

From Concentration Profiles to Concentration Maps. New tools for the study of loss distributions

Andrea Fontanari^{a,b}, Pasquale Cirillo^{a,*}, Cornelis W. Oosterlee^{b,a}

^a TU Delft, The Netherlands

^b CWI, The Netherlands

ARTICLE INFO

Article history:

Received November 2016

Received in revised form November 2017

Accepted 9 November 2017

Available online 15 November 2017

JEL classification:

C43

C46

G32

Keywords:

Concentration measures

Value-at-Risk

Expected Shortfall

Concentration Profile

Gini index

ABSTRACT

We introduce a novel approach to risk management, based on the study of concentration measures of the loss distribution. We show that indices like the Gini index, especially when restricted to the tails by conditioning and truncation, give us an accurate way of assessing the variability of the larger losses – the most relevant ones – and the reliability of common risk management measures like the Expected Shortfall. We first present the *Concentration Profile*, which is formed by a sequence of truncated Gini indices, to characterize the loss distribution, providing interesting information about tail risk. By combining Concentration Profiles and standard results from utility theory, we develop the *Concentration Map*, which can be used to assess the risk attached to potential losses on the basis of the risk profile of a user, her beliefs and historical data. Finally, with a sequence of truncated Gini indices as weights for the Expected Shortfall, we define the *Concentration Adjusted Expected Shortfall*, a measure able to capture additional features of tail risk. Empirical examples and codes for the computation of all the tools are provided.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

This paper introduces a way of dealing with tail risk in loss distributions, for risk management purposes, on the basis of a class of tools derived from concentration measures. The objects of study are losses in general, for which we assume data is available, with no particular reference to the source of risk.

Following a common convention in risk management (Hull, 2012), we model losses as a positive random variable Y , bounded or unbounded to the right, with continuous distribution function $F(y)$, restricting our attention to a static framework, in which the (unconditional) loss distribution is given and does not vary over time. The choice of a static approach is not a limitation in many fields of risk management, where the time horizon on which losses are defined is relatively large (Resti and Sironi, 2007), and the dependence among losses is not a major source of concern¹; an example being the one-year credit loss distribution, commonly used by banks under the Basel framework (Basel Committee

on Banking Supervision – BCBS, 2006, 2011) with the so-called historical simulation approach (Hull, 2012).

Value-at-Risk (VaR) and Expected Shortfall (ES) represent two important risk measures in modern risk management (Basel Committee on Banking Supervision – BCBS, 2006, 2011; Embrechts et al., 2003; Hull, 2012; McNeil et al., 2015). Despite their popularity, these measures are not really able to convey reliable information on how losses are dispersed in the tail: it is indeed not difficult to image several distributions sharing the same VaR and ES , but with different risk profiles because of a diverse tail behavior. A measure of the dispersion of the losses beyond VaR is therefore needed, as a way of assessing the precision of the ES in representing tail risk. Empirical studies (Embrechts et al., 2003; Hull, 2012; Kleiber and Kotz, 2003; McNeil et al., 2015) show that losses tend to follow skewed heavy-tailed distributions, in which the right tail is so fat that often the assumption of a finite second moment is too stringent, thus suggesting that measures of dispersion based on the variance should be avoided.

Our proposal is to make use of concentration (or inequality) measures (Eliazar and Sokolov, 2012; Kleiber and Kotz, 2003; Yitzhaki and Schechtman, 2013) to analyze the dispersion of risk in the tail, and we focus our attention on the Lorenz curve (Lorenz, 1905) and the corresponding Gini index (Gini, 1912), for which we

* Correspondence to: Applied Probability Group, EEMCS Faculty, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands.

E-mail address: P.Cirillo@tudelft.nl (P. Cirillo).

¹ Several interesting dynamic approaches dealing with dependence and time evolution also exist, for example in operational risk (Chavez-Demoulin et al., 2016; Moscadelli, 2004), or in market and credit risk (see McNeil et al., 2015, and references therein), but we do not deal with them here.

derive a risk management interpretation.² In particular we show that the Gini index does not only provide a robust measure for the precision of the ES, assessing how losses are dispersed beyond VaR, but it can also be used as an alternative measure for the fat-tailedness of the loss distribution itself.

We show how a given sequence of truncated Gini indices, which we call *Concentration Profile* (CP), can be used to characterize losses, allowing (1) for the identification of parametric families of distributions, and (2) for the observation of features that are not immediately available from data, like the actual behavior of risk in the tail. We provide a full description of the use of the CP, offering quick heuristics for the everyday business, but also more technical applications like goodness-of-fit tests and extreme value theory.

We then introduce the so-called *Risk Concentration Map*, a graphical tool identifying the main risk factors contained in the CP, mapping them into an easily readable plot in which, through the use of a risk/utility function approach (Edwards, 1992), we can attach a concise risk score to every CP. The map can be used to study different loss distributions in terms of their tail risk, comparing portfolios with different scales and magnitudes, given that the proposed approach is scale-free.

Finally, the *Concentration Adjusted Expected Shortfall* ($CAES_\alpha$) is introduced as the product of the Expected Shortfall at confidence level α , i.e. ES_α , and the corresponding truncated Gini index. This quantity proves to be useful in better characterizing tail risk, complementing the information provided by the CP.

The paper is structured as follows: in Section 2 we briefly review some basic concepts about concentration measures, while in Section 3 some common measures of risk used in risk management are analyzed and put in relation with these concentration measures; in Section 4, we introduce and study the Concentration Profile, and in Section 5 we describe the Concentration Map; in Section 6 some additional extensions based on the Concentration Adjusted Expected Shortfall are discussed, whereas in Section 7 empirical results on simulated and actual data are provided; finally Section 8 closes the paper. For the sake of completeness, some Appendices A–E contain the more technical details of our work, like proofs and explanatory calculations. Python codes for the computation of the new tools are also provided as extra online materials.

2. Basic concentration quantities

2.1. The Lorenz curve

Introduced by Max Lorenz in 1905 (Lorenz, 1905), the Lorenz curve is a pivotal tool in the study of economic inequality and the distribution of wealth in the society (Kleiber and Kotz, 2003).

Consider a positive continuous random variable Y , belonging to the \mathcal{L}^1 class, i.e. $\mu = \mathbb{E}(Y) < \infty$, and let $F(y) = \mathbb{P}(Y \leq y)$ be its cumulative distribution function. Define the quantile function of Y as $Q(\alpha) = F^{-1}(\alpha)$, where $F^{-1}(\alpha) = \inf\{y : F(y) \leq \alpha\}$ with $0 \leq \alpha \leq 1$. The Lorenz curve $L(x)$ is formally given by

$$L(x) = \frac{\int_0^x Q(\alpha) d\alpha}{\int_0^1 Q(\alpha) d\alpha}, \quad 0 \leq x, \alpha \leq 1. \quad (1)$$

In terms of wealth, the Lorenz curve reads as follows: for a given $x \in [0, 1]$, $L(x)$ tells us that $x \times 100\%$ of the population owns $L(x) \times 100\%$ of the total wealth. Such an interpretation tells that the Lorenz curve is scale-free: the total amount of wealth is not

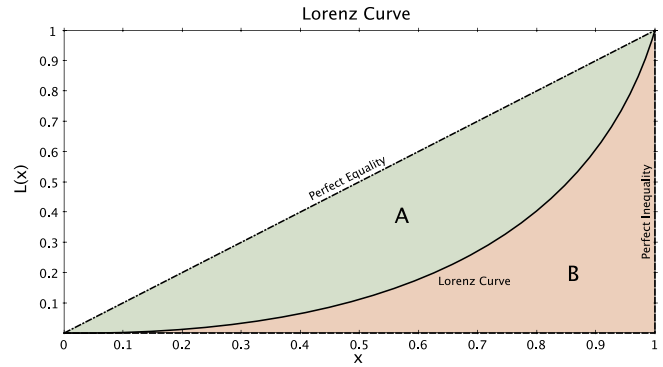


Fig. 1. Graphical representation of a Lorenz curve, of the perfect equality and of the perfect inequality lines. We also show the geometric interpretation of the Gini index as the ratio of area A over $A+B$.

taken into consideration, whereas the way it is distributed among the individuals is the key information.

Mathematically, the Lorenz curve $L : [0, 1] \rightarrow [0, 1]$ defined in Eq. (1) is a continuous, non-decreasing, convex function, almost everywhere differentiable in $[0, 1]$, such that $L(0) = 0$ and $L(1) = 1$. The curve $L(x)$ is bounded from above by the so-called perfect equality curve, i.e. $L_{pe}(x) = x$, and from below by the perfect inequality curve, i.e.

$$L_{pi}(x) = \begin{cases} 0 & 0 \leq x < 1, \\ 1 & x = 1. \end{cases}$$

The perfect equality line L_{pe} indicates the theoretical situation in which everyone possesses the same amount of wealth in the economy, while the perfect inequality line L_{pi} states that only one individual owns all the wealth in the society. A visual representation of a possible Lorenz curve is given in Fig. 1, where we also provide L_{pe} and L_{pi} .

Given its strong relation with the quantile function Q , the Lorenz curve can fully characterize a distribution (Kleiber and Kotz, 2003): if we know the exact functional form of $L(x)$, we can recover $F(y)$ up to a constant. However, graphically, it is not possible to discriminate among distributions just looking at their Lorenz curves (Cirillo, 2013): a curve like that in Fig. 1 may give an indication of how unequal a society is, but it cannot identify by which underlying distribution such inequality is generated.

2.2. The Gini index

Given a Lorenz curve, and following Eliazar and Sokolov (2012), we define a general concentration (or inequality) measure as

$$D_p = \frac{\|L(x) - L_{pe}(x)\|_p}{\|L_{pi}(x) - L_{pe}(x)\|_p}, \quad (2)$$

where the denominator is for normalization purposes, so that

$$0 \leq D_p \leq 1 \quad \forall p \in [1, \infty]. \quad (3)$$

Values of D_p close to 1 indicate relevant variability in the data and high concentration of wealth, while values close to 0 tell the opposite. From Eq. (2) it is clear that, by varying the type of distance $\|\cdot\|_p$, we can define different indices: the Gini index (Gini, 1912) is obtained by fixing $p = 1$, i.e. $G = D_1 \in [0, 1]$.

In the literature there exist many equivalent representations of the Gini index (Eliazar and Sokolov, 2012), here we adopt the following:

$$G = \frac{\mathbb{E}(|Y_1 - Y_2|)}{2\mathbb{E}(Y)} = \frac{\mathbb{E}((Y_1 - Y_2)^+)}{\mathbb{E}(Y)} \in [0, 1], \quad (4)$$

² The use of concentration measures in finance and risk management is not completely new: an interesting application of the Gini index as a substitute of the more common standard deviation is for instance the Mean-Gini efficient portfolio frontier developed Shalit and Yitzhaki (1984, 2005).

where $Y_i, i = 1, 2$ are i.i.d copies of the same random variable $Y \sim F(y)$, and $(Y_1 - Y_2)^+ = \max\{Y_1 - Y_2; 0\}$. Formula (4) defines the Gini index as the normalized average of the distance between two random independent observations taken from the underlying distribution $F(y)$; the Gini index is therefore a measure of variability for a random variable and its realizations, as observed in Shalit and Yitzhaki (1984).

Let us consider two common variance-based measures like the variance-to-mean ratio, defined as:

$$VM = \frac{\mathbb{E}(Y - \mathbb{E}(Y))^2}{\mathbb{E}(Y)},$$

or the well-known coefficient of variation

$$CV = \frac{\sqrt{\mathbb{E}(Y - \mathbb{E}(Y))^2}}{\mathbb{E}(Y)}.$$

Comparing them with the Gini index, we observe the following interesting facts:

- The Gini index is bounded between 0 (perfect equality) and 1 (perfect inequality), while measures like VM and CV are unbounded; being normalized to the unit interval allows for easier comparison and analysis. Also notice that the VM is not scale-free, while the Gini index is; and even if the CV is scale-free, its existence is not guaranteed, as we observe in the next item.
- The Gini index is an \mathcal{L}^1 -measure, meaning that it can be computed for all random variables admitting a finite mean, with no further requirement. This is not true for measures like VM or CV for which the second moment also needs to be finite. This is a restriction when dealing with fat-tailed data, as losses often are Chavez-Demoulin et al. (2016), Fiordelisi et al. (2014), Kleiber and Kotz (2003) and McNeil et al. (2015).
- The Gini index is a quasi-convex measure (Blackorby and Donaldson, 1980). With a little abuse of notation, if $G(x)$ is the Gini index of a data set (x_1, \dots, x_n) from a distribution $F(x)$, and $G(y)$ is the Gini index associated with another distribution $H(y)$ and (y_1, \dots, y_n) , it can be shown that

$$G(\lambda x + (1 - \lambda)y) \leq \max(G(x), G(y)), \quad \lambda \in [0, 1].$$

In words: the Gini index of a data set obtained as linear convex combination of two data sets, e.g. $(\lambda x_1 + (1 - \lambda)y_1, \dots, \lambda x_n + (1 - \lambda)y_n)$, cannot be larger than the largest Gini index from the two original data sets; or, financially, the Gini index of a convex portfolio of losses cannot exceed any of the original Gini indices.

Quasi-convexity is a realistic relaxation of convexity (Cambini and Martein, 2009), with important consequences for sub-additivity and risk diversification, and thus risk management in general, e.g. Frittelli and Maggis (2011), Tian and Jiang (2015) and references therein. In particular, by quasi-convexity we may handle distributions of risk that are not necessarily closed under convolution (Embrechts et al., 2003; Shalit and Yitzhaki, 1984, 2005).

- Differently from measures of dispersion like the VM or the CV, the Gini index does not assume an underlying symmetric structure in the data, and it is therefore more appropriate to study the dispersion of asymmetric random variables like those representing losses (Kleiber and Kotz, 2003). As Eliazar and Sokolov (2012) suggest, the Gini index can also be considered as an \mathcal{L}^1 alternative to the skewness coefficient, for measuring the asymmetry in the data, in particular to the right. As shown in Chotikapanich (2008), the numerator in Eq. (4) moves in the same direction as the skewness coefficient, when the latter is defined (i.e. finite).

In Appendix A, Tables 3 and 4, we have collected the Lorenz curves and the Gini indices of some notable loss distributions, from

the Pareto to the Weibull, together with the parameterizations we use in this work.

3. Basic concepts of risk management

When modeling losses in risk management, it is important to keep in mind two relevant stylized facts observed in the empirical literature:

- When considering losses as nonnegative quantities, the loss distribution is asymmetric and right-skewed (Fiordelisi et al., 2014; Kleiber and Kotz, 2003).
- The loss distribution is usually fat-tailed (Chavez-Demoulin et al., 2016; Embrechts et al., 2003; Kleiber and Kotz, 2003; McNeil et al., 2015), and the Paretianity of the right tail often implies the non-existence of the moments of order greater than or equal to two.³

Given the stylized facts above, one is usually not interested in studying the entire distribution of losses, but rather a part of it, the right tail, where the larger losses concentrate. Most Basel regulations (Basel Committee on Banking Supervision – BCBS, 2006, 2011), but also Solvency II (2009) for insurance companies, deal with the large unexpected losses, the few game-changers, not the many small negligible losses we can easily hedge, thus suggesting to deal with truncated random variables and distributions, rather than with the original ones.

Below we provide some basic quantities that we will use in the rest of the paper.

Definition 1. Given a positive random variable Y with c.d.f. $F(y)$ and p.d.f. $f(y)$, its (left-)truncated version $Y_u = Y|Y \geq u$ has c.d.f.

$$F_u(y) = \frac{F(y) - F(u)}{1 - F(u)}, \quad u \leq y \leq \infty,$$

and p.d.f

$$f_u(y) = \frac{f(y)}{1 - F(u)}.$$

The quantities $F_u(y)$ and $f_u(y)$ are known as exceedance distribution and exceedance density of the random variable Y , respectively.

Definition 2. Given a confidence level $\alpha \in (0, 1)$, the Value-at-Risk (VaR_α) is the statistical quantile of the loss distribution function $F(y)$ defined as

$$VaR_\alpha = \inf\{y \in \mathbb{R} : P(Y \geq y) \leq 1 - \alpha\} = \inf\{y \in \mathbb{R} : F(y) \geq \alpha\}.$$

Definition 3. Given a VaR_α , the Expected Shortfall ES_α , for a positive Y with c.d.f. $F(y)$, is given by

$$ES_\alpha = \mathbb{E}(Y | Y > VaR_\alpha).$$

Interestingly, the ES_α is the mean of the truncated random variable Y_u , when the truncation occurs in $u = VaR_\alpha$. Just notice that

$$\mathbb{E}(Y | Y > u) = \int_u^{+\infty} y f_u(y) dy, \quad u > 0.$$

VaR_α and ES_α are two fundamental measures of risk in modern risk management (Basel Committee on Banking Supervision – BCBS, 2011; Hull, 2012; McNeil et al., 2015). It is well-known that, while the ES_α is a coherent risk measure (positive homogeneous,

³ In the case of operational risk, even the first moment, the mean, may be infinite (Chavez-Demoulin et al., 2016; Cirillo and Taleb, 2016; Moscadelli, 2004), but we ignore this radical case here. Naturally, in such a situation, the Gini index itself would not be defined.

monotone, translation invariant and sub-additive), the Var_α is not, unless we restrict our attention to elliptic loss distributions and co-monotonic portfolios (Embrechts et al., 2003; Resti and Sironi, 2007). Therefore, in the recent years, ES_α appears to be preferred by regulators (Basel Committee on Banking Supervision – BCBS, 2011; Solvency II, 2009), even if both measures have been criticized by experts (McNeil et al., 2015), for their incapacity of dealing with the dispersion of losses in the tails.⁴ It is in fact not difficult to imagine several loss distributions with different tails, but sharing the same Var_α and ES_α values. This is why a measure of dispersion of the losses in the tail beyond the Var_α level is of interest, a measure to understand how reliable the ES_α is in representing the losses above the Var_α threshold.

By construction, a higher value of the Gini index indicates that a larger number of losses are present far in the right tail, while a lower value indicates a distribution of losses which is concentrated around the same values. Therefore, two distributions sharing the same Var_α and ES_α but with different tails are likely to have a different Gini index.

Given our interest for the right tail and the large losses, the Gini index in the formulation above is not optimal, for it takes into account the entire support of the distribution. We need a truncated Gini index, which measures the dispersion above the Var_α , so that we can define a reliable measure of tail risk and ES_α precision.

4. The concentration profile

The Lorenz curve $L(x)$ is not a viable tool in everyday risk management, because $L(x)$ does not provide a unique value, but rather a continuum of information, and a graphical inspection of the Lorenz curve does not identify any particular distribution (Kleiber and Kotz, 2003). Conversely, a concentration measure like the Gini index contains important information about the governing distribution, but it is not sufficient to completely characterize the right tail. There is no one-to-one mapping between the Gini index and the Lorenz curve: the same Gini index may correspond to different Lorenz curves, see Yitzhaki and Schechtman (2013) and Appendix B for more details.

However, the Gini index can be the basic ingredient of a more sophisticated tool, able to discriminate among different tail risk profiles, and to provide a unique characterization of the distribution. We build this tool on a sequence of truncated Gini indices and we call it *Concentration Profile* (CP). We first sort all observations in the sample of losses in increasing order, and then recursively exclude points at the left-hand side, computing the Gini index for the remaining samples. In other words, we truncate the distribution of losses at every single loss, from left to right, obtaining a collection of exceedance distributions, and for each of these distributions we compute the Gini index, which will be a truncated Gini index with respect to the original distribution.

4.1. Mathematical construction

Let us introduce some notation.

Definition 4. Consider a positive random variable Y with distribution $F(y)$ and right endpoint y_F , where $y_F = \sup\{y : F(y) < 1\}$ (infinite or finite but very large), and a confidence level $\alpha \in [0, 1]$. We call a *risk subclass* at level α the truncated support of $F(y)$, defined as

$$S_\alpha = [F^{-1}(\alpha), y_F] = [Q(\alpha), y_F].$$

⁴ As far as ES_α is concerned, there are also issues related to back-testing (Chavez-Demoulin et al., 2014; Du and Escanciano, 2016; Hull, 2012).

As $\alpha \rightarrow 1$, the left point of S_α approaches y_F , therefore $\{S_\alpha\}_{\alpha \in [0, 1]}$ defines a decreasing nested sequence of subsets of the support of the distribution $F(y)$. If a Var_α is defined, we clearly have $S_\alpha = [Var_\alpha, +y_F]$.

For every α , let us define the truncated Lorenz curve $L_\alpha(x)$ as

$$L_\alpha(x) = \frac{L(\alpha + (1 - \alpha)x) - L(\alpha)}{1 - L(\alpha)}, \quad 0 \leq \alpha, x \leq 1, \quad (5)$$

where $L(\cdot)$ is the Lorenz curve. The full derivation of $L_\alpha(x)$ is available in Appendix C.

We can then define the truncated distance index

$$D_\alpha^T = \frac{\|L_\alpha(x) - L_{pe}\|_p}{\|L_{pi} - L_{pe}\|_p} \quad 0 \leq D_\alpha^T \leq 1, \quad (6)$$

and, setting $p = 1$, the truncated Gini index

$$\begin{aligned} G(\alpha) &= 1 - 2 \int_0^1 L_\alpha(x) dx \\ &= 1 - 2 \int_0^1 \frac{L(\alpha + (1 - \alpha)x) - L(\alpha)}{1 - L(\alpha)} dx. \end{aligned} \quad (7)$$

In terms of Var_α and ES_α , Eq. (7) can be re-written as

$$G(\alpha) = \frac{\mathbb{E}(|Y_1 - Y_2| \mid Y_1 \wedge Y_2 > Var_\alpha)}{2ES_\alpha} \in [0, 1].$$

The truncated Gini index $G(\alpha)$ inherits all the properties of the usual Gini index introduced in Section 2, and is therefore a measure of the tail dispersion. High values of $G(\alpha)$, i.e. closer to 1, imply that the losses in S_α are dispersed (e.g. most losses are around Var_α , but a few are far away in the tail, well beyond ES_α), resulting in a persistent thickness of the tail in the subclass, and a lower precision of the corresponding ES_α , while values close to 0 suggest that losses that are less dispersed within S_α , so that ES_α tends to be a reliable measure of the expected tail risk. This way of reading $G(\alpha)$ becomes more clear, if we give $G(\alpha)$ a probabilistic interpretation. Using the Markov inequality, we obtain the following bound, which indicates how higher values of $G(\alpha)$ imply a higher probability of loss dispersion:

$$\mathbb{P}(|Y_1 - Y_2| > 2ES_\alpha) \leq G(\alpha) \leq 1. \quad (8)$$

Given the notion of truncated Gini index $G(\alpha)$, we introduce the Concentration Profile, as follows.

Definition 5. Given a sequence of risk subclasses $\{S_\alpha\}_{\alpha \in [0, 1]}$, we call a *Concentration Profile* (CP) the sequence $\{G(\alpha)\}_{\alpha \in [0, 1]}$.

Graphically speaking, a CP is obtained by plotting $G(\alpha)$ against increasing values of α . The Concentration Profile is therefore contained in the $[0, 1] \times [0, 1]$ square. Through the observation of the behavior of its CP, we may distinguish features of the loss distribution and the associated risk. In Fig. 2 the theoretical CP for four main classes of loss distributions is shown: Pareto, lognormal, Weibull and exponential (where exponential is just a Weibull with shape parameter $\gamma = 1$).

Fig. 2 anticipates an interesting fact, which we further discuss in the next subsection: the Pareto distribution is the only distribution characterized by a constant CP, where the height of the line only depends on the shape parameter of the distribution itself.

4.2. Characterization of the concentration profile

The CP uniquely identifies a distribution: if we know the analytical behavior of $\{G(\alpha)\}_{\alpha \in [0, 1]}$, we can recover the corresponding distribution function $F(y)$. This can be done by exploiting the strict relationship between $G(\alpha)$ and the quantile function $Q(\alpha)$, and thus $F(y)$.

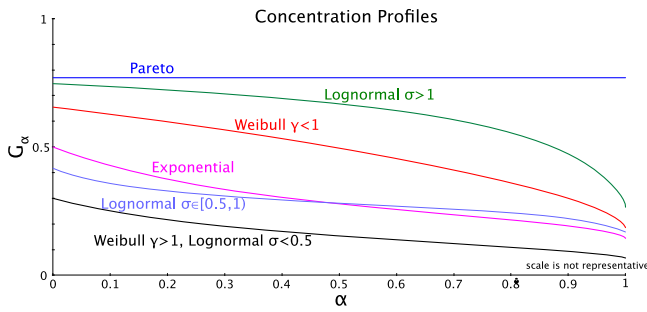


Fig. 2. Theoretical Concentration Profiles for different distributions. Apart from the exponential case, the parameters of the distribution influence the Concentration Profile. For the Pareto distribution, the horizontal line is characteristic, but the height of the line depends on the shape parameter.

Table 1

Theoretical Concentration Profiles for some loss distributions, here $\Gamma(\cdot, \cdot)$ is the incomplete Gamma function $\Gamma(a, b) = \int_b^{+\infty} e^{-x} x^{a-1} dx$ and $\text{erfc}^{-1}(x)$ is the inverse error function.

Distribution	Concentration Profile $G(\alpha)$
Weibull	$G(\alpha) = \frac{-2^{-\frac{1}{\gamma}} \Gamma(1 + \frac{1}{\gamma}, -2 \log(1-\alpha)) - (-1+\alpha) \Gamma(1 + \frac{1}{\gamma}, -\log(1-\alpha))}{(1-\alpha) \Gamma(1 + \frac{1}{\gamma}, -\log(1-\alpha))}$
Lognormal	$G(\alpha) = \frac{\int_0^1 (-1+2x-\alpha) e^{-\sqrt{2\sigma} \text{erfc}^{-1}(2x)} dx}{(1-\alpha) \int_0^1 e^{-\sqrt{2\sigma} \text{erfc}^{-1}(2x)} dx}$
Exponential	$G(\alpha) = \frac{1}{2-2 \log(1-\alpha)}$
Pareto	$G(\alpha) = \frac{1}{(2\rho-1)}$ for $\rho \in (1, +\infty)$

Theorem 6. If $G(\alpha)$ is differentiable, then

$$Q(\alpha) = \frac{(1-\alpha)G'(\alpha) - G(\alpha) + 1}{(1-\alpha)(1+G(\alpha))} \times \exp\left\{-\int \frac{(1-\alpha)G'(\alpha) - G(\alpha) + 1}{(1-\alpha)(1+G(\alpha))} d\alpha\right\}, \quad (9)$$

where $G'(\alpha) = \frac{dG(\alpha)}{d\alpha}$ and $Q(\alpha)$ is the quantile function of a general distribution.

Theorem 6 is a restatement, using $G(\alpha)$, of the main result in Nair et al. (2002), to which we refer for the proof. Examples of Concentration Profiles, obtained via Eq. (7), are given in Table 1 for different distributions.

Another property of every CP is that it represents a characterizing tool for specific classes of distributions. As anticipated in Fig. 2, the shape of a CP can be used to discriminate among distributions. In order to better understand this use of the CP, we identify some limiting cases for light and fat tails.

We focus on the Generalized Pareto Distribution or GPD (Embrechts et al., 2003), which is a well-known distribution family used in extreme value theory (de Haan and Ferreira, 2006), playing an important role in defining the properties of tails. Formally, $Z \geq 0$ follows a one-parameter GPD if its distribution function H is such that

$$H(z) = \begin{cases} 1 - (1 + \xi z)^{-1/\xi} & \xi \neq 0, \\ 1 - e^{-z} & \xi = 0. \end{cases} \quad (10)$$

As discussed in Embrechts et al. (2003), one can introduce a more flexible location-scale version of the GPD, by replacing the argument z by $(z - \nu)/\beta$, with $\nu \in \mathbb{R}$ and $\beta > 0$. However, for ease of notation, in what follows we refer to Eq. (10) when we speak about the GPD.

The GPD can be shown to be the limit distribution of a large number of distribution tails, provided that (1) we move far to the right (Embrechts et al., 2003; Pickands, 1975), setting a high threshold defining the tail, and (2) some non-restrictive technical conditions are met (Beirlant and Goegebeur, 2004; de Haan and

Ferreira, 2006; Gnedenko, 1943). This is known as the GPD approximation.

Depending on the values of ξ , the GPD encompasses relevant tail scenarios in risk management. For $0 < \xi < 1$, we deal with fat-tailed distributions with unbounded support, and fat-tailed distributions with bounded support but very large and often unknown right endpoint, theoretically less dangerous – having all moments – but in practice as problematic as the first ones, for the right endpoint is hardly observed in the data (we refer to Cirillo and Taleb, 2016, for more details) and not easily estimated (Embrechts et al., 2003). We restrict our attention to $\xi < 1$ because for $\xi \geq 1$ the mean of the GDP is not finite, hence we cannot define the Gini index. For $\xi = 0$, the GPD is conversely an exponential distribution, therefore a thin-tailed non-risky case, which we can use as a benchmark; loss distributions in the $\xi = 0$ class are for example the lognormal and the Weibull. We here omit the case $\xi < 0$, which deals with distributions with a finite, not large right endpoint, like the Beta or the Uniform, not particularly relevant to model losses (de Haan and Ferreira, 2006; Embrechts et al., 2003; Kleiber and Kotz, 2003).

Theorem 7. Consider the GPD, its Concentration Profile is given by the following formula:

$$G(\alpha) = \begin{cases} \frac{\xi}{2 - (1-\alpha)^\xi(-2+\xi)(-1+\xi) - \xi} & \xi \in (0, 1), \\ \frac{1}{2 - 2 \log(1-\alpha)} & \xi = 0. \end{cases} \quad (11)$$

for $\alpha \in [0, 1]$.

For every $\alpha \in [0, 1]$, Eq. (11) tells us that:

- For the light-tailed domain with $\xi = 0$, $G(\alpha)$ exhibits a decreasing behavior.
- For the fat-tailed domain with $0 < \xi < 1$, as the shape parameter ξ grows towards 1, the slope of the CP in Eq. (11) tends to zero. For the special case of a purely fat-tailed distribution like the Pareto, we get the horizontal line in Fig. 2. The Paretian CP is immune to changes in the truncation level α and only depends on the tail parameter ρ (corresponding to $1/\xi$), as also shown in Table 1.

When considering Paretian tails, the CP thus tells that risk does not vary: Paretian losses do maintain their riskiness notwithstanding the VaR_α level we might be interested in, and this risk is naturally higher than the fatter the tail, as also shown in Fig. 3a, where smaller values of ρ (larger values of ξ) correspond to a riskier CP, with a higher constant level of $G(\alpha)$. Conversely, if we look at Fig. 2, the lognormal, the Weibull and the exponential distributions all show a decreasing CP, indicating that losses above a high values of VaR_α tend to be less dispersed. This is consistent with a known result in extreme value theory: for light-tailed distributions, in the $\xi = 0$ limiting case, as $\alpha \rightarrow 1$, $ES_\alpha \approx \text{Var}_\alpha$ (Chavez-Demoulin et al., 2014; Embrechts et al., 2003).

Let us consider in more detail the exponential CP, whose analytical formula is given in Table 1 and the full derivation in Appendix E. It does not depend on any parameter, meaning that all exponential distributions share the same profile. This feature is particularly helpful when we use the CP as a goodness-of-fit tool: since the exponential CP starts with $G(0) = 0.5$, for $\alpha = 0$, and decreases towards zero first convexly and successively concavely, with a point of flex at $\alpha = 0.63$, any actual loss distribution whose empirical $G(0)$ is far away from 0.5 cannot be of the exponential type. This is a useful heuristic we can derive from the CP.

The exponential CP is also useful when dealing with Weibull distributions (the exponential being a Weibull with shape parameter $\gamma = 1$). Any Weibull distribution with shape parameter $\gamma > 1$

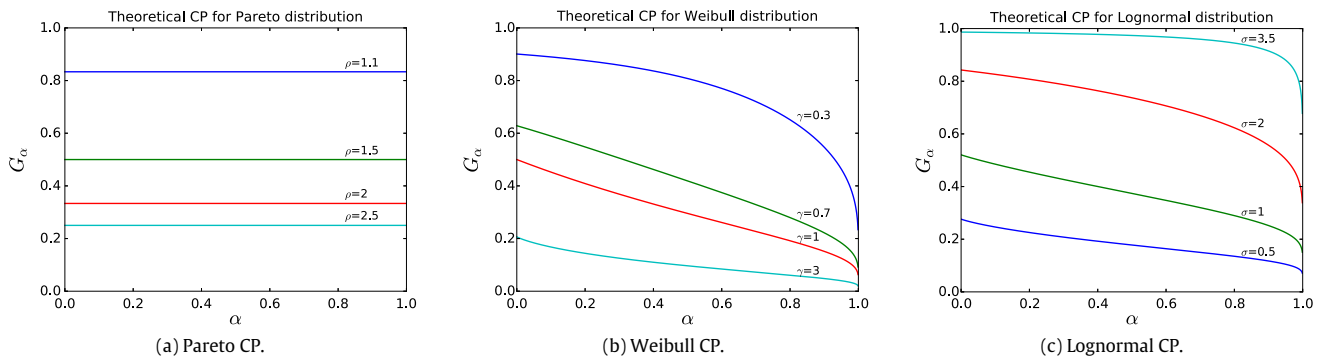


Fig. 3. Theoretical Concentration Profiles for the Pareto, the Weibull and the lognormal distributions, for different values of their relevant parameters.

has a CP which is below the exponential CP, without intersection, while for $\gamma < 1$ the Weibull CP lies above the unique exponential CP. This behavior agrees with theoretical results in distribution theory, pointing out a phase transition in the Weibull distribution between $\gamma < 1$ and $\gamma > 1$ (Kleiber and Kotz, 2003). In terms of risk, when the underlying distribution is of the Weibull type, we can read the values of $\gamma < 1$ as more risky than exponential, and vice versa for $\gamma > 1$. In terms of heuristics, when dealing with data (more details in Section 7), if the empirical CP intersects the exponential distribution CP then the data is not likely from a Weibull distribution.

Regarding the lognormal family, the corresponding CP also has some relationship with the exponential distribution: for $\sigma < 1$, the starting point of the lognormal CP is $G(0) < 0.5$; when $\sigma > 1$, we have $G(0) > 0.5$. Compared to the Weibull CP, the lognormal CP is flatter, indicating that risk tends to decrease more slowly in the tail – in accordance with extreme value theory, where the lognormal distribution is considered more risky than the Weibull distribution (Embrechts et al., 2003; McNeil et al., 2015). To find a lognormal CP which always lies below the exponential CP, without intersection, we need $\sigma < 0.5$. To find a lognormal CP which does not intersect the exponential CP from above, we need $\sigma > 1$.

While the CP may in theory exhibit any type of behavior, for the most used loss distributions (see Kleiber and Kotz, 2003, for a review) the relationship with the truncation level α is non-increasing; a CP with increasing portions is only obtained by mixtures, introducing bumps in the right tail. Additionally, the distributions discussed in this work are representative for many other distributions, an example being the Gamma distribution, whose behavior can be represented by a combination of the Weibull and the exponential.

In Fig. 3, we show the theoretical CP of the Pareto, the Weibull and the lognormal distributions for different values of their relevant parameters.

5. The concentration map

A way of assessing the risk entailed in a given loss distribution is to look at its CP: higher and slowly decreasing values of $G(\alpha)$, for $\alpha \in [0, 1]$, indicate a persistent tail risk, so that the Expected Shortfall may not give reliable information about the risky losses above Value-at-Risk. However, the CP is based on a continuum of points, which is not appealing for risk managers, that prefer concise single measures (just like Var_α and ES_α). In this section we introduce a way of extracting risk information from a CP, creating a tool which simplifies the assessment of risk: the *Concentration Map*.

5.1. Risk drivers

The first step for constructing a Concentration Map for a given loss distribution is to look for the main risk drivers of the corresponding CP, i.e. the quantities summarizing most of the risk.

Definition 8. Given a Concentration Profile $\{G(\alpha)\}_{\alpha \in [0, 1]}$, risk driver $r_1 \in [0, 1]$ is given by its starting value: $r_1 = G(0)$.

The definition of r_1 follows from the fact that, at least for unimodal size distributions, the CP is a non-increasing function of the truncation level α . Given different CPs with same behavior except for their starting point, we expect the one with a higher $G(0)$ to be riskier.

For example, consider the theoretical CP of a Pareto distribution as in Fig. 3a and Table 1. Such a CP is a constant function of parameter $\rho \in (1, +\infty)$, i.e. $G(\alpha) = G(0)$ for every $\alpha \in [0, 1]$. Since $\frac{dG(0)}{d\rho} = -\frac{2}{(-1+2\rho)^2} < 0, \forall \rho \in (1, +\infty)$, we see that in the Paretian case the higher the starting value of the CP, the lower the value of ρ , and the fatter and more risky the loss distribution. The same reasoning holds for any non-increasing CP.

Definition 9. Given a Concentration Profile $\{G(\alpha)\}_{\alpha \in [0, 1]}$, risk driver $r_2 \in [0, 1]$ is given by the difference between the risk driver $r_1 = G(0)$, and $G(\alpha)$, with $\alpha \rightarrow 1$, i.e.,

$$r_2 = \lim_{\alpha \rightarrow 1} |G(0) - G(\alpha)|.$$

From Theorem 7 we know that, for light-tailed distributions, the CP converges to zero as $\alpha \rightarrow 1$. This does not necessarily happen when the loss distribution is fat-tailed and moments can be infinite. Measuring the gap between the initial and final values of the CP is therefore another quantitative way of assessing the fatness of the tail. In particular, the smaller the gap, the larger the mass present in the tail, and the higher the probability of extreme losses. In Fig. 4, we show a graphical representation of r_1 and r_2 .

5.2. The map

Once identified the main risk drivers, we need to combine them to rank different CPs, i.e. different loss distributions, on the basis of their embedded risk. We look for a risk function $R(r_1, r_2) : [0, 1] \times [0, 1] \rightarrow [0, 1]$ which assigns a number in $[0, 1]$ to the combination of risk drivers, summarizing the riskiness of the corresponding CP. This risk function must satisfy the following constraints, given the interpretations of r_1 and r_2 :

$$\frac{dR(r_1, r_2)}{dr_1} > 0 \quad \text{and} \quad \frac{dR(r_1, r_2)}{dr_2} < 0. \quad (12)$$

Let Γ be the set of functions $R(x, y) : [0, 1] \times [0, 1] \rightarrow [0, 1]$ that respect the constraints in Eq. (12). If we additionally impose

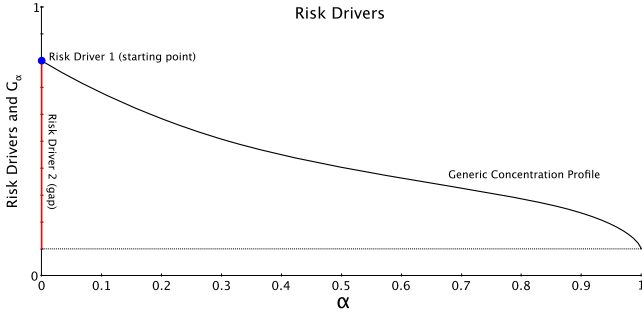


Fig. 4. Risk drivers' visualization with respect to a generic Concentration Profile.

the requirement of a non-increasing CP (i.e. $\frac{dG(\alpha)}{d\alpha} \leq 0$), which is the case for unimodal distributions used in risk management, we reduce the set Γ to functions whose domain is the \mathbb{R}^2 simplex with vertices $[0, 0]$, $[1, 1]$, $[1, 0]$ since, under this additionally constraint, we have $r_1 \leq r_2$.

Looking for a suitable risk function, we rely on well-known results from utility and risk theory, e.g. Edwards (1992). For example, we can choose a Cobb–Douglas risk function,

$$R(r_1, r_2) = r_1^a(1 - r_2)^b \quad \text{with } a, b \in \mathbb{R}^+. \quad (13)$$

This function allows a risk manager to decide which risk driver is relevant, by acting on the parameters a and b . The values of these parameters can be based on historical data or expert judgments. The choice of a Cobb–Douglas is motivated by its convenient properties in terms of risk-driver substitution, and the possibility of deriving isorisk curves that are easy to interpret (Barucci, 2003; Edwards, 1992). Other functions may be used as well.

Figs. 5a and 5b show two examples of Concentration Maps based on the Cobb–Douglas risk function for different parameter sets. Each map reads as follows: on the y-axis we find the values of the first risk driver r_1 , while on the x-axis the values of $\lim_{\alpha \rightarrow 1} G(\alpha)$ are given (in the picture, we choose $\alpha = 0.99$ to have a few observations in the tail). The second risk driver r_2 is computed as the absolute gap, and serves as input in the risk function $R(r_1, r_2) = r_1^a(1 - r_2)^b$, which induces a partial ordering on the Concentration Map. Different combinations of values for the risk drivers may lead to the same output for the risk function, as shown by the colored isorisk curves in the graph.

Looking at the Concentration Maps in Figs. 5a and 5b, we observe some interesting facts:

- Every Pareto distribution with shape parameter $\rho \in (1, \infty)$ lies on the hypotenuse of the map. In particular, a $\text{Pareto}(\rho \approx 1)$ corresponds to the point $(1, 1)$, while a $\text{Pareto}(\rho \rightarrow +\infty)$ tends to $(0, 0)$. This is because the CP associated to the Pareto distribution is constant, so that the risk driver $r_2 = 0$ (since $G(0) = \lim_{\alpha \rightarrow 1} G(\alpha)$), and the x-axis values for the map are the same as those on the y-axis.
- Given their parameter-independent CP, all exponential distributions lie in the cell $(\frac{1}{2-2\log(1-\alpha)}, \frac{1}{2})$, which we can indicate as the exponential cell. The x-coordinate depends on the chosen α level: for $\alpha = 0.99$ the coordinates of the exponential cell become $(0.09, 0.5)$.
- On the basis of the previous two items, we draw the conclusion that any color on the map which is equal to the one placed on the exponential cell or on the Pareto segment will share the same risk, due to the isorisk property of the Cobb–Douglas risk function (Edwards, 1992). For example, the points placed on the hypotenuse, from $(0.33, 0.33)$ to $(1, 1)$ – what we call the transition cells in Fig. 5a and 5b,

correspond to Pareto distributions with $\alpha \in (1, 2]$, whose variance is not finite. The areas of the map sharing the same color of a Pareto distribution with infinite variance will also share the same risk. This is an important feature of the Concentration Map as risk characterization tool.

6. Concentration adjusted expected shortfall

Within a given risk subclass S_α , the quantity $G(\alpha)$ provides a measure of the precision of the ES_α associated to the same subclass. It seems therefore natural to weight the value ES_α by $G(\alpha)$, to define the *Concentration Adjusted Expected Shortfall*, or $CAES_\alpha$.

$G(\alpha)$ belongs to $[0, 1]$, and, for $\alpha \rightarrow 1$, it tends to zero for low-risk light-tailed distributions while it remains constant for Paretian distributions; for light tails the variability of the losses decreases quickly with the truncation level, while it is more persistent for fat tails. Knowing that in general ES_α grows in the truncation level,⁵ a question is whether the increase in the ES_α can be compensated by an increase in its precision, as measured by $G(\alpha)$. A high ES_α is not necessarily a problem, if it also becomes more representative for the tail losses, so that tail risk decreases.

Definition 10. Given a Concentration Profile $\{G(\alpha)\}_{\alpha \in [0,1]}$, and a sequence of Expected Shortfalls $\{ES_\alpha\}_{\alpha \in [0,1]}$, the Concentration Adjusted Expected Shortfall is defined as

$$CAES_\alpha = \{ES_\alpha \cdot G(\alpha)\}_{\alpha \in [0,1]}.$$

The $CAES_\alpha$ thus adjusts the ES_α for the variability of the corresponding risk subclass S_α , and differently from the ES_α sequence, which is increasing for every distribution, it shows a diverse behavior for different distributions. We identify three main situations:

$$\lim_{\alpha \rightarrow 1} CAES_\alpha = \infty, \quad (14)$$

$$\lim_{\alpha \rightarrow 1} CAES_\alpha = c, \quad (15)$$

and

$$\lim_{\alpha \rightarrow 1} CAES_\alpha = 0, \quad (16)$$

with $c \in \mathbb{R}$.

The case in Eq. (14) indicates that the increase in the ES_α is too large to be compensated by a decrease in the variability, or that the variability does not decrease quickly enough. This behavior corresponds for example to the Pareto distribution with any value of $\rho \in (1, +\infty)$, the Weibull distribution with $\gamma < 1$ and lognormal distribution with $\sigma \geq 0.3$. In the second and third cases of Eqs. (15) and (16), the decrease in the variability is sufficient to compensate for an increase of the ES_α , and the value of the ES_α thus becomes representative for the tail losses, as $\alpha \rightarrow 1$. The situation of Eq. (15) occurs for example for the exponential distribution, while that of Eq. (16) manifests itself for Weibulls with $\gamma > 1$ and lognormals with $\sigma < 0.3$.

An extra remark is to be made for the lognormal distribution. Fig. 6b shows how the lognormal $CAES_\alpha$ changes with σ : for $\sigma \in (0.3, 0.7)$, the lognormal $CAES_\alpha$ exhibits a unique u-shaped behavior. This feature can be used in practice to rule out specific values of the lognormal parameter σ when fitting distributions to data.

Fig. 6a presents plots of the theoretical behavior of the $CAES_\alpha$ for the Pareto, the Weibull and the exponential, while Fig. 6b deals with the lognormal.

⁵ For example, it is known that for a Pareto distribution, the quantity $ES_\alpha - VaR_\alpha$ (also known as mean excess) increases linearly in the VaR_α , according to van der Wijk's law (Cirillo, 2013).

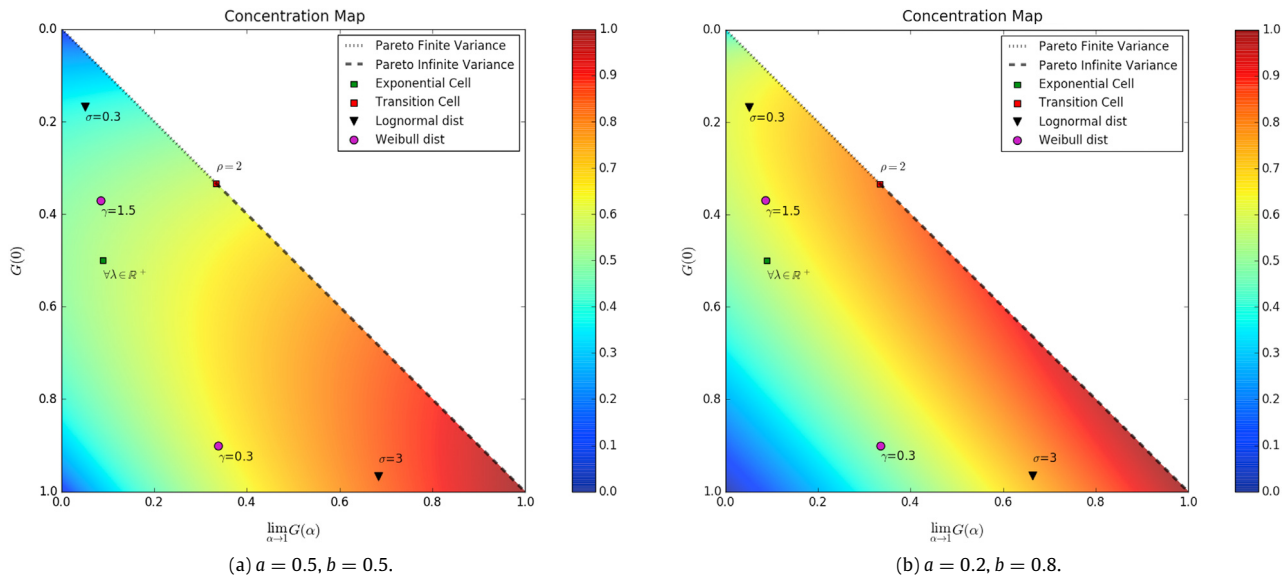


Fig. 5. Concentration Maps, with underlying Cobb–Douglas risk functions of parameters a and b . The different points represent theoretical Concentration Profiles of useful distributions: Pareto, exponential, Weibull and lognormal. $\alpha = 0.99$ has been taken as empirical limit value for $\lim_{\alpha \rightarrow 1} G(\alpha)$.

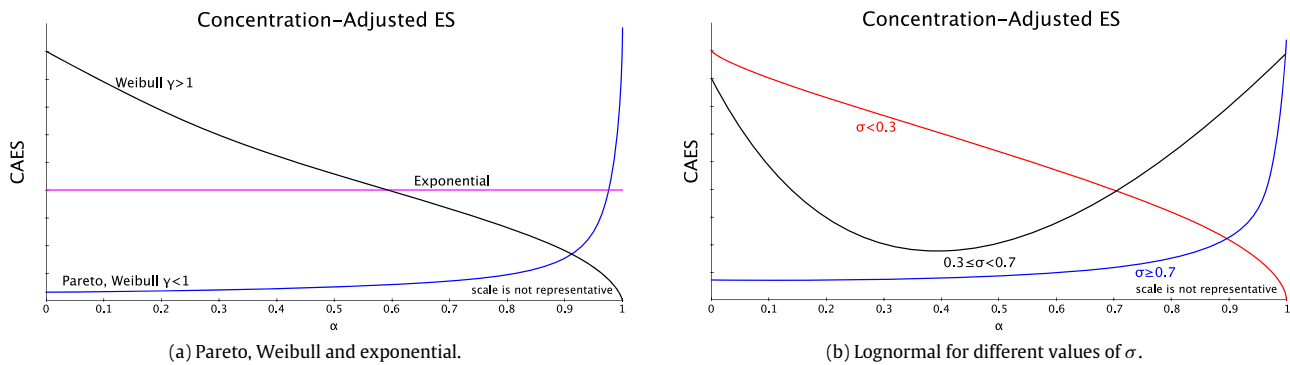


Fig. 6. Theoretical behavior of the Concentration Adjusted Expected Shortfall $CAES_\alpha$ for different distributions.

As we shall see in the next section, the combination of CP and $CAES_\alpha$ can be used to better discriminate among Weibull and lognormal distributions, also guessing the value of their shape parameter, by looking at a limited number of graphs.

7. Applications

In this section we present some applications of the new tools introduced. We start by generating data from two known distributions and we show how the Concentration Profile can be used as a goodness-of-fit tool to identify them. Then, we use two well-known data sets as a basis for our analysis. Finally, in Section 7.3, we show how the Concentration Profile may also find an application in the field of extreme value theory, when one is interested in defining the threshold above which fat tails may manifest themselves.

7.1. Lognormal or Pareto?

We start by generating two samples of size $n = 500$ from a lognormal with parameters $\mu = 0$ and $\sigma = 1$ (finite mean, finite variance) and a Pareto with $\rho = 1.5$ (finite mean but infinite variance). In Fig. 7 the histograms are shown.

Using the empirical Concentration Profile, it is possible to fit the correct distribution. To estimate the empirical CP, notice that

the CP is basically a sequence of Gini indices computed on a decreasing number of ordered observations. Therefore, under normal circumstances, no additional effort must be made in estimation with respect to the classical Gini index, since no bias arises from the truncation procedure. The left point of the domain where the truncation occurs can be successfully estimated by the empirical quantile $\hat{q}_{i:n}$. The estimation of the Concentration Profile (CP) can therefore be performed using any existing estimator for the Gini index $\hat{G}(i : n)$, with n being the total number of observations, and $i \in [1, n - 1]$ the number of data points to be taken into account in the estimation (Yitzhaki and Schechtman, 2013). For example, one could use

$$\hat{G} = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n \sum_{i=1}^n x_i}. \quad (17)$$

After ordering the observations from the smallest to the largest, we obtain the CP by estimating n times the Gini index, each time excluding the first n smallest observations. To get accurate estimates, it is important to leave a sufficient number of observations in the right tail, and we denote this number by k . The number of ordered observations to spare depends on the underlying data generating process. The fatter the tails, the more observations should be left out the estimation cycle, and borrowing heuristics from extreme value theory (Pickands, 1975), we set k between 1% and 5% of the original ordered data points.

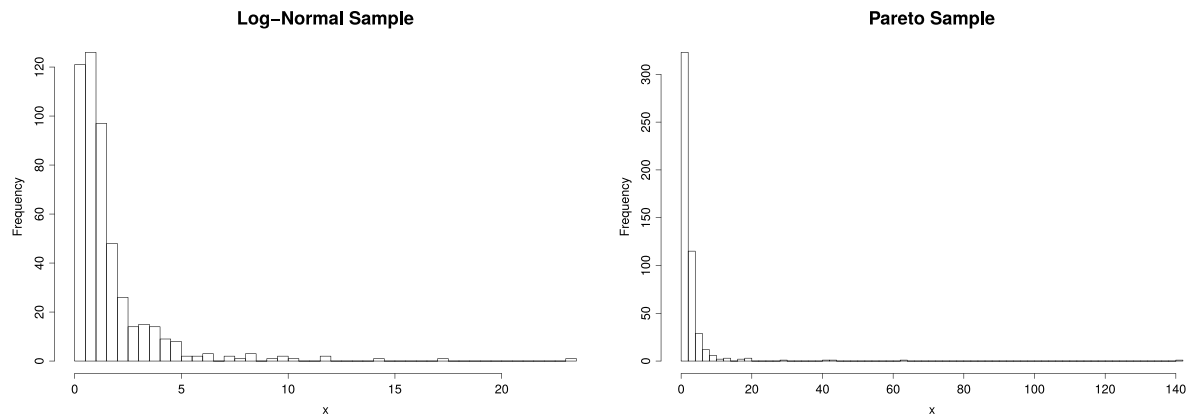


Fig. 7. On the left, the histogram of 500 lognormally distributed data points with $\mu = 0$ and $\sigma = 1$. On the right, the histogram of 500 Pareto distributed data points with $\rho = 1.5$.

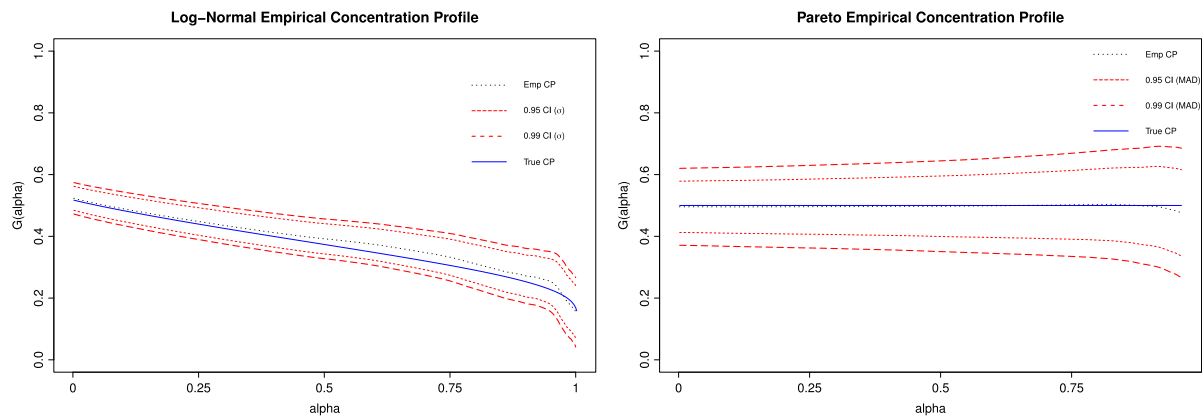


Fig. 8. Empirical Concentration Profile (dotted black line) with 95% and 99% bootstrapped confidence intervals (red dotted lines) compared to the theoretical Concentration Profile (blue straight line). Confidence intervals are built using the standard deviation for the lognormal data and the mean absolute deviation for the Paretian ones. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

When estimating the Gini index, attention must be paid to consistent estimates. According to Gastwirth (1972), the estimator in Eq. (17) (as several others) is consistent and asymptotically normal if and only if the second moment of the data generating process is finite. This assumption can be too stringent in risk management.

From Table 4 in Appendix A we know that the Gini index for the Pareto distribution is a continuous function of the shape parameter ρ . Estimators for the shape parameter of the Pareto distribution exist and do not require any assumption on the finiteness of the second moment (Embrechts et al., 2003). Therefore when we suspect an infinite variance, an estimator for the Gini index (and for the Concentration Profile) can be derived from the tail parameter ρ , by plugging the estimated $\hat{\rho}$ in the theoretical expression of the Gini index. The resulting estimator becomes $\hat{G} = \frac{1}{2\hat{\rho}-1}$, and to get $\hat{\rho}$ one can rely on maximum likelihood estimation (MLE) or other techniques (Beirlant and Goegebeur, 2004; Embrechts et al., 2003). Confidence intervals for the CP can be computed using resampling techniques like bootstrap, jackknife or k -jackknife (Efron and Tibshirani, 1994; Efron, 1982).

Fig. 8 shows the results on the simulated data. In both cases we observe that the theoretical CP is contained in the 95% confidence interval obtained via bootstrap. For both the lognormal and the Pareto distributions, the parameters have been estimated via MLE. For the Paretian data, the (approximate) confidence intervals have been obtained using the mean absolute deviation (MAD, defined as $E|X - E[X]|$ for a generic random variable X), because for these data the theoretical second moment, and hence the standard deviation, is not defined ($\rho = 1.5$). As known, the MAD is a

robust measure of variability, commonly used when the standard deviation is not available (Konno and Yamazaki, 1991; Maronna et al., 2006). Totally compatible results are also obtained by using percentile bootstrap confidence intervals for the Paretian CP. Fig. 8 suggests the usefulness of the CP as a graphical tool for identifying distributions.

To further demonstrate the ability of the CP in discriminating among distributions, we repeat the previous exercise in a simulation experiment, generating two sets of 1000 samples (500 observations) from the same Pareto and lognormal distributions. For each sample we check whether the theoretical CP (with its parameter estimated on the sample) is strictly included in the confidence intervals of the empirical CP (95% and 99%).

For the lognormal case, the empirical coverage is satisfactory, being above 92%, i.e. for at least 92% of times the CP is able to identify the lognormal behavior of data, and no crossing of the confidence intervals is observed. Allowing for small deviations, like touching or slightly crossing the confidence intervals, especially at the right-hand side of the CP, the coverage can be improved, up to the expected percentages.

For the Paretian case, conversely, the coverage is around 75%. This seems unsatisfactory, but it can be justified by the fact that, for $\rho = 1.5$, the variance of the Pareto is not defined, and a sample of 500 observations may not be sufficient to correctly estimate ρ , given the large fluctuations in the samples (Fontanari et al.). Also in this case the empirical coverage can be improved by allowing minor departures from the confidence intervals, but always looking for an almost constant CP. However, and this is the

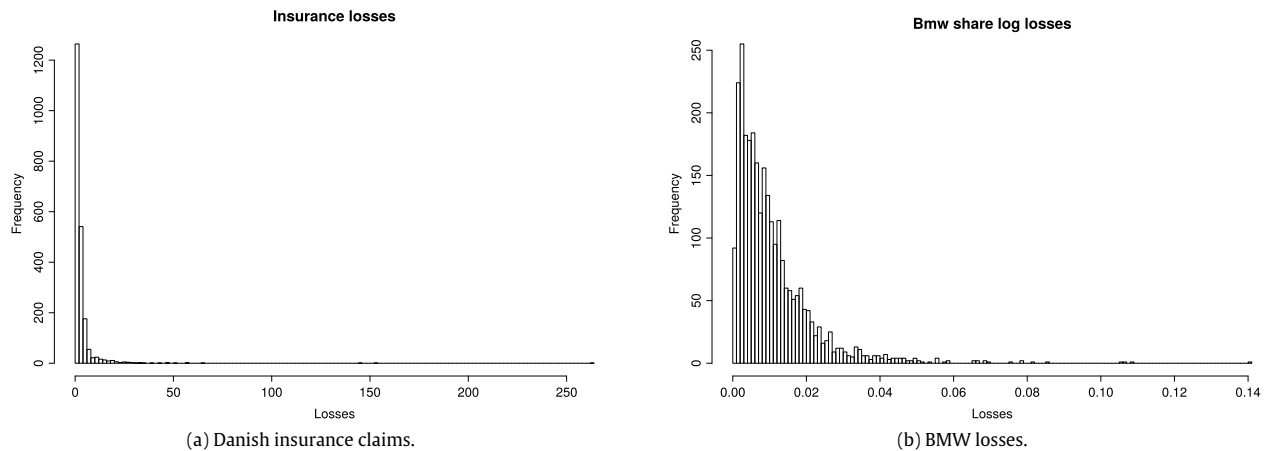


Fig. 9. Histograms of the Danish insurance losses and of the BMW losses.

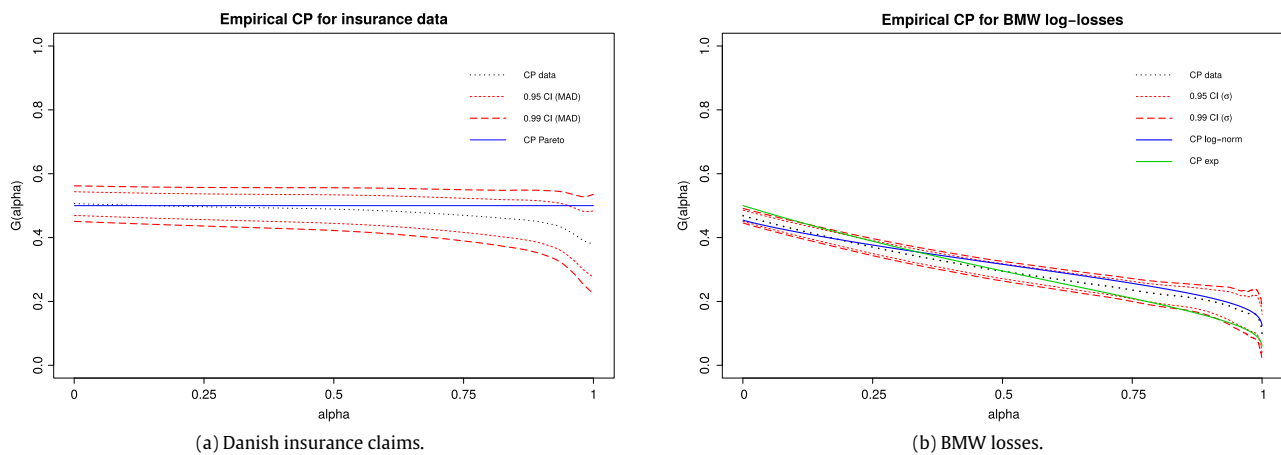


Fig. 10. Empirical Concentration Profiles with 95% and 99% bootstrapped confidence intervals (1000 samples). For the Danish insurance data (left), confidence intervals are built using bootstrapping and the MAD as a robust measure of variability, while for the BMW data (right) we use block-bootstrapping with non-overlapping block of different sizes (size 200 in the figure) and the standard deviation.

most important result, if instead of $\rho = 1.5$ we choose $\rho = 2.1$, so that the theoretical variance is finite, then the empirical coverage for the Pareto reaches the levels of the lognormal case. Similarly if instead of 500 observations, we deal with 2000 or more data points. The quality of the estimated shape parameter $\hat{\rho}$ is therefore fundamental: an unreliable $\hat{\rho}$ can generate a theoretical Paretian CP that is out of the confidence intervals, even if a visual inspection of the CP plot shows a constant empirical CP. This is the most common situation when the theoretical CP does not fall within the bootstrapped confidence intervals, because of an unreliable $\hat{\rho}$: one is still able to observe Paretianity, but she should not rely on the CP for a confirmation about $\hat{\rho}$.

A last experiment is to use the lognormal samples of the simulation study to check whether a mis-specified Weibull CP would fall within the usual confidence intervals. What we get is an empirical coverage of 0%: the theoretical Weibull CP never falls within the confidence intervals, suggesting that the underlying data is not Weibull distributed.

7.2. Real data example

We deal with two well-known data sets in the *evir* package of the statistical language R, used in, e.g., [de Haan and Ferreira \(2006\)](#), [Embrechts et al. \(2003\)](#) and [McNeil et al. \(2015\)](#). The first contains 2167 Danish fire insurance claims (name: danish), and it

is a typical example of a Pareto loss distribution with estimated shape parameter $\rho \in (1, 2)$, so that we expect a finite mean but an infinite variance (here the estimation of ρ is more reliable despite the infinite variance, given the larger sample size). The second data set contains 6146 daily losses and returns on the BMW share price (name: BMW), but here we only consider the 2769 losses, which are known to be lognormally distributed. The empirical loss functions of the data are shown in [Fig. 9](#) as histograms.

[Fig. 10](#) shows the empirical CPs for the Danish insurance and BMW data, together with their bootstrapped 95% and 99% confidence intervals (on the basis of 1000 samples). In the case of Paretian losses (Danish insurance data) the confidence intervals are computed using the MAD. Regarding the BMW data, there is temporal dependence in them ([Embrechts et al., 2003](#)). For this reason, to not completely lose the dependence structure, a simple block-bootstrap approach ([Davison and Hinkley, 1997](#)) can be used,⁶ with non-overlapping blocks of 50, 100 and 200 observations each (the last block only contains the residual observations to be precise), observing no appreciable difference in the final results depending on block size. We have bootstrapped the entire

⁶ More advanced approaches like the moving, the circular ([Davison and Hinkley, 1997](#)) or the stationary block bootstrap ([Politis and Romano, 1994](#)) could have been used, and we actually tried some of them, but our qualitative results about the shape of the CP do not vary.

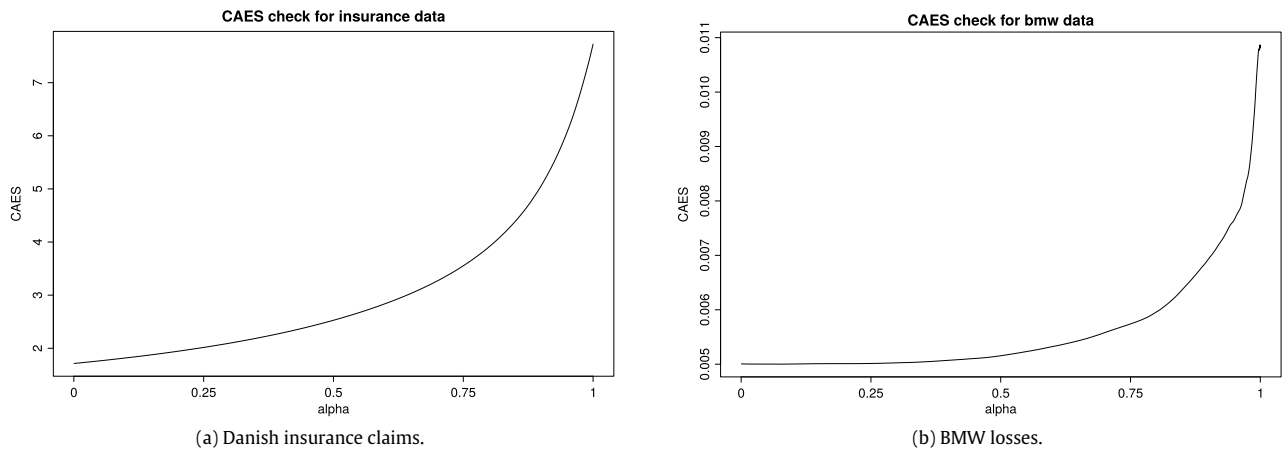


Fig. 11. Empirical Concentration Adjusted ES_α for the Danish insurance claims and the BMW share daily losses.

time series of profits and losses (6146 daily log returns), and only after resampling we have selected the losses (obtaining 1000 data sets of losses of variable size). Just resampling the losses would have introduced non-trivial problems in the data, given that the time series of losses is not equally spaced. These problems of time dependence tend to disappear if we only focus on the very large losses, as extreme value theory teaches us (de Haan and Ferreira, 2006), but it is better not to underestimate them when defining the CP for the lower thresholds.

In Fig. 10a, the empirical CP of the fire insurance claim data is almost constant. Given the insights in Section 4, we conclude that the data is Pareto distributed. An additional check to the Paretianity of the data set can be done using the $CAES_\alpha$, for which we expect a diverging behavior. The result is presented in Fig. 11a.

For the BMW losses, the empirical CP exhibits the following features: it starts around the point 0.5, it decreases rapidly towards zero, and exhibits a change in the slope, being slightly convex at the beginning and ending concave. By the results in Section 4, we can exclude the Pareto distribution since the CP is not constant.

With the starting point of the BMW empirical CP around $G(0) = 0.5$, and recalling that the Weibull distribution starting at this point coincides with the exponential, we can exclude Weibulls with $\gamma \neq 1$ from the candidates. Remaining possibilities are either the exponential or the lognormal with shape parameter $\sigma \rightarrow 1$. To discriminate between these two options we use the $CAES_\alpha$, given that its behavior for these distributions is different: for the exponential the $CAES_\alpha$ is constant, while for the lognormal it exhibits the complex behavior in Fig. 6b. The results are shown in Fig. 11b, where the $CAES_\alpha$ diverges, suggesting lognormally distributed data.

To double check the lognormal hypothesis, we fitted a lognormal distribution to the data, using the method of moments for its parameters, getting $\sigma = 0.88$ (with a s.e. of 0.0065). We then used this value to generate the theoretical lognormal CP in Fig. 10b, comparing it with the empirical one. The result is positive, as we cannot observe any statistically significant difference.

Paretian data are usually much more risky than light-tailed ones. However, comparing the Var_α and the ES_α of the Danish and the BMW data sets is useless since they measure different phenomena with different scales. Moreover, as stated before, Var_α and ES_α alone may be not indicative of the potential tail losses, since they do not provide information about the dispersion. Computing the truncated Gini index $G(\alpha)$ thus appears to be a solution, as it provides information about the precision of the ES_α and the length of the tail beyond Var_α . Furthermore, being a scale free measure, it can be used to compare the two loss distributions even if they represent different phenomena.

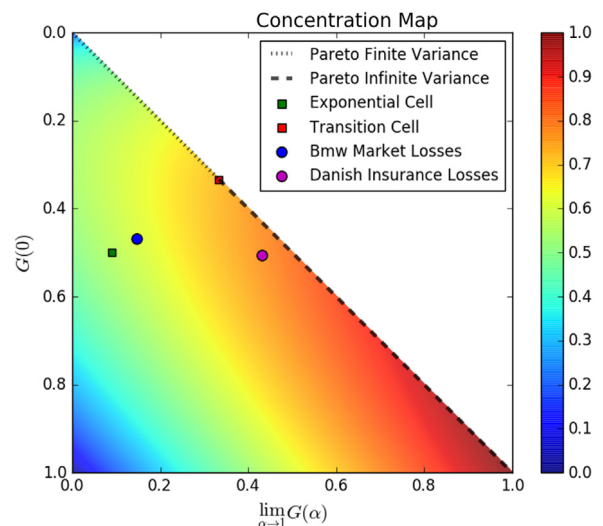


Fig. 12. Risk evaluation of the Danish and BMW data sets via the Concentration Map. Cobb–Douglas parameters: $a = 0.3$, $b = 0.7$.

Table 2 collects the Var_α , ES_α , $G(\alpha)$, VM_α and CV_α for both data sets, with $\alpha \in \{0.9, 0.95, 0.99\}$. As expected, looking at Var_α and at ES_α , the insurance losses seem riskier than the market ones, even if the different scales do not allow for a proper quantification of the term “riskier”. Conversely, the scale free truncated Gini index gives us the opportunity of an informative comparison: Danish losses are definitely more risky, since their Gini index is about twice the value of that of BMW.

Both the variance-to-mean (VM) and the coefficient of variation (CV) cannot be used for the Danish insurance losses, because of the infinite variance. When they can be used, as for the BMW losses, it seems more appropriate to use the CV, which – as the Gini index – is scale free. For different values of α , the quantity VM_α does not vary sensibly, not providing useful insight.

To conclude, in Fig. 12 we show a comparison based on the Concentration Map, with an underlying Cobb–Douglas with parameters $a = 0.3$ and $b = 0.7$. The Danish fire claims are confirmed to be more risky.

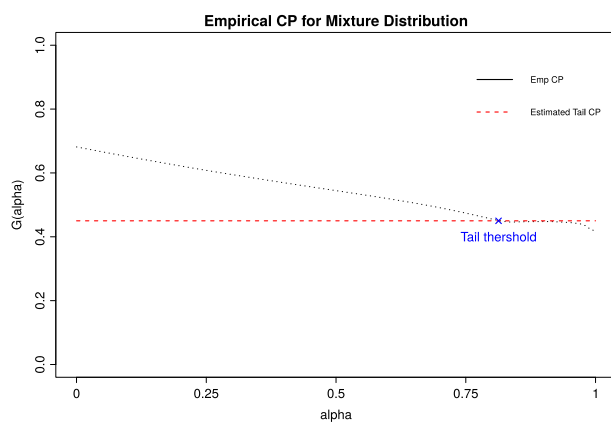
7.3. Identifying thresholds in extreme value theory

In the context of extreme value theory (Fisher and Tippett, 1928; Gnedenko, 1943; Pickands, 1975), a significant issue is the

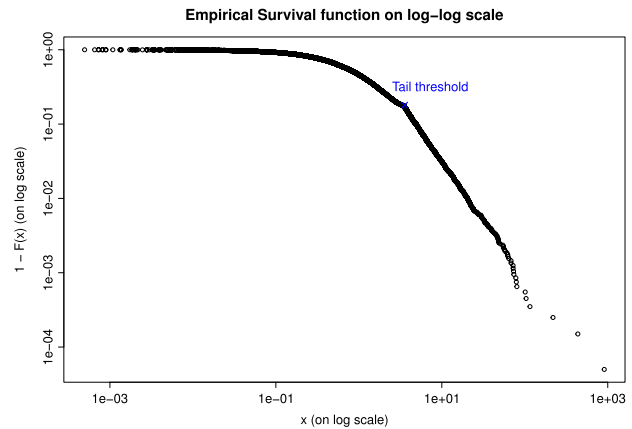
Table 2

Comparison of the main risk measures for the Danish and BMW data sets. For the insurance losses it is not possible to compute the variance-to-mean ratio VM nor the coefficient of variation CV , for the second moment is not finite.

α	Danish insurance losses					BMW log-losses				
	VaR_α	ES_α	$G(\alpha)$	VM_α	CV_α	VaR_α	ES_α	$G(\alpha)$	VM_α	CV_α
0.9	5.541	15.56	0.437	–	–	0.022	0.034	0.201	0.007	0.444
0.95	9.972	24.08	0.390	–	–	0.029	0.044	0.176	0.007	0.387
0.99	26.04	58.58	0.387	–	–	0.049	0.070	0.157	0.007	0.306



(a) Empirical Concentration Profile of a sample from the mixture distribution in Eq. (18).



(b) Zipf plot (log–log plot of the survival function) for the same data.

Fig. 13. Threshold identification for Paretian tails using the empirical CP and the Zipf plot. In the CP figure, a dashed red line has been added to underline the constant behavior of the Concentration Profile at the tail level. The blue crosses indicate the suggested thresholds for Paretianity.

estimation of the threshold above which the possible Paretian tail starts. This has applications in the computation of the tail index, the tail quantile and other relevant quantities (Embrechts et al., 2003). However no unique methodology has been developed to estimate the right threshold so far, but many different tools are used to provide hints about the best value, see for example Cirillo (2013). Here we propose the *Concentration Profile* as an additional instrument.

As the Concentration Profile is a sequence of truncated Gini indices computed for increasing thresholds, progressively discharging the previous observations, we expect that if the data exhibits a Paretian right tail, then, from a certain α -level onward, the CP should be approximately constant.

An example performed on simulated data will clarify this. In Fig. 13a, the empirical CP of a sample (1000 points) from the following mixture is plotted:

$$Y_{mix} = \beta Y + (1 - \beta)Z, \quad (18)$$

where Y is a standard exponential distribution, Z is a Pareto distribution with $\rho = 1.5$, and the mixing parameter is $\beta = 0.85$. By construction, this data set exhibits Paretian tails. In particular, the Paretian behavior starts around quantile $Q(0.825) = 3.2445$ (denoted by a blue cross in Fig. 13a).

To assess the quality of our findings we can compare our methodology with other techniques used in EVT to detect the tail threshold (see Cirillo, 2013; Embrechts et al., 2003, for a review). In Fig. 13b, we report the empirical survival function $S(x) = 1 - F(x)$ plotted in a log–log scale (Cirillo, 2013). From this plot, known as Zipf plot, we see how the straight line behavior which characterizes the Paretian tail is present from approximately the same level identified via the empirical CP. Similar results are found using the mean-excess plot and other commonly used techniques (Cirillo, 2013).

To provide a further real-world example, let us consider 272 non-equally spaced market losses in a portfolio related to Goldman

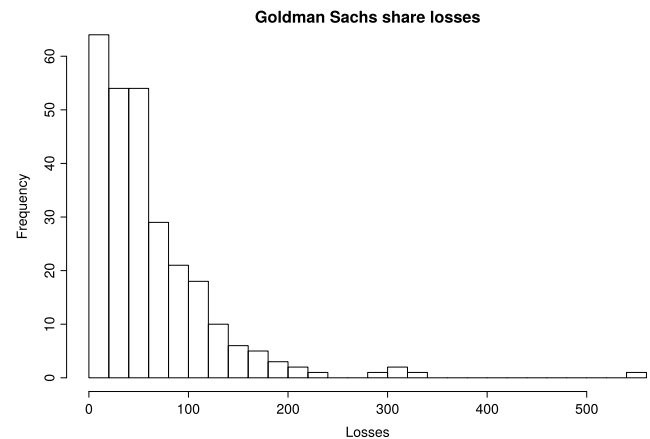


Fig. 14. Histogram of losses for a portfolio on Goldman Sachs shares (272 observation in the period 2010–2013).

Sachs shares in the period 2010–2013. The histogram of the loss distribution is shown in Fig. 14, and its Concentration Profile is given in Fig. 15a. Once again, in order to identify the Paretian threshold, we look for the point at which the concentration profile starts exhibiting an approximately horizontal behavior. In Fig. 15a, we identify this level as the quantile associated to $\alpha \approx 0.79$, as $Q(0.79) = 90.1827$. To ease the identification we also provide a dashed horizontal line representing the theoretical CP of a Pareto distribution with shape parameter $\rho = 2.8$, as estimated from data using the GPD approximation and MLE. In Fig. 16a the last section of the Concentration Profile is zoomed in, to show the horizontal behavior. As before, Zipf plots are provided for comparison: the Paretian threshold seems to be around the 79% quantile, as suggested by the Concentration Profile.

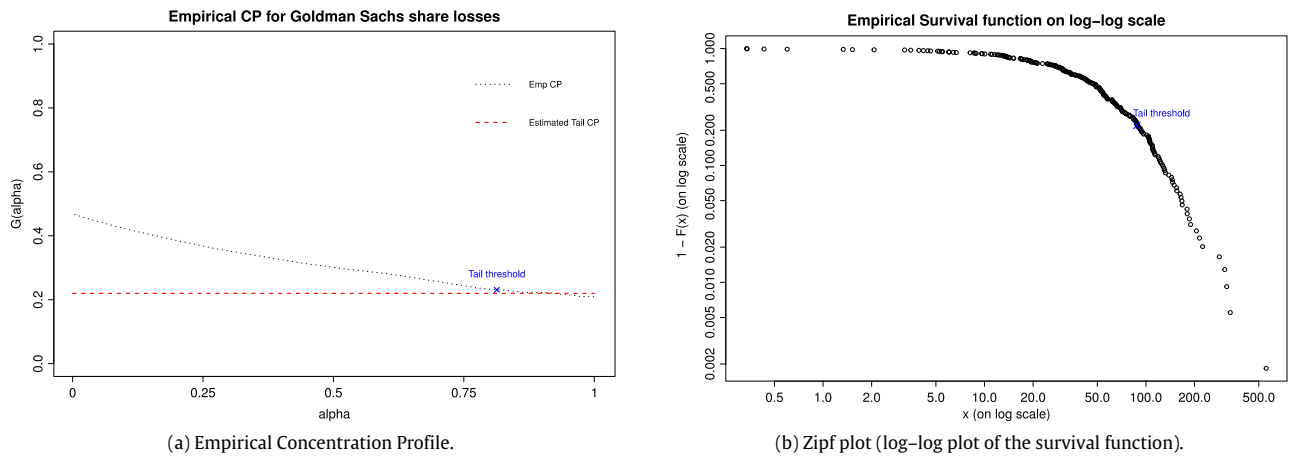


Fig. 15. Empirical Concentration Profile and Zipf plot for the losses of the portfolio on Goldman Sachs shares. The dashed red line represents the theoretical Concentration Profile generated by a Pareto with shape parameter $\rho = 2.8$ ($\xi = 0.36$). Blue crosses indicate the suggested thresholds for Paretnianity.

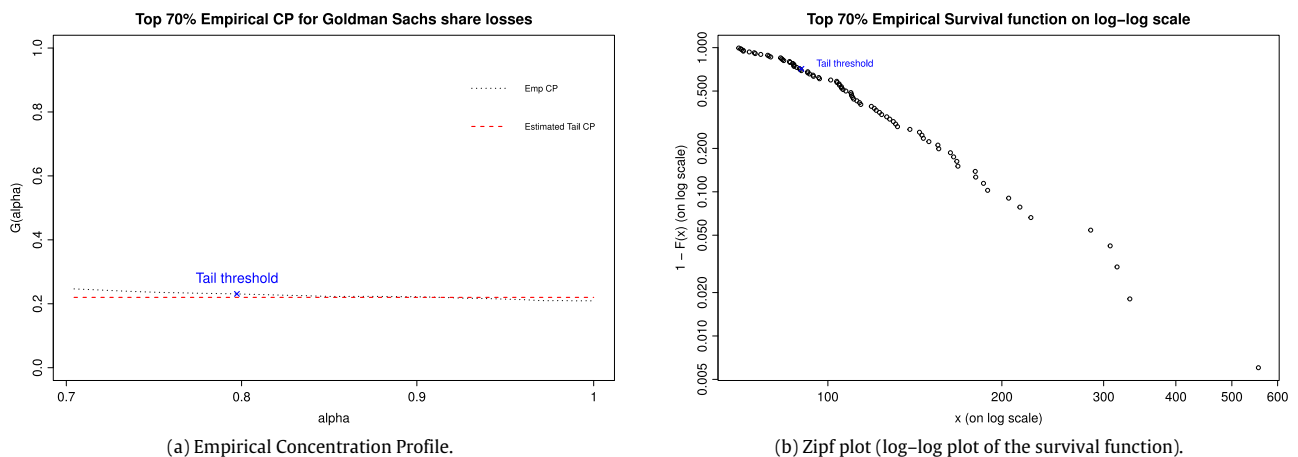


Fig. 16. Zoomed-in versions of both the empirical Concentration Profile and the Zipf plot of Fig. 15 to better visualize the right tail. The blue crosses indicate the suggested thresholds for Paretnianity.

8. Conclusions

In this work we introduced a novel approach to risk management, based on the study of concentration measures of the loss distribution. In particular, we showed that a truncated sequence of Gini indices – the Concentration Profile – represents an accurate way of assessing the variability of the larger losses, the precision of common risk management measures like the ES_α , and tail risk in general. Combining the Concentration Profile and standard results from utility and risk theory, we have developed a *Concentration Map*, which can be used to assess the risk attached to potential losses on the basis of the risk profile of a risk manager. Finally, we have used a sequence of truncated Gini indices as weights for the ES_α defining the so-called *Concentration Adjusted Expected Shortfall*, a measure able to capture additional features of tail risk. Using simulated and empirical data, we showed how to use our methods in practice, from risk management to extreme value theory.

It is worth stressing that Concentration Profiles and Maps may also be obtained by substituting the Gini index with other measures of concentration, like the Pietra index (Pietra, 1915). However, by preliminary studies we noticed that the increase in mathematical complexity is not compensated by an improvement in the applicability. Using an Occam's razor principle, we therefore preferred the simpler (and better known) Gini index.

Acknowledgments

All authors acknowledge the support of the FP7 Marie Curie ITN Project WakeUpCall No. 643045.

Pasquale Cirillo also acknowledges the support of the FP7 Marie Curie Action under the Grant PCIG13-GA-2013-618794.

Appendix A. Gini indices and Lorenz curves of some notable distributions

Table 3 contains some distributions often used for modeling losses (Hull, 2012; Kleiber and Kotz, 2003). We collect them here, as they are also useful for our discussion.

For the distributions in Table 3, we collect in Table 4 the corresponding closed form formulas for their Lorenz curves and Gini indices.

Appendix B. Same Gini index but different Lorenz curves: an example

In Fig. 17 two different Lorenz curves are shown. Their functional forms are

$$L_1(x) = (2x - 1)1_{\{x \geq 0.5\}},$$

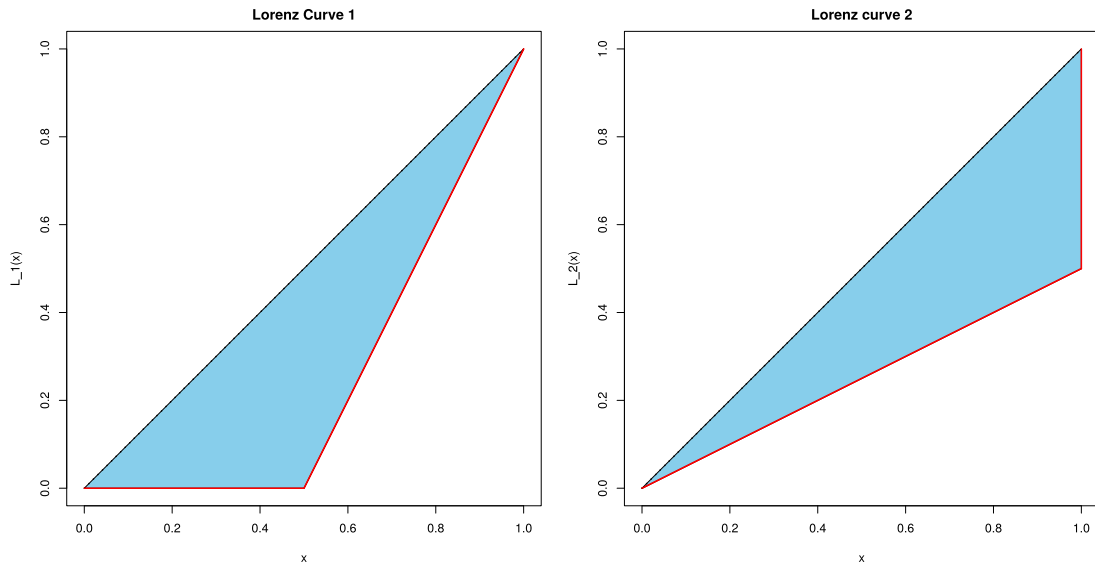


Fig. 17. In the first graph Lorenz curve $L_1(x)$ is shown; in the second $L_2(x)$. The area shaded between L_{pe} and the two Lorenz curves is the same, giving the same Gini index, however the asymmetry is clearly different.

Table 3

C.d.f., p.d.f., support and shape parameter of some of the most used distributions in risk management.

Distribution	c.d.f.	p.d.f.	Support	Shape
Weibull	$F(y) = 1 - e^{-(\frac{y}{\lambda})^\gamma}$	$f(y) = \frac{\gamma}{\lambda} (\frac{y}{\lambda})^{\gamma-1} e^{-(\frac{y}{\lambda})^\gamma}$	$y \in \mathbb{R}^+$	$\gamma \in \mathbb{R}^+$
Lognormal	$F(y) = \Phi(\frac{\ln(y)-\mu}{\sigma})$	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\ln(y)-\mu)^2}{2\sigma^2}}$	$y \in \mathbb{R}^+$	$\sigma \in \mathbb{R}^+$
Exponential	$F(y) = 1 - e^{-\frac{y}{\lambda}}$	$f(y) = \lambda e^{-y\lambda}$	$y \in \mathbb{R}^+$	None
Pareto	$F(y) = 1 - (\frac{y}{y_0})^{-\rho}$	$f(y) = \frac{\rho y_0^\rho}{y^{\rho+1}} 1_{[y \geq y_0]}$	$y \in [y_0, +\infty)$	$\rho \in \mathbb{R}^+$

Table 4

Gini index and Lorenz curve for the same distributions of Table 3.

Distribution	Gini index	Lorenz curve $x \in [0, 1]$
Weibull	$1 - 2^{-\frac{1}{\gamma}}$	$L(x) = 1 - \frac{\Gamma(-\log(1-x), 1 + \frac{1}{\gamma})}{\Gamma(1 + \frac{1}{\gamma})}$
Lognormal	$2\Phi(\frac{\sigma}{\sqrt{2}}) - 1$	$L(x) = \Phi(\Phi^{-1}(x) - \sigma)$
Exponential	$\frac{1}{2}$	$L(x) = x + x(-\log(1-x)) + \log(1-x)$
Pareto	$\frac{1}{2\rho-1}$	$L(x) = 1 - (1-x)^{\frac{\rho-1}{\rho}}$

for the one on the left, and

$$L_2(x) = 0.5x,$$

for that on the right.

Fixing $x = 0.9$, gives $L_1(x) = 0.8$ and $L_2(x) = 0.45$. This means that according to L_1 the top 10% losses account for 20% of the total loss. These numbers change under L_2 , where the same top 10% accounts for 55% of all losses! In both cases, the Gini index is the same: $G_1 = \frac{1}{2} = G_2$. Graphically speaking, it is sufficient to notice that the two blue triangles in Fig. 17 occupy the same area. This simple example shows that the Gini index alone is not able to capture the dissimilarity between two loss distributions, despite their very different risk profiles.

This example can also be used to heuristically justify the use of the Concentration Profile. Let us truncate the original data at $x = 0.5$, i.e. let us ignore the first 50% of the two samples. The new Lorenz curves are given in Fig. 18.

Using Eq. (5) it is not difficult to prove that the two curves are $L_1^{0.5}(x) = x$ and $L_2^{0.5}(x) = \frac{1}{3}x$. The second situation is more risky, given the higher inequality. This is also reflected by the new Gini indices, which now differ: $G_1^{0.5} = 0 < G_1$ and $G_2^{0.5} = \frac{2}{3} > G_2$. This denotes a different risk concentration in the tails.

In a Concentration Profile, we perform the same computation as above at every point of the support of the loss distribution. We thus get a sequence of truncated Gini indices, which is informative about tail risk.

Appendix C. Derivation for truncated Lorenz curve

In order to derive the Lorenz curve for truncated random variables, i.e. Eq. (5), recall that the p.d.f. of a random variable X left-truncated on $[u, +\infty)$ is given by:

$$f_u(x) = \frac{f(x)}{1 - F(u)}. \quad (19)$$

where $f(x)$ is the p.d.f. of X .

Let us define the Cumulative Mean Function $M(t)$ for the truncated random variable, i.e.

$$M(t) := \frac{\int_u^t \frac{xf(x)}{1-F(u)} dx}{\int_u^{+\infty} \frac{xf(x)}{1-F(u)} dx}, \quad (20)$$

with $t \in [u, +\infty)$, which becomes

$$M(t) := \frac{\int_u^t xf(x) dx}{\int_u^{+\infty} xf(x) dx}.$$

The Lorenz curve is related to the Cumulative Mean Function via the change of variable $t = F^{-1}(x)$. Applying this to Eq. (20), we get the Lorenz curve associated to the truncated random variable X_u on $[u, +\infty)$,

$$L_u^*(x) = \frac{\int_\alpha^x F^{-1}(u) du}{\int_\alpha^1 F^{-1}(u) du}, \quad (21)$$

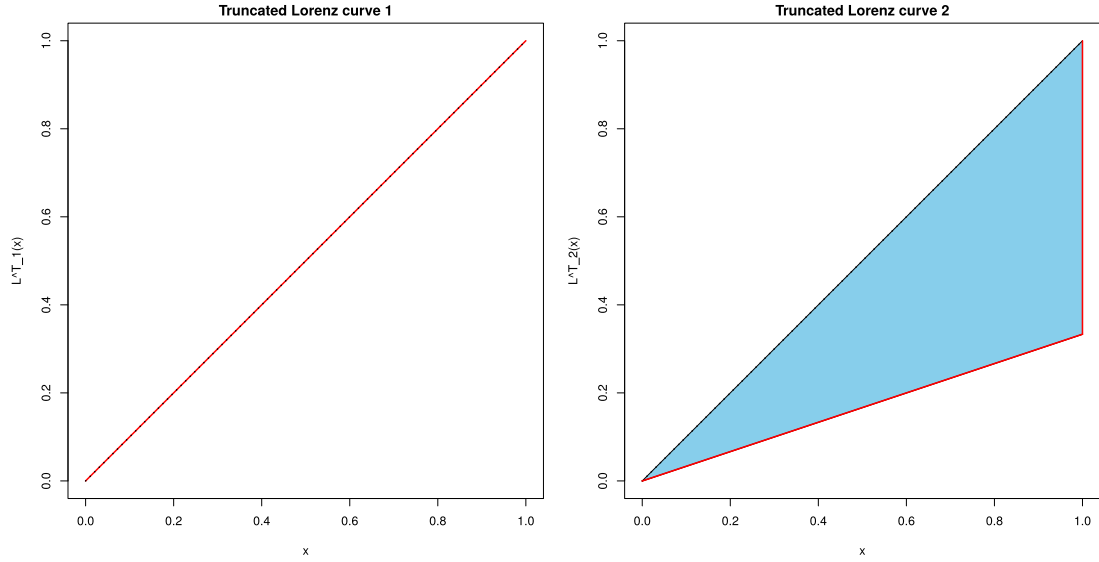


Fig. 18. The rescaled versions of the Lorenz curves in Fig. 17, where the upper 50% of the domain is considered. The shaded area is absent in the left side meaning that the Gini index of $L_1^{0.5}(x)$ is zero.

where α is the confidence level associated to truncation in u , i.e. $F(u) = \alpha$. The curve $L_\alpha(x)$ is therefore defined on the risk subclass S_α .

Now, set $z = \frac{x-\alpha}{1-\alpha}$, and plug it in Eq. (21), so that

$$L_\alpha(z) = L_\alpha^*(z(1-\alpha) + \alpha). \quad (22)$$

Notice that Eq. (22) can be re-written as

$$L_\alpha(z) = L_\alpha^*(z(1-\alpha) + \alpha) = \frac{\int_\alpha^{z(1-\alpha)+\alpha} F^{-1}(u) du}{\int_\alpha^1 F^{-1}(u) du}. \quad (23)$$

Recalling that the Lorenz curve associated with a random variable with domain in \mathbb{R}^+ is given by Eq. (1), we can adjust the right-hand side of Eq. (23) so that

$$\begin{aligned} L_\alpha(z) &= \frac{\int_\alpha^{z(1-\alpha)+\alpha} F^{-1}(u) du}{\int_\alpha^1 F^{-1}(u) du} \\ &= \frac{\int_0^{z(1-\alpha)+\alpha} F^{-1}(u) du - \int_0^\alpha F^{-1}(u) du}{\int_0^1 F^{-1}(u) du - \int_0^\alpha F^{-1}(u) du}. \end{aligned} \quad (24)$$

By dividing the right-hand side by $\frac{\int_0^1 F^{-1}(u) du}{\int_0^1 F^{-1}(u) du}$, and by recalling that $L(\alpha) = \frac{\int_0^\alpha F^{-1}(u) du}{\int_0^1 F^{-1}(u) du}$ and that $L(1) = 1$, we obtain

$$L_\alpha(z) = \frac{L(z(1-\alpha) + \alpha) - L(\alpha)}{1 - L(\alpha)}. \quad (25)$$

Appendix D. Proof of Theorem 7

The quantile function of a one-parameter GPD is given by

$$F^{-1}(v) = \frac{(1-v)^{-\xi} - 1}{\xi}, \quad 0 \leq v \leq 1. \quad (26)$$

From this, using Eq. (1), we get

$$L(x) = \frac{\int_0^x \frac{(1-v)^{-\xi} - 1}{\xi} dv}{\int_0^1 \frac{(1-v)^{-\xi} - 1}{\xi} dv}. \quad (27)$$

Solving (27), we obtain

$$L(x) = \frac{1 - (1-x)^{(1-\xi)} - x + \xi x}{\xi}, \quad (28)$$

which is the closed form solution for the Lorenz curve of the GPD.

We now obtain the truncated Lorenz curve by evaluating (28) in $\alpha + (1-\alpha)x$, that is Eq. (29) in Box I.

Using Eqs. (7) and (29), we observe $G(\alpha)$ given in Box II from which we derive

$$G(\alpha) = \frac{\xi}{2 - (1-\alpha)\xi(-2+\xi)(-1+\xi) - \xi}. \quad (30)$$

In order to prove that, in the case of a light-tailed distribution, the CP is decreasing, we have to show that $\frac{dG(\alpha)}{d\alpha} < 0$ for every $\alpha \in [0, 1]$. We observe

$$\frac{dG(\alpha)}{d\alpha} = \frac{1}{2(\alpha-1)(\log(1-\alpha)-1)^2}, \quad (31)$$

which is always negative.

Left to show is that the CP of a fat-tailed distribution gets flatter as $\xi \rightarrow 1$, with limiting case the constant. It is sufficient to show that: $\frac{d(\frac{dG(\alpha)}{d\alpha})}{d\xi} < 0$ and $\lim_{\xi \rightarrow 1} \frac{dG(\alpha)}{d\alpha} = 0$. Therefore

$$\frac{dG(\alpha)}{d\alpha} = -\frac{(\xi-1)\xi^2(1-\alpha)^{\xi-1}}{(\xi-2)((\xi-1)(1-\alpha)^\xi + 1)^2}, \quad (32)$$

whose limit for $\xi \rightarrow 1$ goes to 0.

Appendix E. Derivation for the exponential distribution Concentration Profile

We here provide the derivation of the CP for the exponential distribution. The derivation of the CP for other distributions only involves the computation of some more cumbersome integrals. For the lognormal distribution the actually-usable solution can only be obtained numerically.

Our aim is to use formula (7) to obtain a workable form for the CP.

Recalling Table 4, the Lorenz curve for the exponential distribution is given by:

$$L(x) = x + x(-\log(1-x)) + \log(1-x).$$

In order to get its truncated version, we exploit formula (5), so that Eq. (33) in Box III with $\alpha \in [0, 1]$.

$$L_{\alpha}(x) = \frac{-(1-\alpha)^{\xi}((\alpha-1)(x-1))^{-\xi} + x(1-\alpha)^{\xi}(((\alpha-1)(x-1))^{-\xi} + \xi - 1) + 1}{\xi(1-\alpha)^{\xi} - (1-\alpha)^{\xi} + 1}. \quad (29)$$

Box I.

$$G(\alpha) = 1 - 2 \int_0^1 \frac{-(1-\alpha)^{\xi}((\alpha-1)(x-1))^{-\xi} + x(1-\alpha)^{\xi}(((\alpha-1)(x-1))^{-\xi} + \xi - 1) + 1}{\xi(1-\alpha)^{\xi} - (1-\alpha)^{\xi} + 1} dx,$$

Box II.

$$L_{\alpha}(x) = \frac{\alpha + (1-\alpha)x + (\alpha + (1-\alpha)x)(-\log(1 - (\alpha + (1-\alpha)x))) + \log(1 - (\alpha + (1-\alpha)x)) - (\alpha + \alpha(-\log(1-\alpha)) + \log(1-\alpha))}{1 - (\alpha + \alpha(-\log(1-\alpha)) + \log(1-\alpha))}, \quad (33)$$

Box III.

After some algebra we get

$$L_{\alpha}(x) = \frac{-x + \log(1-\alpha) + (-1+x)\log((-1+x)(-1+\alpha))}{-1 + \log(1-\alpha)}. \quad (34)$$

We can now exploit Eq. (34) in formula (7), and we get

$$G(\alpha) = 1 - 2 \int_0^1 \frac{-x + \log(1-\alpha) + (-1+x)\log((-1+x)(-1+\alpha))}{-1 + \log(1-\alpha)} dx. \quad (35)$$

The integral in (35) can be solved directly, thus getting the formula provided in Table 1.

Appendix F. Supplementary materials

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.insmathco.2017.11.003>.

References

- Basel Committee on Banking Supervision - BCBS, 2006. Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. Bank for International Settlements.
- Basel Committee on Banking Supervision - BCBS, 2011. Basel III: A Global Regulatory Framework for more Resilient Banks and Banking Systems. Bank for International Settlements.
- Barucci, E., 2003. Financial Markets Theory. Springer.
- Beirlant, J., Goegebeur, Y., 2004. Local polynomial maximum likelihood estimation for Pareto-type distributions. *J. Multivariate Anal.* 89, 97–118.
- Blackorby, C., Donaldson, D., 1980. Ethical indices for the measurement of poverty. *Econometrica* 48, 1053–1060.
- Cambini, A., Martein, L., 2009. Generalized Convexity and Optimization. Springer.
- Chavez-Demoulin, V., Embrechts, P., Hofert, M., 2016. An extreme value approach for modeling Operational Risk losses depending on covariates. *J. Risk Insur.* 83, 735–776.
- Chavez-Demoulin, V., Embrechts, P., Sardy, S., 2014. Extreme-quantile tracking for financial time series. *J. Econometrics* 181, 44–52.
- Chotikapanich, Duangkamon, 2008. Modeling Income Distributions and Lorenz Curves. Springer.
- Cirillo, P., 2013. Are your data really Pareto distributed?. *Physica A* 392, 5947–5962.
- Cirillo, P., Taleb, N.N., 2016. Expected shortfall estimation for apparently infinite-mean models of operational risk. *Quant. Finance* 16, 1–10.
- Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and their Applications. Cambridge University Press.
- Tian, D., Jiang, L., 2015. Quasiconvex risk statistics with scenario analysis. *Math. Financ. Econ.* 9, 111–121.
- Du, Z., Escanciano, J.C., 2016. Backtesting Expected Shortfall: Accounting for Tail Risk. *Manage. Sci.* 63, 940–958.
- Edwards, W., 1992. Utility Theories: Measurements and Applications. Springer.
- Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Efron, B., Tibshirani, R.J., 1994. An Introduction To the Bootstrap. In: Chapman et Hall Monographs on Statistics and Applied Probability.
- Eliazar, I., Sokolov, I.M., 2012. Measuring statistical evenness: a panoramic overview. *Physica A* 391, 1323–1353.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 2003. Modelling Extremal Events for Insurance and Finance. Springer.
- Fiordelisi, F., Soana, M.-G., Schwizer, P., 2014. Reputational losses and operational risk in banking. *Eur. J. Finance* 20, 105–124.
- Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Camb. Phil. Soc.* 24, 180–190.
- Fontanari, A., Taleb, N.N., Cirillo, P., Gini estimation under infinite variance. Available at SSRN: <https://ssrn.com/abstract=3005184>.
- Frittelli, M., Maggis, M., 2011. Dual representation of quasi-convex conditional maps. *SIAM J. Financial Math.* 2, 357–382.
- Gastwirth, J.L., 1972. The Estimation of the Lorenz Curve and Gini Index. *Rev. Econ. Stat.* 54 (3).
- Gini, C., 1912. Variabilità E Mutabilità. Reprinted in: Variabilità e Mutabilità, E Pizetti and T Salvemini, Memorie di Metodologica Statistica, Libreria Eredi Virgilio Veschi, Rome, 1955.
- Gnedenko, B., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math.* 44, 423–453.
- de Haan, L., Ferreira, A., 2006. Extreme Value Theory: An Introduction. Springer.
- Hull, J., 2012. Risk Management and Financial Institutions. Wiley.
- Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. Wiley.
- Konno, H., Yamazaki, H., 1991. Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Manage. Sci.* 37, 519–531.
- Lorenz, M.O., 1905. Methods of measuring the concentration of wealth. *Publ. Amer. Stat. Assoc.* 9, 209–219.
- Maronna, R., Martin, R.D., Yohai, V., 2006. Robust Statistics - Theory and Methods. Wiley.
- McNeil, A., Frey, R., Embrechts, P., 2015. Quantitative Risk Management, second ed.. 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. Bank of Italy.
- Nair, N.U., Sankaran, P.G., Kumar, B.V., 2002. Characterization of distributions by properties of truncated Gini index and mean difference. *Metron* 70, 173–191.
- Pickands, J., 1975. Statistical inference using extreme order statistics. *Ann. Statist.* 3, 119–131.
- Pietra, G., 1915. Atti Del Reale Istituto Veneto Di Scienze, Lettere Ed Arti, Tomo LXXIV Parte II.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *J. Amer. Statist. Assoc.* 89, 1303–1313.

- Resti, A., Sironi, A., 2007. Risk Management and Shareholders' Value in Banking: from Risk Measurement Models To Capital Allocation Policies. Wiley.
- Shalit, H., Yitzhaki, S., 1984. Mean-gini, portfolio theory, and the pricing of risky assets. *J. Finance* 39, 1449–1468.
- Shalit, H., Yitzhaki, S., 2005. The mean-Gini efficient portfolio frontier. *Journal of Financial Research* 28, 59–75.
- Solvency II. 2009. European Union Directive 2009/138/CE.
- Yitzhaki, S., Schechtman, E., 2013. *The Gini Methodology*. Springer.