases: (i) the losses are Pareto in the tail region (Proposition 3); (ii) the number and weights of Pareto losses are random (Proposition 4); (iii) the Pareto losses are triggered by catastrophic events (Theorem 2). These results provide an important implication in risk management, i.e., the diversification of Pareto losses without finite mean may increase the risk assessment of a portfolio (Proposition 6).

The equilibrium of a risk exchange model is analyzed in Theorem 3, where agents can take extra Pareto losses with compensations. In particular, if every agent is associated with an initial position of a Pareto loss without finite mean, 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 4).

The diversification effects are investigated by numerical studies where two open technical questions arise. The first question is whether (4) holds, that is,

$$\frac{1}{k}\sum_{i=1}^{k} X_i \leq_{\mathrm{st}} \frac{1}{\ell}\sum_{i=1}^{\ell} X_i, \tag{31}$$

holds for all $k, \ell \in \mathbb{N}$ such that $k \leq \ell$, where $X_1, \dots, X_l$ are iid Pareto losses without finite mean. The statement is true if ℓ is a multiple of k, as shown in Proposition 2. The second question is whether

$$\mathrm{VaR}_p\left(\sum_{i=1}^{n} \theta_i X_i\right) \geq \sum_{i=1}^{n} \theta_i \mathrm{VaR}_p(X_i) \tag{32}$$

holds for $(\theta_1, \dots, \theta_n) \in \Delta_n$ and independent ultra heavy-tailed Pareto losses $X_1, \dots, X_n$ with possibly different tail parameters. From the numerical results in Section 6, both (31) and (32) are anticipated to hold; a proof seems to be beyond the current techniques.

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, 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

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

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

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

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

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*, 62(5):1431–1457.

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.

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

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.

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

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

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

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

Hardy, G. H., Littlewood, J. E. and Pólya, G (1934). *Inequalities*. Cambridge University Press.

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

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

Kley, O., Klüppelberg, C. and Reinert, G. (2016). Risk in a large claims insurance market with bipartite graph structure. *Operations Research*, 64(5):1159–1176.

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

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.

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.

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.

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. *The American Economic Review*, 60(3):457–

459.

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

Appendices

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 $\rho : \mathcal{X}_\rho \to \mathbb{R}$:

(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 ultra 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

$$\mathrm{RVaR}_{p,q}(X) = \frac{1}{q-p}\int_p^q \mathrm{VaR}_u(X)\mathrm{d}u.$$

For $p \in (0, 1)$, $\lim_{q \downarrow p^+} \mathrm{RVaR}_{p,q}(X) = \mathrm{VaR}_p(X)$. 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 (21). 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 ρ_h be a distortion risk measure with distortion function h. Suppose that ρ_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 (21) we get

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

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 $h(t) = \lambda \mathbb{1}_{\{0<t\leq 1\}} + (1-\lambda)\mathbb{1}_{\{t=1\}}$ for $t \in [0,1]$, which is the distortion function of $\rho_h = \lambda\, \text{ess-inf} + (1-\lambda)\, \text{ess-sup}$. □

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 all theorems, propositions, and lemmas

B.1 Proofs of the results in Section 2

Proof of Theorem 1. For $(u_1, \dots, u_n) \in (0,1)^n$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) \in \Delta_n$, define the generalized weighted average $M_{r,\boldsymbol{\theta}}(u_1, \dots, u_n) = (\theta_1 u_1^r + \cdots + \theta_n u_n^r)^{\frac{1}{r}}$, where $r \in \mathbb{R}$. Note that (2) can be equivalently written as

$$M_{r,\boldsymbol{\theta}}(U_1, \dots, U_n) \leq_{\text{st}} U, \tag{A.1}$$

where $U, U_1, \dots, U_n$ are iid uniform random variables on $(0,1)$, and $r = -1/\alpha \in (-\infty, -1]$. It is well known that $M_{r,\boldsymbol{\theta}} \leq M_{s,\boldsymbol{\theta}}$ for $r \leq s$; see Theorem 16 of Hardy et al. (1934). Hence,

$M_{r,\boldsymbol{\theta}}(U_1,\dots,U_n) \le M_{-1,\boldsymbol{\theta}}(U_1,\dots,U_n)$ for all $r \le -1$. Therefore, for (A.1) to hold for all $r \le -1$, it suffices to show that $M_{-1,\boldsymbol{\theta}}(U_1,\dots,U_n) \le_{\mathrm{st}} U$.

If some $\theta_1,\dots,\theta_n$ are 0, we can reduce the dimension of the problem. Hence, we will assume $\min_{i\in[n]} \theta_i > 0$ in the proof below. There is nothing to show if only one $\theta_i > 0$ which reduces to dimension 1.

We first show the case of $n = 2$. For a fixed $p \in (0,1)$ and $\boldsymbol{\theta} = (\theta_1,\theta_2) \in \Delta_2$ where $\min\{\theta_1,\theta_2\} > 0$, let $\delta = \theta_2/(p^{-1} - 1 + \theta_2)$. For $(u_1,u_2) \in (0,1)^2$, if $u_2 \le \delta$, then

$$\theta_1 u_1^{-1} + \theta_2 u_2^{-1} \ge \theta_1 + \theta_2\delta^{-1} = 1 - \theta_2 + p^{-1} - 1 + \theta_2 = p^{-1}.$$

Hence, $M_{-1,\boldsymbol{\theta}}(u_1,u_2) \le p$ if $u_2 \le \delta$. Then, for iid uniform random variables U_1 and U_2 on $(0,1)$, we have

$$\begin{aligned}
\mathbb{P}\left(M_{-1,\boldsymbol{\theta}}(U_1,U_2) \le p\right) &= \mathbb{P}\left(\theta_1 U_1^{-1} + \theta_2 U_2^{-1} \ge p^{-1}\right) \\
&= \mathbb{P}(U_2 \le \delta) + \mathbb{P}\left(\theta_1 U_1^{-1} \ge p^{-1} - \theta_2 U_2^{-1}, U_2 > \delta\right) \\
&\ge \mathbb{P}(U_2 \le \delta) + \mathbb{P}\left(\theta_1 U_1^{-1} \ge p^{-1} - \theta_2, U_2 > \delta\right) \\
&= \delta + \theta_1(1-\delta)(p^{-1} - \theta_2)^{-1} \\
&> \delta + \theta_1(1-\delta)p \\
&= \theta_1 p + p\delta(p^{-1} - 1 + \theta_2) = p.
\end{aligned}$$

Hence, we have shown the result when $n = 2$. Next, let $n \ge 2$, and $\boldsymbol{\theta} = (\theta_1,\dots,\theta_{n+1}) \in \Delta_{n+1}$ where $\min_{i\in[n+1]} \theta_i > 0$. Let $U, U_1,\dots,U_{n+1}$ be iid uniform random variables on $(0,1)$. Assume that $U^{-1} \le_{\mathrm{st}} \theta_1/(\sum_{i=1}^n \theta_i)U_1^{-1} + \dots + \theta_n/(\sum_{i=1}^n \theta_i)U_n^{-1}$. As first-order stochastic dominance is closed under convolutions (e.g., Theorem 1.A.3 (a) of Shaked and Shanthikumar (2007)), we have

$$\theta_1 U_1^{-1} + \dots + \theta_{n+1}U_{n+1}^{-1} \ge_{\mathrm{st}} \left(\sum_{i=1}^n \theta_i\right) U^{-1} + \theta_{n+1}U_{n+1}^{-1} \ge_{\mathrm{st}} U^{-1},$$

Thus, $M_{-1,\boldsymbol{\theta}}(U_1,\dots,U_{n+1}) \le_{\mathrm{st}} U$. Moreover, for $p \in (0,1)$,

$$\begin{aligned}
\mathbb{P}\left(M_{-1,\boldsymbol{\theta}}(U_1,\dots,U_{n+1}) \le p\right) &= \mathbb{P}\left(\theta_1 U_1^{-1} + \dots + \theta_{n+1}U_{n+1}^{-1} \ge p^{-1}\right) \\
&\ge \mathbb{P}\left(\left(\sum_{i=1}^n \theta_i\right) U^{-1} + \theta_{n+1}U_{n+1}^{-1} \ge p^{-1}\right) > p.
\end{aligned}$$

By induction, we have the desired result. □

Proof of Proposition 1. Note that (1) implies that $\mathrm{ES}_p(X) \le \mathrm{ES}_p(\sum_{i=1}^n \theta_i X_i)$ for all $p \in (0,1)$, where ES_p is defined in Section 4. Since ES_p is convex and $X_1, \dots, X_n$ are identically distributed, we have

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

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

Proof of Proposition 2. Let $Y_j = (\sum_{i=n(j-1)+1}^{jn} X_i)/n$, $j = 1, \dots, m$. By Theorem 1, $X_j' \le_{\mathrm{st}} Y_j$ for $j = 1, \dots, m$, where $X_1', \dots, X_m' \sim \mathrm{Pareto}(\alpha)$ are independent. Note that $Y_1, \dots, Y_m$ 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

$$X_1 + \cdots + X_m \simeq_{\mathrm{st}} X_1' + \cdots + X_m' \le_{\mathrm{st}} Y_1 + \cdots + Y_m = \frac{X_1 + \cdots + X_{mn}}{n}.$$

Dividing both sides by m yields the desired inequality. □

Proof of Proposition 3. Let $X_1, \dots, X_n$ be iid $\mathrm{Pareto}(\alpha)$ random variables. Note that for $t \ge x$, by using Theorem 1 and $Y \ge_{\mathrm{st}} X$, we have

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

The statement on strictness also follows from Theorem 1. □

Proof of Proposition 4. By Theorem 1 and the law of total expectation, it is easy to verify that, for $n = 2, 3, \dots$, $\mathbb{P}(\sum_{i=1}^n W_i X_i / \sum_{i=1}^n W_i \le t) < \mathbb{P}(X \le t)$, $t > 1$. As N is independent of $\{W_i X_i\}_{i \in \mathbb{N}}$, for $t > 1$,

$$\begin{aligned}\mathbb{P}\left(\frac{\sum_{i=1}^N W_i X_i}{\sum_{i=1}^N W_i} \le t\right) &= \mathbb{P}(N = 0) + \sum_{n=1}^\infty \mathbb{P}\left(\frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i} \le t\right) \mathbb{P}(N = n) \\ &\le \mathbb{P}(N = 0) + \mathbb{P}(N \ge 1)\left(1 - \frac{1}{t^\alpha}\right) = \mathbb{P}\left(X \mathbb{1}_{\{N \ge 1\}} \le t\right).\end{aligned}$$

It is obvious that the inequality is strict if $\mathbb{P}(N \geq 2) \neq 0$. To show the second inequality in (7), note that for each realization of $N = n$ and $(W_1, \dots, W_N) = (w_1, \dots, w_n) \in \mathbb{R}^n$, $\sum_{i=1}^n w_i X \leq_{\rm st} \sum_{i=1}^n w_i X_i$ holds by Theorem 1. Hence, the second inequality in (7) holds. □

B.2 Proofs of the results in Section 3

Proof of Lemma 1. The result is clearly true if $c_1 = \dots = c_n = 0$. If any components of $(c_1, \dots, c_n)$ are zero, the problem simply reduces its dimension. Hence, we assume that $(c_1, \dots, c_n) \in (0, 1]^n$ for the rest of the proof. For $t \geq 1 \geq \max_{i \in [n]} c_i$,

$$\mathbb{P}\left(\sum_{i=1}^n c_i X \mathbb{1}_{B_i} \leq t\right) = \sum_{i=1}^n \left(1 - \frac{c_i^\alpha}{t^\alpha}\right) \mathbb{P}(B_i) + \mathbb{P}\left(\bigcap_{i \in [n]} B_i^c\right).$$

Since $B_1, \dots, B_n$ are mutually exclusive, $\sum_{i=1}^n \mathbb{P}(B_i) = \mathbb{P}\left(\bigcup_{i \in [n]} B_i\right) = 1 - \mathbb{P}\left(\bigcap_{i \in [n]} B_i^c\right)$. Moreover, as $c_i \in (0, 1]$ and $\alpha \in (0, 1]$, $c_i^\alpha \geq c_i$ for $i \in [n]$. Therefore,

$$\mathbb{P}\left(\sum_{i=1}^n c_i X \mathbb{1}_{B_i} \leq t\right) = 1 - \sum_{i=1}^n \frac{c_i^\alpha}{t^\alpha} \mathbb{P}(B_i) \leq 1 - \frac{1}{t^\alpha} \sum_{i=1}^n c_i \mathbb{P}(B_i) = 1 - \frac{1}{t^\alpha} \mathbb{P}(A) = \mathbb{P}(X \mathbb{1}_A \leq t).$$

For $t \in [0, 1)$, $\mathbb{P}\left(\sum_{i=1}^n c_i X \mathbb{1}_{B_i} \leq t\right) \leq \mathbb{P}\left(\sum_{i=1}^n c_i X \mathbb{1}_{B_i} \leq 1\right) \leq 1 - \mathbb{P}(A) = \mathbb{P}(X \mathbb{1}_A \leq t)$. This yields the desired result. □

Proof of Theorem 2. For $S \subseteq [n]$, let $B_S = \left(\bigcap_{i \in S} A_i\right) \cap \left(\bigcap_{i \in S^c} A_i^c\right)$. For $(\theta_1, \dots, \theta_n) \in \mathbb{R}_+^n$, we write

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

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

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

Note that

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

As $\sum_{i\in S}\theta_i/\lambda \in [0,1]$ for any $S \subseteq [n]$, by Lemma 1, $\sum_{S\subseteq[n]}(\sum_{i\in S}\theta_i/\lambda)X\mathbb{1}_{B_S} \geq_{\rm st} X\mathbb{1}_A$. Hence, $\sum_{i=1}^n \theta_i X_i \mathbb{1}_{A_i} \geq_{\rm st} \lambda X \mathbb{1}_A$. □

B.3 Proofs of the results in Section 4

Proof of Proposition 6. The proof of (i) follows directly from Theorem 1. Statement (ii) follows from Theorem 2 by noting that there exists $j \in [n]$ such that $\mathbb{P}(A_j) \leq \mathbb{P}(A)$, and hence,

$$wX_j\mathbb{1}_{A_j} \leq_{\rm st} wX\mathbb{1}_A \leq_{\rm st} \sum_{i=1}^n w_i X_i \mathbb{1}_{A_i},$$

where X and A are as in (10) with $\lambda = w$ and $(\theta_1,\dots,\theta_n) = (w_1,\dots,w_n)$. □

B.4 Proofs of the results in Section 5

Proof of Theorem 3. (i) Suppose that $\left(\mathbf{p}^*, \mathbf{w}^{i*}, \dots, \mathbf{w}^{n*}\right)$ forms an equilibrium. We let $p = \max_{j\in[n]}\{p_j\}$ and $S = \arg\max_{j\in[n]}\{p_j\}$. For the ith agent, by writing $w = \|\mathbf{w}^i\|$, using Theorem 1 and the fact that ρ_i is mildly monotone, we have that for any $\mathbf{w}^i \in [0,1]^n$,

$$\begin{aligned}\rho_i(L_i(\mathbf{w}^i,\mathbf{p}^*)) &= \rho_i(\mathbf{w}^i\cdot(\mathbf{X}-\mathbf{p}^*) + \mathbf{a}^i\cdot\mathbf{p}^*) \\ &\geq \rho_i(\mathbf{w}^i\cdot\mathbf{X} - wp + \mathbf{a}^i\cdot\mathbf{p}^*) \geq \rho_i(wX_1 - wp + \mathbf{a}^i\cdot\mathbf{p}^*).\end{aligned}$$

By the last statement of Theorem 1, the last inequality is strict if $\mathbf{w}^i$ contains at least two non-zero components. Moreover, $c(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|) = c(w - \|\mathbf{a}^i\|)$. Therefore, we know that the optimizer $\mathbf{w}^{i*} = (w_1^{i*},\dots,w_n^{i*})$ to (19) has at most one non-zero component w_j^{i*}, and $j \in S$. Hence, $w_k^{i*} = 0$ if $k \in [n]\setminus S$ and this holds for each $i \in [n]$. Using $\sum_{i=1}^n \mathbf{w}^{i*} = \sum_{i=1}^n \mathbf{a}^i$ which have all positive components, we know that $S = [n]$, which further implies that $\mathbf{p}^* = (p,\dots,p)$ for $p \in \mathbb{R}_+$. Next, as each $\mathbf{w}^{i*}$ has only one positive component, $(\mathbf{w}^{i*},\dots,\mathbf{w}^{n*})$ has to be an n-permutation of $(\mathbf{a}^1,\dots,\mathbf{a}^n)$ in order to satisfy the clearance condition (20).

(ii) The clearance condition (20) 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 6, for $i \in [n]$,

$$\begin{aligned}
&\min_{\mathbf{w}^i \in \mathbb{R}^n_+} \left\{\rho_i \left(L_i(\mathbf{w}^i, \mathbf{p}^*)\right) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)\right\} \\
&= \min_{\mathbf{w}^i \in \mathbb{R}^n_+} \left\{\rho_i \left(\mathbf{w}^i \cdot \mathbf{X} - (\mathbf{w}^i - \mathbf{a}^i) \cdot \mathbf{p}^*\right) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)\right\} \\
&= \min_{\|\mathbf{w}^i\| \in \mathbb{R}_+} \left\{(\rho_i \left(\|\mathbf{w}^i\| X\right) - (\|\mathbf{w}^i\| - a_i)p) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)\right\} \\
&= \min_{w \in \mathbb{R}_+} \left\{w(\rho_i(X) - p) + a_i p + c_i(w - a_i)\right\}. \qquad \text{(A.2)}
\end{aligned}$$

Note that $w \mapsto w(\rho_i(X) - p) + c_i(w - a_i)$ is convex and with condition (22), its minimum is attained at $w = a_i$. Therefore, $\mathbf{w}^{i*} = \mathbf{a}^{i*}$ is an optimizer to (19), which shows the desired equilibrium statement.

(iii) By (i), $\left(\mathbf{w}^{1*}, \ldots, \mathbf{w}^{n*}\right)$ is an n-permutation of $(\mathbf{a}^1, \ldots, \mathbf{a}^n)$. It means that for any $i \in [n]$, there exists $j \in [n]$ such that a_j is the minimizer of (A.2). As c_i is convex, we have

$$c'_{i+}(a_j - a_i) \geq p - \rho_i(X) \geq c'_{i-}(a_j - a_i), \quad \text{for each} \quad i \in [n].$$

Hence, we obtain (23). □

Proof of Theorem 4. As in Section 5.2, 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 3 (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 ρ_E and ρ_I are mildly monotone and Proposition 6, we can rewrite the optimization problems in (24) and (25) as

$$\min_{\mathbf{u}^j \in \mathbb{R}^n_+} \left\{\rho_E \left(L_E(\mathbf{u}^j, \mathbf{p})\right) + c_E(\|\mathbf{u}^j\|)\right\} = \min_{u \in \mathbb{R}_+} \left\{u \left(\rho_E \left(X\right) - p\right) + c_E(u)\right\}, \qquad \text{(A.3)}$$

and

$$\min_{\mathbf{w}^i \in \mathbb{R}^n_+} \left\{\rho_I \left(L_i(\mathbf{w}^i, \mathbf{p})\right) + c_I(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)\right\} = \min_{w \in \mathbb{R}_+} \left\{w(\rho_I \left(X\right) - p) + ap + c_I(w - a)\right\}, \qquad \text{(A.4)}$$

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.3) 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.4). Using strict convexity of c_E and c_I, we get the following facts.

1. The optimizer u to (A.3) 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.4) 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 (26) 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) If $L_E(a/k) < L_I(-a)$, then by strict monotonicity of L_E and L_I, there is no $u \in (0, a/k]$ such that $L_E(u) = L_I(-ku)$. Since u cannot be larger than a/k, if an equilibrium exists, then $u = 0$; but in this case, by (1.a) and (2.b), we have $L_E(0) \geq p \geq L_{I-}(0)$, which contradicts $L_E(a/k) < L_I(-a)$. Hence, there is no equilibrium.

(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.3) and $w = a - ku^*$ optimizes (A.4). 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.3), 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 (26) 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 3 (ii). Next, we show the "only if" statement. By (2.a), it is clear that the equilibrium price p satisfies $p \le L_I^+(0)$. If $p < L_I^-(0)$, by (1.a), (2.c), and (2.d), the clearance condition (26) cannot be satisfied. Thus, $p \ge L_I^-(0)$. By a similar argument, we have $p \le L_E(0)$. Hence, we get $p \in [L_I^-(0), L_E(0) \wedge L_I^+(0)]$. From (1.a) and (2.b), we have $u = 0$ and $w = a$ and thus the desired result. □

Proof of Proposition 7. The clearance condition (20) is clearly satisfied. As ES is translation invariant, it suffices to show that $\mathbf{w}^{i*}$ minimizes $\mathrm{ES}_q(\mathbf{w}^i \cdot \mathbf{X} - \mathbf{w}^i \cdot \mathbf{p}^*) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)$ for $i \in [n]$. Write $r : \mathbf{w} \mapsto \mathrm{ES}_q(\mathbf{w} \cdot \mathbf{X})$ for $\mathbf{w} = (w_1, \dots, w_n) \in [0,1]^n$. By Corollary 4.2 of Tasche (2000),

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

where $A_{\mathbf{w}} = \{\sum_{i=1}^n w_i X_i \ge \mathrm{VaR}_q(\sum_{i=1}^n w_i X_i)\}$. Moreover, using convexity of r, we have (see McNeil et al. (2015, p. 321))

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

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