Infinite-mean models in risk management: Discussions and recent advances

Yuyu Chen*   Ruodu Wang†

October 25, 2024

Abstract

In statistical analysis, many classic results require the assumption that models have finite mean or variance, including the most standard versions of the laws of large numbers and the central limit theorems. Such an assumption may not be completely innocent, and it may not be appropriate for datasets with heavy tails (e.g., catastrophic losses), relevant to financial risk management. In this paper, we discuss the importance of infinite-mean models in economics, finance, and related fields, with recent results and examples. We emphasize that many results or intuitions that hold for finite-mean models turn out to fail or even flip for infinite-mean models. Due to the breakdown of standard thinking for infinite-mean models, we argue that if the possibility of using infinite-mean models cannot be excluded, great caution should be taken when applying classic methods that are usually designed for finite-mean cases in finance and insurance.

Keywords: heavy tails; extreme value theory; finance; insurance; infinite-mean

1 Introduction

The purpose of this article is to review and discuss recent advances in infinite-mean models and their intriguing behavior and implications in risk management and statistics. The discussion is by no means exhaustive, but rather it reflects the subjective choices of the authors and their limitation of knowledge.

Let us first set straight a simple, but perhaps important and puzzling, question. Whereas whether the world is finite or not may be philosophical or agnostic, it is agreed by most people that the financial world is finite, because the total amount of money in the world is finite. Why

---
*Department of Economics, University of Melbourne, Australia. ✉ yuyu.chen@unimelb.edu.au

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

are we interested in a model with infinite mean, when the risk variables that we try to model are eventually bounded?

The short answer is that these infinite-mean models fit statistically better the reality in many relevant situations, and they provide useful insights on risk management problems, which are sometimes as useful as finite-mean models, but very different in nature, as we will discuss.

Important to infinite-mean models is the principle of power law. Power-law models have been extensively employed to study various heavy-tailed phenomena such as population incomes, city sizes, word frequencies, stock returns, and financial losses. They are also made popular to the public via its connection to the Black Swan of Taleb (2007). Andriani and McKelvey (2007) and Clauset et al. (2009) discussed the applications and empirical evidence of power-law models. Although models with finite moments are commonly assumed in many statistical analyses, in the context of finance and insurance, power-law models with divergent second or even first moments are ubiquitous.

Since empirical observations are finite by nature, sample moments of all orders are always finite. If sample moments converge for large numbers of sample sizes, it is useful to take the "stabilized" values as approximations of the true moments due to the law of large numbers. In view of the finiteness of sample moments, models with infinite moments may seem problematic for applications. However, as Mandelbrot (1997) emphasized, the concept of infinite moments is not "improper" as sample moments do not always converge or may tend to infinity as the sample size increases. Therefore, the finiteness of sample moments cannot be used to exclude the possibility of using models with infinite moments (see also p. 88–92 of Mandelbrot (1997)). In practice, of course, one can avoid infinite moments by either putting bounds to infinite-mean models or replacing their tail parts with lighter tails. As a cost, these corrections may give varying or even misleading results.

Infinite-mean models are by no means strange or pathological. For an inexperienced reader, let us consider a uniformly distributed random variable U on $[0, 1]$, or a normally distributed random variable X with mean 0 and variance 1, both very common in any probabilistic or statistical analysis. What may not be obvious, although mathematically simple to verify, is that $1/U$ and $1/X$ have infinite mean, although they are very natural objects. (Note that X takes the value 0 with probability 0, so the reciprocal $1/X$ is well-defined.)

Among models with infinite moments, those with infinite means are particularly useful in risk management, especially for modeling catastrophic risks and operational risks. Section 3 contains a summary of various examples of infinite-mean models in risk management and related areas. In this paper, we collect some recent results and phenomena of infinite-mean models across different

research domains. It turns out that the assumption of having a finite mean is crucial for many classic results to hold; many of these results are typically flipped for infinite-mean models. We also emphasize that several commonly used methods in statistics, decision analysis, and risk management are not useful for infinite-mean models anymore. For instance, the classic central limit theorem requires a finite second moment and sample moments will not converge for infinite-mean models. Therefore, if infinite-mean models are the outcomes of careful statistical analysis, standard methods should be used with considerable caution or possibly avoided.

The rest of the paper is organized as below. Section 2 gives several examples of power-law distributions. Section 3 exemplified the occurrences of infinite-mean models in many studies. Standard estimation methods of power-law models are briefly described in Section 4. In Section 5, we point out several statistical, decision, and risk models that may fail for infinite-mean models. Sections 6–9 discuss the applications of infinite-mean models in risk aggregation, optimization and risk sharing, multiple hypothesis testing, optimal bundling, firm growth analysis, and linear estimators, respectively. Section 10 concludes the paper.

2 Definitions and properties

Many useful models applied to heavy-tailed phenomena fall into the class of power-law distributions. A random variable X is said to follow a *power-law distribution* (or a regularly varying distribution) with tail parameter $\alpha > 0$, if

$$\mathbb{P}(X > x) = \frac{L(x)}{x^{\alpha}},$$

where L is a *slowly varying* function, that is, $L(tx)/L(x) \to 1$ as $x \to \infty$ for all $t > 0$. We say a power-law distribution is *extremely heavy-tailed* if $\alpha \leq 1$. A distribution is said to be an *infinite-mean distribution/model* if its absolute moment of order 1 is infinite (i.e., a random variable X follows an infinite-mean distribution if $\mathbb{E}[|X|] = \infty$). We are particularly interested in those models for X with $\mathbb{E}[X_+] = \infty$, where $X_+ = \max\{X, 0\}$, for their financial interpretation. We present below some important examples of power-law distributions, all of which are infinite-mean models if their tail parameters are no greater than 1.

The most popular power-law model is arguably the *Pareto distribution*, with its distribution function defined as

$$\mathbb{P}(X \leq x) = 1 - \left(\frac{\theta}{x}\right)^{\alpha}, \quad x \geq \theta,$$

where $\alpha > 0$ is the tail parameter and $\theta > 0$ is the scale parameter. The nth moment of X is infinite if and only if the tail parameter α is in $(0, n]$; in particular, $\mathbb{E}[X] = \infty$ if $\alpha \in (0, 1]$. This parametrization of the Pareto distribution is known as the Pareto distribution of Type I; see Arnold (2015) for other parametrizations of Pareto distributions and the related distributional properties.

An important distribution from the Extreme Value Theory (EVT) is the generalized Pareto distribution (GPD). The *generalized Pareto distribution* with parameters $\xi \in \mathbb{R}$ and $\beta > 0$ is defined as

$$\mathbb{P}(X \leq x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ e^{-x/\beta}, & \text{if } \xi = 0, \end{cases} \tag{1}$$

where $x \in [0, \infty)$ if $\xi \geq 0$ and $x \in [0, -\beta/\xi)$ if $\xi < 0$. If $\xi > 0$, the generalized Pareto distribution is a power-law distribution with a tail parameter being $1/\xi$. If $\xi \geq 1$, then the generalized Pareto distribution has infinite mean. By the Pickands-Balkema-de Haan Theorem (Balkema and de Haan, 1974; Pickands, 1975), the generalized Pareto distributions are the only possible non-degenerate limiting distributions of the excess of random variables beyond a high threshold.

Another important power-law distribution in the EVT literature is the *Fréchet distribution* with the distribution function given by

$$\mathbb{P}(X \leq x) = \exp(-x^{-\alpha}), \quad x > 0,$$

where $\alpha > 0$ is the tail parameter. If $\alpha \leq 1$, then $\mathbb{E}[X] = \infty$. By the Fisher-Tippett-Gnedenko Theorem, the Fréchet distribution (with location-scale transforms) is one of the three possible non-degenerate limiting distributions of the maximum of iid random variables (the limiting distributions are referred to as the extreme value distributions). In particular, if $X_1, X_2, \ldots$ are iid regularly varying random variables, then there exist sequences of constants $\{d_n\}$ and $\{c_n\}$ where $c_n > 0$ such that $(\max\{X_1, \ldots, X_n\} - d_n)/c_n$ converges to the Fréchet distribution (up to location-scale transforms).

Two other examples of extremely heavy-tailed power-law models are the *Cauchy distribution* and the *Lévy distribution*. The standard Cauchy cumulative distribution function is

$$\mathbb{P}(X \leq x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}, \quad x \in \mathbb{R},$$

and the standard Lévy cumulative distribution function is

$$\mathbb{P}(X \leq x) = 2 - 2\Phi\left(\frac{1}{\sqrt{x}}\right), \quad x \geq 0,$$

where Φ is the cumulative distribution function of the standard normal distribution. The tail indices of the Cauchy distribution and the Lévy distribution are 1 and 0.5, respectively. The expectation of the Cauchy distribution does not exist and that of the Lévy distribution is ∞.

The Cauchy distribution is also well-known via its density function, given by

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

This function is also known as "the witch of Agnesi", getting the name from a book of Maria Gaetana Agnesi published in 1748. In any college-level probability courses, the Cauchy distribution is the first example of an infinite-mean distribution, for which the laws of large numbers fail.

Both the Cauchy and the Lévy distributions are special cases of a large class of power-law distributions called stable distributions. The characteristic function of a *stable distribution* with stability parameter $\alpha \in (0,2)$, skewness parameter $\beta \in [-1,1]$, scale parameter $\sigma = 1$ and shift parameter $\mu = 0$, is

$$\varphi(\theta) = \begin{cases} \exp\left(-|\theta|^\alpha (1 - i\beta\operatorname{sgn}(\theta)\tan\frac{\pi\alpha}{2})\right) & \text{if } \alpha \neq 1, \\ \exp\left(-|\theta|(1 + i\beta\frac{2}{\pi}\operatorname{sgn}(\theta)\log|\theta|)\right) & \text{if } \alpha = 1, \end{cases} \quad \theta \in \mathbb{R},$$

where $\operatorname{sgn}(\cdot)$ is the sign function. As the name suggests, a linear combination of iid stable random variables follows the same distribution of the iid random variables, up to location-scale transforms. Another important property of stable distributions is that they are the only possible limiting distributions of sums of iid random variables; see Samorodnitsky (2017) and Nolan (2020) for properties of stable distributions. Stable distributions are usually defined via their characteristic functions because they generally do not have analytical cumulative distribution functions. The few examples with analytical cumulative distribution functions include the normal distribution, the Cauchy distribution ($\alpha = 1$, $\beta = 0$), and the Lévy distribution ($\alpha = 1/2$, $\beta = 1$). Stable distributions with $\alpha < 1$ and $\beta = 1$ are defined on the positive axis and have infinite mean. A stable distribution has infinite absolute expectation if its tail parameter is less than or equal to 1.

3   Empirical evidence in finance and insurance

Compared with distributions with infinite second or higher moments, statistical models with infinite mean are less common in finance and insurance literature and thus need more motivation. Recall that financial asset return data typically show features of finite mean and infinite 4th or 5th

moments; see the summary of Cont (2001). Moreover, Mandelbrot (1997) provided many examples of infinite second moments. Nevertheless, there are still many contexts where infinite-mean models are useful. In the following discussion, we delve into examples in the literature leading to extremely heavy-tailed power-law models. Our list is only for a simple illustration, and it is by no means exhaustive.

Catastrophic losses

With standard seismic theory, Ibragimov et al. (2009) illustrated that the tail indices α of earthquake losses lie in the range $[0.6, 1.5]$. The tail indices α for some wind catastrophic losses are around 0.7, as shown by Rizzo (2009). Chen et al. (2024b) reported that the tail indices α of the wildfire suppression costs in Alberta, Canada are around 0.85. The tail indices of losses caused by nuclear power accidents are observed in the range of $[0.6, 0.7]$ by Hofert and Wüthrich (2012); similar observations were made by Sornette et al. (2013).

Operational losses

By analyzing data collected by the Basel Committee on Banking Supervision, Moscadelli (2004) reported the tail indices α of operational losses in 8 different business lines to lie in the range $[0.7, 1.2]$. Notably, 6 out of the 8 tail indices are less than 1, with 2 being significantly less than 1 at a 95% confidence level. See Nešlehová et al. (2006) for a detailed discussion on the risk management consequences of infinite-mean operational losses. Cyber risk losses exhibit tail indices $\alpha \in [0.6, 0.7]$, as explored by Eling and Wirfs (2019) and Eling and Schnell (2020).

Large insurance losses

Several significant commercial property losses collected from two Lloyd's syndicates have tail indices α considerably less than 1; see Biffis and Chavez (2014). Analyzed by Beirlant et al. (1999), some fire losses collected by the reinsurance broker AON Re Belgium have tail indices α approximately equal to 1. Chen et al. (2024b) showed that the tail parameter α for the marine losses of a French private insurer is around 0.9. Most major damage insurance losses in a standard Swiss Solvency Test document (FINMA (2021, p. 110)) have default tail parameter α in the range $[1, 2]$, with $\alpha = 1$ attained by some aircraft insurance.

Financial data

According to Silverberg and Verspagen (2007), the tail indices of financial returns from some technological innovations are less than 1. The tail part of cost overruns in information technology projects can have tail indices $\alpha \leq 1$, as discussed by Flyvbjerg et al. (2022). The tail indices of detected fraud sizes in firms were estimated to be 0.65 by Cheynel et al. (2024) who also proposed a theoretical mechanism to explain the heavy-tailed phenomenon. Gabaix et al. (2003) showed that the market value of the managed assets in top mutual funds has tail indices around 1.

Other infinite-mean power laws

Word frequencies, city sizes, and firm sizes follow Zipf's law ($\alpha \approx 1$); see Zipf (1949), Gabaix (1999), and Axtell (2001). Interconnected agglomerations of firms in Italy have a tail index around 1; see Andriani and McKelvey (2007). Casualties in major earthquakes and pandemics modelled by Pareto distributions also have tail indices less than 1; see Clark (2013) and Cirillo and Taleb (2020). War sizes and intensities fit well with infinite-mean power law models (Clauset et al. (2009) and Cederman (2003)). Both peak gamma-ray intensity of solar flares and the number of religious followers have tail indices around 0.8 (Clauset et al. (2009)).

4 Identifying infinite mean using extreme value theory

Since data are finite but our interested models have no finite mean, it is important to verify whether a given dataset fits with a model with infinite mean. This is typically done by estimating the tail parameter. It needs to be emphasized that if the possibility of infinite-mean models cannot be excluded, one should consider using statistical methods based on order statistics rather than those based on moments. Below we only present some basic estimation methods for the tail parameter and we refer readers to Embrechts et al. (1997), de Haan and Ferreira (2006), and Reiss and Thomas (2007) for comprehensive discussions and more advanced methods.

Standard methods to estimate the tail parameter of power-law distributed data include the classic Hill method and the GPD method from the EVT (see McNeil et al. (2015) and Embrechts et al. (1997)). Another estimation method is the log-log rank-size estimation (see Ibragimov et al. (2015) and Gabaix and Ibragimov (2011)). In this section, we briefly describe the Hill method and the GPD method; we refer to Embrechts et al. (1997) and McNeil et al. (2015) for more details.

Let $X_1, \ldots, X_n$ be iid non–degenerate random variables following a power-law distribution with tail parameter $\alpha > 0$. Denote the order statistics of $X_1, \ldots, X_n$ by $X_{n,n}, \ldots, X_{1,n}$, from the

smallest to the largest (i.e., $X_{n,n} \leq \cdots \leq X_{1,n}$). The Hill estimator of α, denoted by $\hat{\alpha}_{k,n}$, is defined as

$$\hat{\alpha}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \log X_{j,n} - \log X_{k,n} \right)^{-1}, \quad 2 \leq k \leq n. \tag{2}$$

Thus the Hill estimator is based on the k largest order statistics of an iid sample; some variations of the Hill estimator replace k in (2) by $k-1$. The value of k is generally small and is chosen from the Hill plot which gives the Hill estimates $\hat{\alpha}_{k,n}$ for different values of k. If there is a stable region in the Hill plot where the Hill estimates are close for different k, then the Hill estimates are obtained from the stable region. In practice, one can roughly choose k such that $k/n \in [0.01, 0.05]$ (see McNeil et al. (2015)). We take the Hill plot for wildfire suppression costs in Alberta, Canada as an example (Figure 1 (b) of Chen et al. (2024b)); see Figure 1. By choosing a threshold around the top 5% order statistics of the sample, we can see a stable region after the threshold in Figure 1 and the Hill estimate is around 0.847.
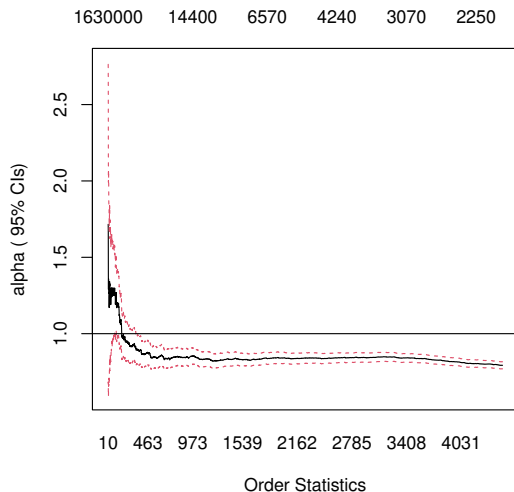

Figure 1: The Hill plot for wildfire suppression costs: The Hill estimates are plotted as black curves with the 95% confidence intervals being red curves. Adapted from Chen et al. (2024b).

The Hill method works well for data whose tail distribution is close to a perfect power function but it can also be misleading in some cases. For instance, the Hill plot for AT&T stock data in Figure 5.8 of McNeil et al. (2015) does not have a stable region. It is also known that the Hill method can be sensitive to the serial dependence in the data. Asymptotic properties of the Hill estimator can be found in Embrechts et al. (1997).

The GPD method assumes that the excess distribution of $X_1, \ldots, X_n$ over some high threshold u is a GPD distribution (1). The tail parameter of the GPD distribution can be estimated by the

maximum likelihood method and the probability-weighted moments method (see Embrechts et al. (1997)).

However, there is a common method for GPD distribution that does not work for the infinite-mean case. To diagnose whether a GPD distribution is a good fit for the excess distribution with threshold u, the sample mean excess function is often used. For $v > 0$, a sample mean excess function is defined as

$$e_n(v) = \frac{\sum_{i=1}^n (X_i - v)\mathbb{1}_{\{X_i > v\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > v\}}}.$$

If a GPD distribution is a good fit, then $e_n(v)$ should be "linear" for $v \geq u$ given large n; this is because the mean excess function $e(v) = \mathbb{E}[X - v | X > v]$, $v > 0$, is linear in v if X is a GPD random variable. Such a graphic diagnosis will not work if the GPD distribution has infinite mean. Klar (2023) proposed a new diagnosis method without moment restrictions.

5 Decision and statistical models that need extra care

5.1 The St. Petersburg paradox

The most famous example of an infinite-mean model in decision making is perhaps the St. Petersburg paradox. Although it is well known, we briefly discuss it here before moving to a more formal analysis.

Consider a lottery that yields a payoff of 2^n dollars with probability 2^{-n}, for $n = 1, \ldots, \infty$. This lottery can be achieved easily by throwing a fair coin many times until a tail has been observed, and a payoff of 2^n is made to the player if it takes n rounds to reach a tail (i.e., there are $n - 1$ heads before this round). It is easy to observe that the expected payoff is $\sum_{n=1}^{\infty} 2^{-n} 2^n = \infty$. Since this payoff has an infinite mean, for a risk-neutral player the fair price of the lottery should be ∞; however, few people are willing to pay more than 20 dollars to play this game, thus the paradox.

Daniel Bernoulli's answer in 1738 (translated version in Bernoulli (1954)) to this paradox is to use a logarithm utility function to model the decision maker, which leads to a finite expected utility of the lottery. However, it is easy to design another lottery with an infinite expected utility for the logarithm utility function. This creates the same trouble as in the original paradox: In theory, the player should prefer such a lottery, and even a tiny fraction of such a lottery (i.e., multiplying the payoff by a very small $\epsilon > 0$), to an arbitrarily large constant payoff, which is clearly paradoxical. There are many other explanations for the paradox, and we omit them. Here, we simply would like to show that an infinite-mean model can lead to controversial situations, even considering a simple lottery.

5.2 Stochastic dominance

Notions of stochastic dominance are commonly used in decision analysis to rank risks; see Levy (1992, 2016) for wide applications of stochastic dominance. Two commonly used stochastic dominance relations are stochastic order and convex order. Let X and Y be two random variables. Then

(i) X is less than Y in *stochastic order*, denoted by $X \leq_{\text{st}} Y$, if $\mathbb{P}(X > x) \leq \mathbb{P}(Y > x)$ for all $x \in \mathbb{R}$;

(ii) X is less than Y in *convex order*, denoted by $X \leq_{\text{cx}} Y$, if $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all convex functions u such that the two expectations exist.

These two notions of stochastic dominance have distinct meanings and provide very different implications in decision makings. In general, $X \leq_{\text{st}} Y$ means that Y is "larger" than X. Indeed, there exists another random variable X' such that X' follows the same distribution of X and $X' \leq Y$ holds almost surely (e.g., Theorem 1.A.1 of Shaked and Shanthikumar (2007)). On the other hand, $X \leq_{\text{cx}} Y$ means that Y is more "spread-out" than X; for instance, it implies $\text{var}(X) \leq \text{var}(Y)$, but generally stronger than that. See Shaked and Shanthikumar (2007) and Müller and Stoyan (2002) for comprehensive treatment of these and other stochastic dominance relations.

When comparing infinite-mean risks, $\leq_{\text{st}}$ order always works whereas $\leq_{\text{cx}}$ may fail because the expectation of any increasing convex transform of a positive infinite-mean random variable is infinity. Both dominance relations work well for finite-mean models but $\leq_{\text{st}}$ order is not very useful in comparing risks with the same mean; e.g., in portfolio selection, it is common that different portfolios have the same mean. If $\mathbb{E}[X] = \mathbb{E}[Y] < \infty$ and $X \leq_{\text{st}} Y$, then X and Y have the same distribution (Theorem 1.2.9 of Müller and Stoyan (2002)). Stochastic order $\leq_{\text{st}}$ provides very strong implications in making decisions as explained below.

5.3 Risk measures and expected utilities

We discuss several decision models that can or cannot be applied to infinite-mean models. A *preference* is a function $\rho : \mathcal{X}_\rho \to \overline{\mathbb{R}} := [-\infty, \infty]$, where the domain $\mathcal{X}_\rho$ is a set of random variables representing risks. For our discussion, we use the term preference in a very broad sense; it includes risk measures in the sense of Artzner et al. (1999) and Föllmer and Schied (2016) and preference models such as the expected utility. For two random losses X and Y, an agent equipped with a preference model ρ prefers X over Y if $\rho(X) \leq \rho(Y)$; for random profits X and Y, the agent prefers

X over Y if $\rho(Y) \leq \rho(X)$. For the rest of this section, assume that all random variables represent losses.

Most commonly used preference models for comparing risks have the following two properties. We write $X \stackrel{\mathrm{d}}{=} Y$ if X and Y have the same distribution.

(a) Law invariance: $\rho(X) = \rho(Y)$ for $X, Y \in \mathcal{X}_\rho$ if $X \stackrel{\mathrm{d}}{=} Y$.

(b) Monotonicity: $\rho(X) \leq \rho(Y)$ for $X, Y \in \mathcal{X}_\rho$ if $X \leq_{\mathrm{st}} Y$.

Decision models that do not respect (a) are often regarded as models for decisions under ambiguity. We do not discuss them and interested readers may refer to Wakker (2010). We will call a preference model satisfying (a) and (b) a monotone risk preference.

If random variables X and Y represent losses, an agent equipped with a monotone risk preference will prefer X over Y as soon as $X \leq_{\mathrm{st}} Y$. We present below some examples of monotone preference models.

The first examples are Value-at-Risk (VaR) and Expected Shortfall (ES), the two regulatory risk measures in the insurance and finance sectors. For a random variable X and $p \in (0,1)$, VaR is defined as

$$\mathrm{VaR}_p(X) = \inf\{t \in \mathbb{R} : \mathbb{P}(X \leq t) \geq p\},$$

and ES is defined as

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

If X has infinite mean, $\mathrm{VaR}_p(X)$ is always finite whereas $\mathrm{ES}_p(X)$ can be infinite and is thus not very useful for infinite-mean models. Many useful risk measures for infinite-mean models are in the class of distortion risk measures such as VaR and Range Value-at-Risk proposed by Cont et al. (2010) and studied by Embrechts et al. (2018); distortion risk measures coincide with dual utilities of Yaari (1987) up to a sign change. By contrast, convex risk measures (Föllmer and Schied (2002)), which have several nice properties, will take infinite values when evaluating infinite-mean losses and hence are not useful (see Filipović and Svindland (2012)).

The next example is the expected utility model. Let u be an increasing function, called a utility function, and $u(w)$ represents the utility of wealth w. Since we are dealing with losses, we flip the sign of the expected utility model, by using an increasing function v (called a disutility function) given by $v(x) = -u(-x)$, and denote the expected utility model by E_v, that is,

$$E_v(X) = \mathbb{E}[v(X)] = \mathbb{E}[-u(-X)].$$

An expected utility agent is said to be *risk-averse* (in the sense of Rothschild and Stiglitz (1970)) if u is concave (v is convex). Note that risk-averse expected utility may not be useful for comparing infinite-mean risks as E_v on them can take infinite values due to the concavity of u. However, there are still some useful examples of expected utility models such as (i) the Markowitz utility function of Markowitz (1952b) (u has a convex-concave structure on the loss side); (ii) the utility function in the cumulative prospect theory of Tversky and Kahneman (1992) (u is convex below a reference point); (iii) agents with limited liability (u is bounded from below).

Although risk-averse expected utility may be infinite for extremely heavy-tailed risks, it does not necessarily mean that there is no way to distinguish the risks for agents with concave utilities. Instead of calculating their expected utility, an alternative way to compare risks is the "overtaking criterion" used by optimal accumulation problems in economics; see von Weizäscker (1965). This "overtaking criterion" is used to avoid the problem of infinite expected utility. The key observation here is that, whenever $\mathbb{E}[v(X)]$ is well-defined, it can be rewritten as

$$\mathbb{E}[v(X)] = \int_0^1 v(F_X^{-1}(p))\mathrm{d}p,$$

where F_X^{-1} is the left quantile function of X, that is, $F_X^{-1}(p) = \mathrm{VaR}_p(X)$ for $p \in (0,1)$. Applying this idea to compare infinite-mean losses, we propose the following formulation. For two losses X and Y, we may say an agent with model E_v prefers X over Y if

$$\int_0^1 \left(v\left(F_X^{-1}(p)\right) - v\left(F_Y^{-1}(p)\right)\right)\mathrm{d}p < 0. \tag{3}$$

In this form, since $X \leq_{\mathrm{st}} Y$ implies $F_X^{-1} \leq F_Y^{-1}$ on $(0,1)$, we can conclude that if $X \leq_{\mathrm{st}} Y$ then X is preferred to Y, even if $\mathbb{E}[v(X)]$ and $\mathbb{E}[v(Y)]$ are not well-defined.

5.4   Laws of large numbers and the central limit theorem

Some classical results for sums of independent random variables in statistical theory may fail for infinite-mean models. Although this observation is very simple, we would like to emphasize it as it can be overlooked by practitioners.

The first example is the law of large numbers. Let $X_1, X_2, \ldots$ be a series of iid random variables and $S_n = \sum_{i=1}^n X_i$. By the strong law of large numbers, if $\mathbb{E}[X_1] = \mu \in \mathbb{R}$, then $S_n/n \to \mu$ almost surely. That is, for a sample of iid observations, the sample mean will converge to the true expectation. However, if $\mathbb{E}[X_1] = \infty$, then $S_n/n \to \infty$ almost surely (see, e.g., Theorem 2.4.5 of Durrett (2019)). As the sample mean will explode for infinite-mean models, many applications

based on the law of large numbers (e.g., Monte Carlo method) do not work.

Another example is the central limit theorem which concerns the limiting distribution of S_n. Assume that $X_1, X_2, \ldots$ are non–degenerate. If $\mathbb{E}[X_1^2] < \infty$, by the classic central limit theorem, the only limiting distribution of S_n (after normalization) is the normal distribution. This result generally does not hold for heavy-tailed random variables as their second or even first moments are usually infinite. In this case, the limiting distribution of S_n (after normalization) for commonly used heavy-tailed random variables is a stable distribution; see, e.g., Embrechts et al. (1997).

6 Some results on aggregating infinite-mean risks

6.1 Asymptotic super-additivity of VaR under regular variation

Suppose that $X_1, \ldots, X_n$ are n identically distributed risks faced by an insurer. It is well known that if $(X_1, \ldots, X_n)$ follows the multivariate normal distribution, VaR_p is sub-additive for $p \in [0.5, 1)$, i.e.,

$$\mathrm{VaR}_p(X_1 + \cdots + X_n) \leq \mathrm{VaR}_p(X_1) + \cdots + \mathrm{VaR}_p(X_n).$$

This means that constructing a diversified portfolio helps to reduce the capital reserve level of the insurer. More generally, the sub-additivity of VaR holds if $(X_1, \ldots, X_n)$ follows the elliptical distribution. For the above claims, see, e.g., Theorem 8.28 of McNeil et al. (2015).

The above benefit from diversification however does not always exist outside the elliptical world. For instance, if $X_1, \ldots, X_n$ are iid and follow a power-law/regularly varying distribution with tail parameter $\alpha > 0$, VaR is asymptotically super-additive for $\alpha < 1$, i.e.,

$$\lim_{p \to 1} \frac{\mathrm{VaR}_p(X_1 + \cdots + X_n)}{\mathrm{VaR}_p(X_1) + \cdots + \mathrm{VaR}_p(X_n)} > 1.$$

The above inequality flips for $\alpha \geq 1$ and VaR is said to be asymptotically sub-additive. Therefore, diversification may increase the insurer's risk if $X_1, \ldots, X_n$ have infinite mean. The asymptotic properties of VaR also hold if $X_1, \ldots, X_n$ are dependent, e.g., multivariate regularly varying $(X_1, \ldots, X_n)$ (see Embrechts et al. (2009) and Mainik and Rüschendorf (2010)). We will extend the discussion of portfolio diversification to a non-asymptotic sense in the following sections.

Besides the marginal behaviors of risks, their dependence structure is also crucial to the diversification effects. For instance, if power-law risks $X_1, \ldots, X_n$ are positively dependent via some Archimedean copula (see Nelsen (2006)), Alink et al. (2004) showed that for infinite-mean (resp. finite-mean) risks, a stronger dependence makes the portfolio $X_1 + \cdots + X_n$ "smaller"

13

(resp. "larger") in the tail part. A similar observation for multivariate regularly varying risks is made in Chapter 13 of Rüschendorf (2013).

An early observation of the detrimental effects of diversification was made by Fama and Miller (1972). It was shown that for infinite-mean stable models $X_1, \ldots, X_n$, diversification makes the portfolio more "spread-out", i.e., $(X_1 + \cdots + X_n)/n$ becomes more dispersive as n increases. This was reasserted by Ibragimov (2005) via probability inequalities, which will be discussed later.

6.2 Stochastic dominance for extremely heavy-tailed risks

Since the pioneering work of Markowitz (1952b), diversification has been regarded as an efficient tool to reduce portfolio risks. In particular, if risks are iid and have a finite mean, diversification will make the portfolio less "spread-out", as studied by Samuelson (1967) and Fama and Miller (1972). However, such results may not hold for infinite-mean risks, as we discussed in Section 6.1.

The comparison among diversified portfolios can be made in a stronger sense. Recently, several comparisons of infinite-mean models have been made via stochastic dominance relations and probability inequalities. These results suggest that diversification will make a portfolio of infinite-mean risks "larger" or more "spread-out".

Let $X_1, \ldots, X_n$ be iid extremely heavy-tailed Pareto random variables. For a nonnegative vector $(\theta_1, \ldots, \theta_n)$, Chen et al. (2024a) showed

$$\theta X_1 \leq_{\mathrm{st}} \theta_1 X_1 + \cdots + \theta_n X_n \quad \text{where } \theta = \sum_{i=1}^n \theta_n. \tag{4}$$

Intuitively, the left-hand side of (4) is a non-diversified portfolio on X_1 with total exposure θ and the right-hand side of (4) is a diversified portfolio with exposure θ_i allocated to X_i for all $i = 1, \ldots, n$. If $X_1, \ldots, X_n$ represent losses, (4) suggests that any decision maker who prefers less loss will choose non-diversification, leading to a *diversification penalty* (diversification will be preferred if $X_1, \ldots, X_n$ represent profits). This result is "unexpected" in the sense that (i) it holds in the strongest form of risk comparison, and (ii) it does not hold for finite-mean models; if $X_1, \ldots, X_n$ have finite mean and $\min_i \theta_i > 0$, (4) holds if and only if $X_1, \ldots, X_n$ are almost surely equal (Proposition 2 of Chen et al. (2024a)). Inequality (4) also holds if $X_1, \ldots, X_n$ are weakly negatively associated super-Pareto random variables. The class of super-Pareto distributions includes special cases of the generalized Pareto, the Burr, the paralogistic, and the log-logistic distributions, all with infinite mean, as studied by Chen et al. (2024a).

With the notion of majorization order, Chen et al. (2024c) generalized (4) by allowing comparison between different diversified portfolios of iid extremely heavy-tailed Pareto random variables.

For two vectors $(\theta_1, \ldots, \theta_n)$ and $(\eta_1, \ldots, \eta_n)$ in $\mathbb{R}^n$, $(\theta_1, \ldots, \theta_n)$ is dominated by $(\eta_1, \ldots, \eta_n)$ in *majorization order* if

$$\sum_{i=1}^{n} \theta_i = \sum_{i=1}^{n} \eta_i \quad \text{and} \quad \sum_{i=1}^{k} \theta_{(i)} \geq \sum_{i=1}^{k} \eta_{(i)} \quad \text{for } k \in [n-1],$$

where $\theta_{(i)}$ and $\eta_{(i)}$ represent the ith smallest order statistics of $(\theta_1, \ldots, \theta_n)$ and $(\eta_1, \ldots, \eta_n)$, respectively; see Marshall et al. (2011). Let $(\theta_1, \ldots, \theta_n)$ and $(\eta_1, \ldots, \eta_n) \in \mathbb{R}_+^n$ be the exposure vectors of two portfolios such that $(\theta_1, \ldots, \theta_n)$ is smaller than $(\eta_1, \ldots, \eta_n)$ in majorization order; the portfolio with exposure vector $(\theta_1, \ldots, \theta_n)$ is more diversified than that with exposure vector $(\eta_1, \ldots, \eta_n)$. Chen et al. (2024c) showed the following inequality

$$\eta_1 X_1 + \cdots + \eta_n X_n \leq_{\text{st}} \theta_1 X_1 + \cdots + \theta_n X_n. \tag{5}$$

Similar to (4), if $X_1, \ldots, X_n$ represent losses, (5) means more diversification leads to a worse portfolio in the sense that diversification makes the portfolio loss "larger" (the observation is flipped if $X_1, \ldots, X_n$ represent profits). Inequality (5) also holds if $X_1, \ldots, X_n$ has a specific type of positive dependence structure modelled by Clayton copulas. By comparison, if $X_1, \ldots, X_n$ are iid with finite mean, by Theorem 3.A.35 of Shaked and Shanthikumar (2007), we have

$$\theta_1 X_1 + \cdots + \theta_n X_n \leq_{\text{cx}} \eta_1 X_1 + \cdots + \eta_n X_n. \tag{6}$$

Hence, diversification makes a portfolio of iid finite-mean losses less "spread-out".

Let $(\theta_1, \ldots, \theta_n)$ and $(\eta_1, \ldots, \eta_n)$ be as before and $X_1, \ldots, X_n$ be iid extremely heavy-tailed stable random variables. By properties of stable distributions, Ibragimov (2005) showed that,

$$\mathbb{P}(\eta_1 X_1 + \cdots + \eta_n X_n \geq x) \leq \mathbb{P}(\theta_1 X_1 + \cdots + \theta_n X_n \geq x) \quad \text{for } x > 0, \tag{7}$$

and the above inequality is flipped if $x < 0$. This result suggests that diversification makes the portfolio more "spread-out", distinct from (5); similar observations can be found in Fama and Miller (1972, p. 271). In particular, (7) implies that (5) holds if $X_1, \ldots, X_n$ are iid positive one-sided stable random variables with infinite mean. On the other hand, if $X_1, \ldots, X_n$ are iid stable random variables with finite mean, the above inequalities will flip; that is, diversification will make the portfolio less "spread-out". The result for finite-mean stable distributions continues to hold for symmetric log-concave distributions (Proschan (1965)); a distribution with density is said to be *log-concave* if its density is log-concave. Inequality (7) also holds for convolutions of risks with joint

α-symmetric distributions (Ibragimov (2009)). Similar probability inequalities for bounded stable random variables are studied by Ibragimov and Walden (2007).

6.3 The sum of infinite-mean risks can be a constant, but not unique

The sum of infinite-mean risks can have strange behavior compared to finite-mean risks. Here we discuss one example that is relevant to risk management.

Suppose that n random variables $X_1, \ldots, X_n$ follow the same distribution F, and they may be dependent. If for some constant $C \in \mathbb{R}$ it holds that $\mathbb{P}(\sum_{i=1}^{n} X_i = nC) = 1$, then we say that C is a center of the distribution F. If $X_1, \ldots, X_n$ have the same finite mean $\mu \in \mathbb{R}$, it is clear that $C = \mu$, since $\mathbb{E}[\sum_{i=1}^{n} X_i] = n\mu$. However, if the mean of $X_1, \ldots, X_n$ does not exist or is infinite, this becomes unclear. A question stated by Wang (2015) was whether the center C of F is unique. This question is answered by Puccetti et al. (2019) for Cauchy distributions. They showed that for every $n \geq 2$, there exist n standard Cauchy random variables adding up to a constant nC if and only if $|C| \leq \log(n-1)/\pi$. Hence, the center is not unique.

The existence of C is formally studied through the notions of complete and joint mixability (Wang and Wang (2011, 2016)), which can be used as a building block for many optimization problems such as minimizing the variance or finding the worst-case value of risk measures in risk aggregation; see Embrechts et al. (2013); Embrechts et al. (2014). The relevance of this problem to risk management is treated in Chapter 8 of McNeil et al. (2015) and its recent progresses are summarized in the monograph Rüschendorf et al. (2024).

7 Infinite-mean models in optimization and risk sharing

We next discuss the implications of infinite-mean models in some optimization problems in risk management.

7.1 Portfolio optimization

Suppose that an agent equipped with a risk preference model ρ faces several iid profits $X_1, \ldots, X_n$ and needs to maximize their preference:

$$\text{To maximize} \quad \rho\left(\sum_{i=1}^{n} \theta_i X_i\right) \quad \text{subject to} \quad \theta_1, \ldots, \theta_n > 0 \quad \text{and} \quad \sum_{i=1}^{n} \theta_i = \theta,$$

where $\theta > 0$ represents the total budget of the agent. What can we say about the agent's optimal strategy? This question has been well studied for finite-mean models since the work of

Markowitz (1952a), and the optimal strategy clearly depends on the agent's risk attitude. If the profits have finite mean, the optimal strategy for risk-averse expected utility agents is full diversification (i.e., $(\theta_1, \ldots, \theta_n) = (\theta/n, \ldots, \theta/n)$) whereas that for risk-seeking expected utility agents is non-diversification (i.e., $(\theta_1, \ldots, \theta_n) = (\theta, \ldots, 0)$); see Samuelson (1967). Indeed, an expected utility agent prefers to diversify over iid finite-mean profits if and only if they are risk-averse (Theorem 4 of Chen et al. (2024c)).

For profits with infinite mean (e.g., profits from technology innovations; see Silverberg and Verspagen (2007)), agents' risk attitudes may not affect their optimal decisions. Assume that the profits are iid extremely heavy-tailed Pareto random variables. In this case, as long as an agent is equipped with a monotone preference model, by (5), full diversification will maximize the agent's preference; see Chen et al. (2024c). Consequently, for risk-averse expected utility agents facing iid profits that follow a Pareto distribution, the optimal strategy is always full diversification; this does not depend on whether the profits have finite mean or not.

Next, let us consider the case where the agent faces iid losses $X_1, \ldots, X_n$. The optimization problem now becomes:

$$\text{To minimize} \quad \rho\left(\sum_{i=1}^{n} \theta_i X_i\right) \quad \text{subject to} \quad \theta_1, \ldots, \theta_n > 0 \quad \text{and} \quad \sum_{i=1}^{n} \theta_i = \theta.$$

If $X_1, \ldots, X_n$ have finite mean and the agent is equipped with a risk measure that is consistent with $\leq_{\mathrm{cx}}$ order,[1] by (6), full diversification is preferred. If $X_1, \ldots, X_n$ are extremely heavy-tailed Pareto risks, the optimal decision is clearly no diversification, as for any monotone risk measure ρ, by (4), we always have $\rho(\theta X_1) \leq \rho(\theta_1 X_1 + \cdots + \theta_n X_n)$. One important consequence is the super-additivity of VaR:

$$\mathrm{VaR}_p(\theta_1 X_1) + \cdots + \mathrm{VaR}_p(\theta_n X_n) \leq \mathrm{VaR}_p(\theta_1 X_1 + \cdots + \theta_n X_n) \quad \text{for all } p \in (0, 1).$$

See McNeil et al. (2015) for the consequences of the super-additivity of VaR. Note that this super-additivity result holds for all significance levels and is thus different from those on regularly varying distributions, which are in some asymptotic sense (Embrechts et al. (2009)). For agents with concave utility functions, although the expected utility model may take infinite value for extremely heavy-tailed risks, the agents can instead use the "overtaking criterion" (3) which implies no diversification is still preferred.

Another interesting case is when the iid losses $X_1, \ldots, X_n$ follow a symmetric stable distribution

---

[1] A risk measure ρ is consistent with $\leq_{\mathrm{cx}}$ order if for $X \leq_{\mathrm{cx}} Y$, $\rho(X) \leq \rho(Y)$.

with tail parameter less than or equal to 1. If the agent is equipped with VaR_p where $p \in [0.5, 1)$, Ibragimov (2009) showed that no diversification minimizes their risk measure. This result is flipped if $p \leq 0.5$ or the tail parameter of the stable distribution is greater than 1; in either case, full diversification is optimal.

7.2 Risk sharing

Risk sharing is the allocation of risks among multiple parties and has been a fundamental tool for insurance companies to mitigate risks and build resilience. In the presence of heavy-tailed risks such as catastrophic losses, however, this standard method may fail and insurance companies may choose not to share risks or even not to participate in the market; see, e.g., Ibragimov et al. (2009).

Chen et al. (2024b) studied the effects of risk sharing in a risk exchange model. In the risk exchange economy, each agent is equipped with some position of an individual risk and the risks are iid. The agents can exchange their risks with other agents by paying/charging premia. The goal of each agent is to minimize their risk assessment after risk exchange.

We fix some notations. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the risk vectors for n agents. The initial exposure vector of agent i, denoted by $\mathbf{a}^i$, satisfies $\mathbf{a}^i \cdot \mathbf{X} = a_i X_i$ where $a_i > 0$; that is, agent i has an initial position $a_i X_i$. Let $\mathbf{w}^i \in \mathbb{R}_+^n$ be the exposure vector of agent i on $\mathbf{X}$ after risk sharing and $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{R}_+^n$ be the premium vector for $\mathbf{X}$. The loss position of agent $i \in [n]$ after risk sharing is

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

Agent $i \in [n]$ is equipped with a risk measure ρ_i and a convex cost function c_i with $c_i(0) = 0$. The risk assessment of agent i is $\rho_i\left(L_i(\mathbf{w}^i, \mathbf{p})\right) + c_i(\|\mathbf{w}^i\| - \|\mathbf{a}^i\|)$. An *equilibrium* of the market is a tuple $\left(\mathbf{p}^*, \mathbf{w}^{1*}, \ldots, \mathbf{w}^{n*}\right) \in (\mathbb{R}_+^n)^{n+1}$ if the following two conditions are satisfied.

(a) Individual optimality:

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

(b) Market clearance:

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

If the risks are super-Pareto random variables (i.e., extremely heavy-tailed), even if there is some risk exchange at equilibrium, the agents will merely exchange their entire positions with each other; hence there is no risk sharing. A similar observation is made by Ibragimov et al. (2011) based

on a different model. Moreover, the premia for all risks are the same. By contrast, if the risks have finite mean and the agents use ES as their risk measure, the agents will share risks proportionally according to their initial risk exposures in an equilibrium. The premia in this case are not the same as well.

The situation becomes quite different if the agents with initial infinite-mean risks can transfer their risks to some external agents. The external agents do not have any risks in the initial phase and can be regarded as institutional investors in the market model. Similarly, an equilibrium is attained in this setting if all agents, including the external agents, can minimize their risk assessments. If the external agents have a stronger risk tolerance than the internal agents, both parties may benefit by transferring losses to the external agents at an equilibrium. Nevertheless, there is still no risk sharing, i.e., no agent will hold more than one risk at the equilibrium.

In practice, due to the limited liabilities of insurers, one may bound each loss below some positive threshold. If the risks are positive random variables (e.g., Pareto risks), the bounded risks will have finite mean. In the model of Chen et al. (2024b), agents equipped with ES will prefer diversification over bounded risks. However, as studied in a different economic model of Ibragimov et al. (2009), agents may not always prefer diversification of bounded risks. Their results show that insurers may fall into a "nondiversification trap" for bounded extremely heavy-tailed losses; that is, even if insurers can benefit from high-level diversification, they may still prefer non-diversification as a low degree of diversification is harmful (the value of diversification is U-shaped in the number of risks).

8   Infinite-mean models in multiple hypothesis testing

Infinite-mean risk models have recently been a popular tool in multiple hypothesis testing, a statistical problem that is relevant to many fields including risk management, such as testing the suitability of the credit rating of structured finance products from ratings of different tranches.

In the context of multiple hypothesis testing, a fundamental question is how to combine several p-values, or test statistics in general, into one p-value. A random variable P is called a *p-variable* if $\mathbb{P}(P \leq p) \leq p$ for all $p \in (0,1)$; realizations of p-variables are called p-values. For our discussions, it suffices to assume that the p-variables to merge are uniformly distributed on $(0,1)$, denoted by $U_1, \ldots, U_n$.

Many useful methods to combine p-values are via the following generalized weighted averaging function

$$M_\phi^{\mathbf{w}}(p_1, \ldots, p_n) := \psi\left(w_1\phi(p_1) + \cdots + w_n\phi(p_n)\right), \tag{8}$$

where $\mathbf{w} = (w_1, \ldots, w_n) \in [0,1]^n$ with $\sum_{i=1}^n w_i = 1$, $\phi : [0,1] \to [-\infty, \infty]$ is a continuous strictly monotonic function, and ψ is its inverse. Several previous methods in the class of (8) use quantile function F^{-1} or $F^{-1}(1-\cdot)$ as ϕ to transform p-values, where F is a cumulative distribution function; that is, transformations $\phi(U_1), \ldots, \phi(U_n)$ follow distribution F. Some examples are

(i) Fisher's combination method (Fisher (1948)): $\phi(x) = -\log(x)$ for $x \in [0,1]$, and $\mathbf{w} = (1/n, \ldots, 1/n)$;

(ii) the harmonic mean method (Wilson (2019)): $\phi(x) = 1/x$ for $x \in [0,1]$;

(iii) the generalized mean method (Vovk and Wang (2020)): $\phi(x) = x^r$ for $x \in [0,1]$, where $r \in [-\infty, \infty]$ (for $r \in \{-\infty, 0, \infty\}$, the averaging function is obtained by taking limits, and this includes the classic Bonferroni correction with $r = -\infty$);

(iv) the Cauchy combination method (Liu and Xie (2020)): $\phi(x) = \tan\left(\pi\left(x - \frac{1}{2}\right)\right)$ for $x \in [0,1]$.

Fisher's combination method and the generalized mean method with $r \geq 0$ utilize light-tailed distributions (the exponential and beta distributions) to transform p-variables. In contrast, heavy-tailed transformations (the Pareto and Cauchy distributions) are used by the harmonic mean method, the Cauchy combination method, and the generalized mean method with $r < 0$.

For different choices of combining functions $M_\phi^{\mathbf{w}}$ and dependence assumptions of p-variables, a threshold function g is assigned such that the test is valid; that is $\mathbb{P}(M_\phi^{\mathbf{w}}(U_1, \ldots, U_n) \leq g(p)) \leq p$, $p \in (0,1)$; see Chen et al. (2023) for thresholds of the above methods given different dependence assumptions. However, the dependence assumption of p-variables is very difficult to verify and the validity of a test procedure cannot be guaranteed if there is any model misspecification on the dependence assumption. For instance, Fisher's combination method assumes p-values are independent and its type-I error rate can explode even if the model misspecification is small. Recently, the harmonic mean and the Cauchy combination methods have received much interest due to their nice properties against model misspecification on dependence assumptions. Coincidentally, both methods use infinite-mean distributions to transform p-variables and their performances are similar and closely connected to the famous Simes method (Simes (1986)) in several senses; see Chen et al. (2023) for the comparisons of the three methods. Other combining methods based on heavy-tailed transformations are studied by, e.g., Gui et al. (2023) and Fang et al. (2023). It is by now clear that transformations to an extremely heavy-tailed distribution have various advantages over other choices, making the methods of Wilson (2019) and Liu and Xie (2020) quite popular. Under independence or some other nice dependence models, the threshold g for the Cauchy combination method can be chosen as precisely $g(p) = p$, $p \in (0,1)$, and the one with the harmonic mean often

needs to be corrected substantially; see Chen et al. (2024d). These studies show how infinite-mean models are useful in statistics even when the underlying problem does not have such a feature.

9   Infinite-mean models in some other settings

We briefly discuss, in a few different contexts, how infinite-mean models lead to different phenomena, compared to usual finite-mean models. In each of these settings, the corresponding conclusion flips when switching from finite-mean models to infinite-mean models.

9.1   Optimal bundling

Product bundling is a strategy to sell goods as collections. It turns out that the valuation distributions may play an important role in determining the optimal bundling strategies. We briefly describe the Vickrey auction model studied by Palfrey (1983) below. Consider a single seller who provides m goods to n buyers. Let 2^M be the power set of $M = [m] := \{1, \ldots, m\}$. A *bundling decision* $\beta = \{B_1, \ldots, B_l\}$ is a partition of $[m]$. For $j \in [n]$, the jth buyer's valuations for goods are represented by $\mathbf{X}_j = (X_{1j}, \ldots, X_{mj})$ where X_{ij} is their valuation for good $i \in [m]$ and $X_{1j}, \ldots, X_{mj}$ are iid. The valuations of buyers are independent of each other, i.e., $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent. Each buyer does not know the valuations of the other buyers; the only available information is their own valuations and the distributions of the other buyers' valuations. For a bundle $B \in 2^M$, its valuation of buyer j is $v_j(B) = \sum_{i \in B} X_{ij}$. In the Vickrey auction, all buyers will simultaneously submit sealed bids for bundles of goods. The buyer with the highest bid wins the goods and pays the second highest bid. Buyer $j \in [n]$ with realized valuations $\mathbf{X}_j = \mathbf{x}_j \in \mathbb{R}^m$ has the following expected surplus for bundle $B \in 2^M$

$$\mathbb{E}[S_j(B, \mathbf{x}_j)] = \mathbb{P}\left(\max_{s \in [n]/\{j\}} v_s(B) < v_j(B)\right)\left(v_j(B) - \mathbb{E}\left(\max_{s \in [n]/\{j\}} v_s(B) \Big| \max_{s \in [n]/\{j\}} v_s(B) < v_j(B)\right)\right).$$

The expected surplus for a bundling decision is the sum of the expected surplus for each bundle. A buyer prefers a bundling decision if its expected surplus is larger.

In a market of two buyers, Palfrey (1983) showed that separate auctions ($\beta = \{\{1\}, \ldots, \{m\}\}$, i.e., one good is sold at a time) are preferred by buyers regardless of the valuation distributions. In the case of more than two buyers, the optimal bundling to the buyers depends on the valuation distributions. If the valuations are bounded, Palfrey (1983) showed that the buyers never prefer separate auctions. However, if the valuations are positive stable random variables with infinite mean, separate auctions are preferred, as shown in Theorem 4.1 of Ibragimov and Walden (2010).

9.2 Firm growth analysis

Ibragimov (2014) reconsidered the demand-driven innovation and spatial competition overtime model of Jovanovic and Rob (1987) in the presence of extremely heavy-tailed customer signals. We briefly describe below the setting studied by Ibragimov (2014), a simplified version of the original model.

In each period, a firm has two decisions to make. In period t, the products of the firm are differentiated by a parameter $\hat{\theta}_t \in \mathbb{R}_+$. There exists an "ideal" product $\theta_t \in \mathbb{R}_+$ but this information is not available to the firm until they decide $\hat{\theta}_t$. Once the firm commits to $\hat{\theta}_t$ and learns the "ideal" product θ_t, the price of the product, denoted by $p(\hat{\theta}_t, \theta_t)$, is determined. Then, the firm needs to choose the level of output y_t. After that, the firm will receive customers' signals about next period's "ideal" product $s_{i,t} = \theta_{t+1} + \epsilon_{i,t}$, $i = 1, \ldots, N$, where $\epsilon_{i,t}$, $i = 1, \ldots, N$, are iid shocks and N is a random sample size with a known distribution given y_t. The signals will be used to estimate next period's "ideal" product θ_{t+1}, via, e.g., the sample mean of the signals $\hat{\theta}_{t+1} = \sum_{i=1}^{N} s_{i,t}/N = \theta_{t+1} + \sum_{i=1}^{N} \epsilon_{i,t}/N$. To choose y_t, the firm needs to solve the following problem

$$V(p(\hat{\theta}_t, \theta_t)) = \max_{y_t} \left\{ y_t p(\hat{\theta}_t, \theta_t) - c(y) + \beta V(x) \mathrm{d}F(x|y_t) \right\},$$

where c is a convex cost function, $\beta \in \mathbb{R}$ is a discounting factor, and $F(x|y_t)$ is the distribution function of $p(\hat{\theta}_{t+1}, \theta_{t+1})$ given y_t.

Let y_t^1 and y_t^2 be the outputs of two firms in period t. Denote by $\mathrm{RP}_t = \mathbb{P}(y_{t+1}^1 > y_{t+1}^2 | y_t^2 > y_t^1)$, called the reversal rank probability. Suppose the firms use the sample mean to estimate product design. If the signals follow a log-concave distribution, as implied by the results of Jovanovic and Rob (1987), then $\mathrm{RP}_t \leq 1/2$ and RP_t diminishes as $y_t^2 - y_t^1$ increases. The results also hold if the signals follow a symmetric stable distribution with a tail parameter being strictly greater than 1 (Ibragimov (2014)). These results suggest that if customers' signals are not extremely heavy-tailed, the firms' outputs exhibit some positive persistence (i.e., firms with large outputs are likely to have large outputs in the next period). However, if the signals follow an extremely heavy-tailed symmetric stable distribution, the previous results are reversed and firms' outputs show a pattern of anti-persistence. Ibragimov (2014) also showed that if the sample median is used as the estimator, firms' outputs are persistent regardless of whether the signals are extremely heavy-tailed or not.

9.3   Efficiency of linear estimators

Consider the model

$$X_i = \mu + \epsilon_i, \quad i = 1, 2, \ldots,$$

where μ is the location parameter and ϵ_i, $i = 1, 2, \ldots$, are iid symmetric random noises with mode 0. If ϵ_i, $i = 1, 2, \ldots$, have finite mean, then μ is the population mean. A linear estimator for μ is $\hat{\theta}_{\mathbf{a}} = \sum_{i=1}^{n} a_i X_i$ where $\mathbf{a} = (a_1, \ldots, a_n) \in [0, 1]^n$ and $\sum_{i=1}^{n} a_i = 1$. We write $\hat{\theta}_n := \hat{\theta}_{\mathbf{a}}$ for $\mathbf{a} = (1/n, \ldots, 1/n)$. If ϵ_i, $i = 1, 2, \ldots$, have finite second moments, it is common to use variance to compare two unbiased estimators. However, this approach fails for heavy-tailed noises. We only present here one possible solution that is the most relevant to our discussion. A formal treatment of robust statistical methods can be found in Huber and Ronchetti (2009).

To solve this issue, Ibragimov (2007) proposed to compare estimators via peakedness. For two random variables X and Y, X is said to be more *strictly peaked* about $\theta \in \mathbb{R}$ than Y if $\mathbb{P}(|X - \theta| > \epsilon) < \mathbb{P}(|Y - \theta| > \epsilon)$ for all $\epsilon > 0$. An estimator $\hat{\theta}^1$ is said to be more *P-efficient* than another estimator $\hat{\theta}^2$ if $\hat{\theta}^1$ is strictly more peaked about θ than $\hat{\theta}^2$. If $\hat{\theta}_n$ becomes more P-efficient as n increases, $\hat{\theta}_n$ is said to exhibit *monotone consistency*. Ibragimov (2007) considered the case when ϵ_i, $i = 1, 2, \ldots$, follow a symmetric stable distribution. If the stable distribution has finite mean, it is shown that $\hat{\theta}_n$ is the most P-efficient linear estimator and $\hat{\theta}_n$ exhibits monotone consistency. However, if the stable distribution does not have finite mean, then $\hat{\theta}_n$ is the least P-efficient linear estimator and the P-efficiency of $\hat{\theta}_n$ decreases as n increases. These results hold for more general classes of noise distributions in Ibragimov (2007).

10   Conclusion

Heavy-tailed data, prevalent in a wide range of applications, can violate the assumption of a finite first moment, which is the theoretical foundation of many classic results such as the central limit theorem. This paper reviews a few recent advancements for infinite-mean models and discusses the breakdown of standard thinking due to infinite mean in several research areas including but not limited to portfolio diversification, risk sharing, and multiple hypothesis testing. Most of the reviewed results for infinite-mean models show a significant divergence compared to their counterparts in finite-mean cases, making it necessary to rethink the impacts of traditional methods in the presence of infinite mean. This is illustrated by many examples in this paper. Moreover, many classic methods for finite-mean models are warned to be inappropriate for infinite-mean models. For instance, the regulatory risk measure Expected Shortfall may yield infinite value when applied

to infinite-mean losses, making its meaning as a capital reserve difficult to interpret. Consequently, in the presence of infinite-mean models, decision makers need to take a step back and carefully reconsider the possible consequences of applying classic methods developed in a finite-mean world.

Acknowledgements

References

Alink, S., Löwe, M. and Wüthrich, M. V. (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics*, 35(1):77–95.

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.

Arnold, B. C. (2015). *Pareto Distributions*. Second Edition. CRC Press

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

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

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

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

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36.

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

Cederman, L.-E. (2003). Modeling the size of wars: From billiard balls to sandpiles. *American Political science review*, 97(1):135–150.

Chen, Y., Embrechts, P. and Wang, R. (2024a). An unexpected stochastic dominance: Pareto distributions, dependence, and diversification. *Operations Research*, forthcoming.

Chen, Y., Embrechts, P. and Wang, R. (2024b). Risk exchange under infinite-mean Pareto models. *arXiv:2403.20171*.

Chen, Y., Hu, T., Wang, R. and Zou, Z. (2024c). Dominance between combinations of infinite-mean Pareto random variables. *arXiv:2404.18467*.

Chen, Y., Liu, P., Tan, K. S. and Wang, R. (2023). Trade-off between validity and efficiency of merging p-values under arbitrary dependence. *Statistica Sinica*, 33:851–872.

Chen, Y., Wang, R., Wang, Y. and Zhu, W. (2024d). Sub-uniformity of harmonic mean p-values. *arXiv:2405.01368*.

Cheynel, E., Cianciaruso, D. and Zhou, F. (2024). Fraud power laws. *Journal of Accounting Research*, forthcoming.

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.

Clauset, A., Shalizi, C. R. and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236.

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

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

Durrett, R. (2019). *Probability: Theory and Examples*. Fifth Edition. Cambridge University Press.

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., 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., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking & Finance*, 37(8):2750–2764.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R. and Beleraj, A. (2014). An academic

response to Basel 3.5. *Risks*, 2(1):25-48.

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

Fang, Y., Chang, C., Park, Y. and Tseng, G. C. (2023). Heavy-tailed distribution for combining dependent p-values with asymptotic robustness. *Statistica Sinica*, 33:1115-1142.

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`.

Fisher, R. A. (1948). Combining independent tests of significance. *American Statistician*, **2**(30).

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

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

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

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

Gabaix, X., Gopikrishnan, P., Plerou, V. and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270.

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

Gui, L., Jiang, Y. and Wang, J. (2023). Aggregating dependent signals with heavy-tailed combination tests. *arXiv*: 2310.20460.

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

Huber, P. J. and Ronchetti E. M. (2009). *Robust Statistics*. Second ed., Wiley Series in Probability and Statistics. Wiley, New Jersey. First Ed.: Huber, P. (1980).

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

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

Ibragimov, R. (2007). Efficiency of linear estimators under heavy-tailedness: convolutions of α-symmetric distributions. *Econometric Theory*, 23(3):501–517.

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

Ibragimov, R. (2014). On the robustness of location estimators in models of firm growth under heavy-tailedness. *Journal of Econometrics*, 181(1):25–33.

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.

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

Jovanovic, B. and Rob, R. (1987). Demand-driven innovation and spatial competition over time. *Review of Economic Studies*, 54(1):63–72.

Klar, B. (2023). A Pareto tail plot without moment restrictions. *arXiv:2310.01076*.

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.

Liu, Y. and Xie, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529):393–402.

Mainik, G. and Rüschendorf, L. (2010). On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, 14:593–623.

Mandelbrot, B. B. (1997). *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, New York.

Markowitz, H. (1952a). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Markowitz, H. (1952b). The utility of wealth. *Journal of Political Economy*, 60(2):151–158.

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

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

Moscadelli, M. (2004). The modelling of operational risk: Experience with the analysis of the data

collected by the Basel committee. Technical Report 517. *SSRN*: 557214.

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

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

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.

Nolan, J. P. (2020). *Univariate Stable Distributions: Models for Heavy Tailed Data*. Springer Cham.

Palfrey, T. R. (1983). Bundling decisions by a multiproduct monopolist with incomplete information. *Econometrica*, 51(2):463–483.

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

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

Puccetti, G., Rigo, P., Wang, B. and Wang, R. (2019). Centers of probability measures without the mean. *Journal of Theoretical Probability*, 32(3):1482–1501.

Reiss, R.-D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values*. Third Edition. Springer.

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.

Rüschendorf, L. (2013). *Mathematical Risk Analysis*. Springer Berlin, Heidelberg.

Rüschendorf, L., Vanduffel, S. and Bernard, C. (2024). *Model Risk Management: Risk Bounds Under Uncertainty*. Cambridge University Press.

Samorodnitsky, G. (2017). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Routledge.

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

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.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*,

73(3):751–754.

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. (2007). *The Black Swan: The Impact of the Highly Improbable*. Random House, New York.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of Uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

von Weizäscker, C. C. (1965). Existence of optimal programs of accumulation for an infinite time horizon. *Review of Economic Studies*, 32(2):85–104.

Vovk, V. and Wang, R. (2020). Combining p-values via averaging. *Biometrika*, 107(4):791–808.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, 41(3):808–826.

Wang, R. (2015). Current open questions in complete mixability. *Probability Surveys*, 12:13–32.

Wilson, D. J. (2019). The harmonic mean p-value for combining dependent tests. *Proceedings of the National Academy of Sciences*, 116(4):1195–1200.

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

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley Press, Cambridge.