Lorenz-based quantitative risk management

Fontanari, Andrea

DOI
10.4233/uuid:0c5b50a5-4514-431d-a31a-b1f4ae2c0713

Publication date
2019

Document Version
Final published version

Citation (APA)
Fontanari, A. (2019). *Lorenz-based quantitative risk management*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:0c5b50a5-4514-431d-a31a-b1f4ae2c0713

LORENZ-BASED QUANTITATIVE RISK MANAGEMENT

LORENZ-BASED QUANTITATIVE RISK MANAGEMENT

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Tuesday 10 December 2019 at 15:00 o'clock

by

Andrea FONTANARI

Master of Science in Economic and Social Sciences,
Bocconi University, Milan, Italy,
born in Trento, Italy.

This dissertation has been approved by the promotors

promotor: Prof. dr. ir. C. W. Oosterlee
copromotor: Dr. P. Cirillo

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. C. W. Oosterlee	Delft University of Technology, promotor
Dr. P. Cirillo	Delft University of Technology, copromotor

Independent members:

Prof. dr. F. H. J. Redig	Delft University of Technology
Prof. dr. A. Pascucci	University of Bologna, Italy
Prof. dr. M. Bonetti	Bocconi University, Italy
Prof. dr. P. J. C. Spreij	University of Amsterdam and Radboud University
Prof. dr. ir. M. H. Vellekoop	University of Amsterdam
Prof. dr. ir. G. Jongbloed	Delft University of Technology, reserve member

This research was funded by the European Commission through European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 643045.

ISBN 978-94-6380-657-2

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

I am not interested in research,
I am interested in understanding.

David Blackwell

SUMMARY

In this thesis, we address problems of quantitative risk management using a specific set of tools that go under the name of Lorenz curve and inequality indices, developed to describe the socio-economic variability of a random variable.

Quantitative risk management deals with the estimation of the uncertainty that is embedded in the activities of banks and other financial players due, for example, to market fluctuations. Since the well-being of such financial players is fundamental for the correct functioning of the economic system, an accurate description and estimation of such uncertainty is crucial.

However, this task is complicated by the nature of the randomness involved. In fact, unlike other phenomena, typical of the physical world, where the randomness is given by measurement error and so governed by Gaussian laws, financial data deviate from gaussianity and often exhibit heavy-tailed behavior meaning that rare and disruptive events have a non-negligible chance of happening. Mathematically this translates in phenomena that have highly asymmetric distributions and that may not be L^2-integrable, making most of the standard modeling techniques inaccurate and biased.

The problem of describing and summarizing uncertainty for models departing from gaussianity have been tackled by statisticians and mathematicians when trying to build methods to study socio-economic phenomena. The most celebrated example is probably the Pareto distribution, the prototype for many heavy-tailed models, which was developed to describe the size of human settlements.

Within socio-economic models probably the most successful tool to analyze variability is the Lorenz curve and inequality indices derived from it.

The Lorenz curve is a transformation of a positive valued random variable which maps its quantile function into an increasing convex function space. This type of transformation allows picturing the variability induced by a random variable in a clearer and more compact way than by looking at its probability or cumulative density function. Additionally, by studying the geometry of such transformation it is possible to build measures that capture different aspects of the variability of a random variable. In this thesis, we focus in particular the L^1 functional distance between the Lorenz curve associated to a deterministic constant and the Lorenz curve obtained from the data that goes under the name of Gini index.

This thesis is essentially split up into two parts. In the first one, we deal with the issue of tail variability measurements for portfolio loss distributions using the descriptive power of the Lorenz curve. In the second one, we exploit probabilistic properties of the Lorenz curve such as stochastic orderings, relations with majorization and its representation as a convex distortion, to tackle risk management problems related to dependence and systemic risk.

In Chapter 2 we build, starting from the Gini index, several tools for the estimation of the tail variability of a loss distribution. In particular, by truncating the distribution at

its Value-at-Risk we build another transformation of the quantile function, called Concentration Profile which provides a better interpretation of the reliability of the Expected Shortfall and still recovers the original distribution up to a constant allowing not only to describe variability but also to perform model selection. Real data examples and simulations are provided along with an application of the choice of threshold problem for extreme value theory.

In Chapter 3 we study the behavior of the non-parametric estimator of the Gini index when a heavy-tailed stochastic environment is assumed. In particular, we study its limiting distribution via the application of the Generalized Central Limit theorem for order statistics and we prove how the loss of symmetry of the limiting $\alpha-$stable distribution when the second moment becomes infinite may increase the bias in the estimation. We finally suggest the use of a finite sample correction based on the mode-to-mean distance to improve the consistency of the estimator.

In Chapter 4 we apply the notion of majorization, a partial order on positive real vectors strongly related with the Lorenz curve, to study financial correlation matrices. In particular, we derive a set of axioms that ordering over the space of correlation matrices should have in order to capture financial risk. We prove that the partial order obtained from the majorization of the spectra of correlation matrices respects these axioms and that it can be used to build summary measures. In particular, we show how many summary measures of correlation matrices used in practice are consistent with such order. We further investigate the validity of such order by checking its presence in empirical data. We find out that, by looking at the Industrial Dow Jones, correlation matrices are ordered with respect to each other more consistently right before and during financial turbulence. From this observation, we build a simple warning system that generates a signal by looking at the ordering trends of daily correlation matrices showing that such a system produces much more information than just a single risk measure.

In Chapter 5 we use the geometry of the Lorenz curve, a convex distortion of the identity map, to build Archimedean generators for the construction of bivariate copulas. We show how any non-strict bivariate Archimedean copula can be obtained using a dual of the Lorenz curve as its generator. We further characterize the right-tail properties of such copulas in terms of the univariate random variable associated to Lorenz curve used to span them. We also show how the Gini index is related to the Kendall's τ measure of association and that the Lorenz and the star stochastic orders can be related to stochastic multivariate orderings. Finally, we provide simulations and algebraic formulas for the cumulative distribution function and Kendall functions of some of such generated copulas.

Finally, in the conclusions, further research questions are posed. Among others, we propose to use again the geometric structure of the Lorenz curve to build Pickands dependence functions for extreme value theory applications and to exploit the inequality indices as dependence measures for diagonal and Archimedean copulas.

SAMENVATTING

In dit proefschrift behandelen we problemen van kwantitatief risicobeheer met behulp van een specifieke set hulpmiddelen die onder de naam Lorenz-curve en ongelijkheidsindexen vallen, ontwikkeld om de sociaal-economische variabiliteit van een willekeurige variabele te beschrijven.

Kwantitatief risicobeheer houdt zich bezig met de schatting van de onzekerheid die is ingebed in de activiteiten van banken en andere financiële spelers, bijvoorbeeld door marktschommelingen. Aangezien het welzijn van dergelijke financiële spelers van fundamenteel belang is voor de juiste werking van het economische systeem, is een nauwkeurige beschrijving en schatting van dergelijke onzekerheid cruciaal.

Deze taak wordt echter gecompliceerd door de aard van de willekeur. In tegenstelling tot andere fenomenen, typerend voor de fysieke wereld, waar de willekeur wordt gegeven door meetfouten en dus wordt beheerst door Gauss-wetten, wijken financiële gegevens af van Gaussianiteit en vertonen ze vaak zwaarachtig gedrag, wat betekent dat zeldzame en verstorende gebeurtenissen een niet- verwaarloosbare kans hebben om te gebeuren. Wiskundig vertaalt dit zich in fenomenen die zeer asymmetrische verdelingen hebben en die mogelijk niet L^2- integreerbaar zijn, waardoor de meeste standaard modelleringstechnieken onnauwkeurig en bevooroordeeld zijn.

Het probleem van het beschrijven en samenvatten van onzekerheid voor modellen die van gaussianiteit afwijken, is aangepakt door statistici en wiskundigen bij het proberen methoden te ontwikkelen om sociaal-economische fenomenen te bestuderen. Het meest gevierde voorbeeld is waarschijnlijk de Pareto-distributie, het prototype voor veel zwaarstaartmodellen, dat werd ontwikkeld om de grootte van menselijke nederzettingen te beschrijven.

Binnen sociaal-economische modellen is waarschijnlijk de meest succesvolle tool om variabiliteit te analyseren de Lorenz-curve en daarvan afgeleide ongelijkheidsindexen.

De Lorenz-curve is een transformatie van een willekeurige willekeurige variabele met een positieve waarde, die zijn kwantiele functie in een toenemende convexe functieruimte in kaart brengt. Dit type transformatie maakt het mogelijk om de variabiliteit die wordt geïnduceerd door een willekeurige variabele op een duidelijkere en compactere manier weer te geven dan door te kijken naar de waarschijnlijkheid of de cumulatieve dichtheidsfunctie. Door de geometrie van een dergelijke transformatie te bestuderen, is het bovendien mogelijk om maatregelen te bouwen die verschillende aspecten van de variabiliteit van een willekeurige variabele bevatten. In dit proefschrift richten we ons in het bijzonder op de functionele afstand van L^1 tussen de Lorenz-curve geassocieerd met een deterministische constante en de Lorenz-curve verkregen uit de gegevens die onder de naam Gini-index gaan.

Dit proefschrift is in wezen opgesplitst in twee delen. In de eerste behandelen we de kwestie van staartvariabiliteitsmetingen voor portefeuilleverliesverdelingen met behulp

van de beschrijvende kracht van de Lorenz-curve. In de tweede gebruiken we probabilistische eigenschappen van de Lorenz-curve, zoals stochastische ordeningen, relaties met majorisatie en de weergave ervan als een convexe vervorming, om risicobeheersingsproblemen met betrekking tot afhankelijkheid en systeemrisico aan te pakken.

In Hoofdstuk 2 bouwen we, uitgaande van de Gini-index, verschillende hulpmiddelen voor de schatting van de staartvariabiliteit van een verliesverdeling. In het bijzonder bouwen we door het afkappen van de verdeling naar zijn Value-at-Risk een nieuwe transformatie van de kwantiele functie, genaamd concentratieprofiel, die een betere interpretatie geeft van de betrouwbaarheid van de verwachte tekortkoming en de oorspronkelijke verdeling nog steeds herstelt tot een constante waardoor niet alleen om variabiliteit te beschrijven, maar ook om modelselectie uit te voeren. Echte gegevensvoorbeelden en simulaties worden verstrekt samen met een toepassing van de keuze van het drempelprobleem voor extreme waardetheorie.

In Hoofdstuk 3 bestuderen we het gedrag van de niet-parametrische schatter van de Gini-index wanneer een stochastische omgeving met een zware staart wordt verondersteld. In het bijzonder bestuderen we de beperkende distributie via de toepassing van de algemene centrale limietstelling voor orderstatistieken en we bewijzen hoe het verlies van symmetrie van de beperkende $\alpha-$ stabiele distributie wanneer het tweede moment oneindig wordt, de bias in de schatting. We suggereren ten slotte het gebruik van een eindige monstercorrectie op basis van de modus-tot-gemiddelde afstand om de consistentie van de schatter te verbeteren.

In Hoofdstuk 4 passen we het begrip van majorisatie toe, een gedeeltelijke volgorde op positieve reële vectoren die sterk verband houden met de Lorenz-curve, om financiële correlatiematrices te bestuderen. In het bijzonder leiden we een aantal axioma's af die ordening over de ruimte van correlatiematrices zouden moeten hebben om financieel risico te vangen. We bewijzen dat de gedeeltelijke volgorde die is verkregen uit de majorisatie van de spectra van correlatiematrices deze axioma's respecteert en dat het kan worden gebruikt om samenvattende maatregelen te bouwen. In het bijzonder laten we zien hoeveel samenvattende maten van correlatiematrices die in de praktijk worden gebruikt, consistent zijn met een dergelijke volgorde. We onderzoeken de geldigheid van een dergelijke bestelling verder door de aanwezigheid ervan in empirische gegevens te controleren. We komen erachter dat, door te kijken naar de Industrial Dow Jones, correlatiematrices ten opzichte van elkaar consistenter zijn geordend vlak voor en tijdens financiële turbulentie. Op basis van deze observatie bouwen we een eenvoudig waarschuwingssysteem dat een signaal genereert door te kijken naar de ordeningstrends van dagelijkse correlatiematrices waaruit blijkt dat een dergelijk systeem veel meer informatie produceert dan slechts een enkele risicomaatstaf.

In Hoofdstuk 5 gebruiken we de geometrie van de Lorenz-curve, een convexe vervorming van de identiteitskaart, om Archimedische generatoren te bouwen voor de constructie van bivariate copula's. We laten zien hoe elke niet-strikte bivariate Archimedische copula kan worden verkregen met behulp van een tweevoud van de Lorenz-curve als generator. We karakteriseren verder de eigenschappen van de rechtstaart van dergelijke copula's in termen van de univariate willekeurige variabele geassocieerd met de Lorenz-curve die wordt gebruikt om ze te overspannen. We laten ook zien hoe de Gini-index is gerelateerd aan de τ -maatstaf van de Kendall en dat de Lorenz en de ster-

stochastische orders kunnen worden gerelateerd aan stochastische multivariate bestellingen. Tot slot bieden we simulaties en algebraïsche formules voor de cumulatieve distributiefunctie en Kendall-functies van sommige van dergelijke gegenereerde copula's.

Ten slotte worden in de conclusies verdere onderzoeksvragen gesteld. We stellen onder andere voor om de geometrische structuur van de Lorenz-curve opnieuw te gebruiken om Pickands-afhankelijkheidsfuncties te bouwen voor toepassingen met extreme waardetheorieën en om de ongelijkheidsindexen te gebruiken als afhankelijkheidsmaatstaven voor diagonale en Archimedische copula's.

CONTENTS

1

INTRODUCTION

1.1. INTRODUCTION

In this thesis we study problems of quantitative risk management using tools that may seem unconventional when looking at the common risk management literature. In particular we are interested in the applicability of tools developed in the social sciences to measure the concentration of wealth in the society: objects like the Lorenz curve and the associated indices.

We argue that these tools posses some analytic and geometric properties that are suitable to efficiently represent risk in portfolios, to construct new risk measures, and to investigate relevant facts and statistical regularities of financial markets.

We hope to convince the reader that Lorenz-based methodologies are powerful and useful when applied to financial risk problems. Additionally–and interestingly–we will show that some of these tools have already been used by risk theorists and practitioners, but without a clear acknowledgment nor understanding.

The goal of this thesis is therefore twofold. First, we provide new results on (portfolio) tail risk in Chapters 2 and 3, on systemic risk in Chapter 4, and on dependence in Chapter 5. Second, we aim at bringing more awareness in the risk management community about the use of the Lorenz curve and its derivations as an additional powerful toolbox.

1.1.1. THE BASIC TOOLS

Let us now introduce the basic objects we will deal with in this thesis, explaining how they are related to each other and providing some history behind.

The contents of this work are built on the contributions of two scholars who lived more than a century ago: the Italian statistician Corrado Gini, and the American economist Max Otto Lorenz.

Corrado Gini, born in 1884 in the Venetian town of Motta di Livenza, was an Italian statistician, founder of the Italian journal of statistics *Metron* and of the statistics faculty of the University of Rome. During his life he studied many different topics spanning

from law and economics, to actuarial sciences, demography and statistics. However, Gini's most famous contribution was probably the study and the development of the inequality measure that inherited his name: the Gini index.

During his studies on the concentration of wealth, Gini developed the following intuition regarding the variability of phenomena studied in statistics. He argued that there exist two fundamentally different types of statistical variability according to the origin of the data at use. One is the variability originated from the measurement error, which is typical of the natural physical world, where a true value exists, such as the position of a celestial body or the height of a mountain, but it is hidden to the observer by the inability of having a precise measurement. The second type of variability, that we shall call *socio-economic variability*, belongs conversely to the socio-economical framework and it reflects a true heterogeneity, like when measuring the wealth of people. Gini reckoned that, given their core differences, these two types of variability should be measured differently. Using Gini's own words [1], "a measure for the variability of natural world's data should answer the question about *how much the different measurements differ from the real value*, while a measure for the variability of the socio-economic world's data should provide information about *how much the different objects differ from each other*."

In particular, Gini pointed out that most of the measures of variability used in the literature, such as the variance, the mean absolute deviation or the median absolute deviation were only appropriate for the first type of variability, and he thus advocated for the development of measures able to capture the second type of variability.

Hence Gini developed the concept of *mean difference*, sometimes denoted by Δ, and later called *Gini mean difference* (GMD) in his honor. Tracking the original formulation of this measure is hard since it has been "rediscovered" many times: Yitzhaki collects more than 12 different formulations [2]. In Equation (1.1) we report the formulation which is believed to be the one originally used by Gini in its monograph *Variabilità e Mutabilità* [3], i.e.

$$GMD = \frac{2}{n(n-1)} \sum_{i=1}^{\frac{n+1}{2}} (n+1-2i)(x_{(n-i+1)} - x_{(i)}), \tag{1.1}$$

where $x_{(1)}, x_{(2)}, ..., x_{(n)}$ is a vector of n positive observations sorted in non-decreasing order. An alternative formulation for the GMD is obtained noting that if we allow the summation to count until n and taking the absolute value of the differences $x_{(n-i+1)} - x_{(i)}$, dividing everything by 2 to avoid double counting we obtain the following expression for the GMD which turns to be quite useful when expressing the GMD for continuous random variables, see Equation (1.4):

$$GMD = \frac{1}{n(n-1)} \sum_{i=1}^{n} (n+1-2i)|x_{(n-i+1)} - x_{(i)}| \tag{1.2}$$

At this point it is important to mention that in 1948, 36 years after Gini first developed his variability measures, Hoeffding in his pioneering work on U-statistics [4] proved that also the variance can be expressed in terms of squared differences between single observations, $(x_i - x_j)^2$. This result may seem to be a major critique to Gini's distinction between socio-economic variability and the physical one, however this is not the case.

In fact, Gini's argues that another key feature of measure of socio-economic variability is the presence of a weighting scheme, called *the rank* that penalizes large differences between observations. In its GMD this weighting scheme is given by the term $n+1-2i$ which increases the contribution of high differences between the observations. Such weighting scheme is not present in Hoeffding's representation of the variance.

It is easy to notice that Equation (1.1) is bounded between 0 and $2\bar{x}$, where $\bar{x}$ is the empirical average of the observations, $\bar{x} = \sum_i^n x_i / n$. Therefore, by applying the scaling $2\bar{x}$ to Equation (1.1), we obtain the celebrated Gini index G which is now bounded between 0 and 1:

$$G = \frac{GMD}{2\bar{x}}. \tag{1.3}$$

Note also that by construction the Gini index attains its lower bound 0 if and only if each individual in the society owns the same amount of wealth, $x_i = x_j \ \forall i, j$, while its upper bound 1 is attained if and only if one individual posses the entire wealth, $\exists! \, i \,|\, x_i > 0$ and $x_j = 0 \ \forall j \neq i$.

It is then legitimate to ask what happens when the size of the population grows to infinity. Equation (1.4), answers this question and provides the *continuos* version of the Gini index–see [2] for a proof. One has

$$G = \frac{\mathbb{E}(|X' - X''|)}{2\mathbb{E}(X)}, \tag{1.4}$$

where X' and X'' are two independent identically distributed copies of the non-negative valued random variable X. In this case, the Gini bounds can be interpreted in terms of the underlying random variable X. One has $G = 0$ if and only if X is a deterministic constant, while if $G = 1$ then the underlying random variable has a non-finite first moment. In particular it is important to notice that the statement $G = 1$ makes sense only when understood as a the limit of a sequence of random variables with quantile functions more and more steep at $F^{-1}(1)$. This condition can be easily derived using the Lorenz curve, object that we will introduce next, see [5] for more details.

By looking at Equation (1.1), or equivalently at Equation (1.4), it should be clear why the Gini index provides an answer to the question asked by Gini [3]. In fact, it provides the average distance between observations without necessarily relying on the concept of average, as a fixed point from which one takes the distances of the measurements.

However–I guess–the reader may still not be fully convinced that this is the proper way to measure socio-economic variability. Luckily, it turns out that the Gini index has a deeper interpretation in terms of another object used to study socio-economic variability, the Lorenz curve, and this makes everything clearer.

The Lorenz curve was introduced in just 35 lines in the PhD thesis of Otto Max Lorenz, in 1905, to graphically represent the inequality, or disparity in the distribution of wealth among the individuals of a population [6].

Given a population on n individuals, each of them endowed with a non-negative wealth $(x_i)_{i=1}^n$, the Lorenz curve was initially defined as:

$$L\left(\frac{i}{n}\right) = \frac{\sum_{j=1}^{i} x_{(j)}}{\sum_{j=1}^{n} x_{(j)}} \quad i = 0, ..., n; \tag{1.5}$$

where $x_{(1)}, x_{(2)}, ..., x_{(n)}$ is again a non-decreasing ordered vector, and where, to obtain a continuous function, a linear interpolation between the coordinates $(i/n, L(i/n))$ is taken. Finally, by convention, $L(0/n) = 0$.

Equation (1.5) reads as follows, the $(i/n)100\%$ of the individuals own $L(i/n)100\%$ of the total wealth or, in other words, the $(1-L(i/n))100\%$ of the total wealth is concentrated in the hands of the richest $(1-i/n)100\%$. The Lorenz curve can then be considered the right mathematical tools to verify the so-called Pareto principle [7], and more in general to analyze other similar wealth concentration statements (e.g. *the* 1% *of the society owns* 99% *of the total wealth*) [8].

The original formulation of the Lorenz curve, provided in Equation (1.5), can be computed when a finite number of data is available. When dealing with continuous quantities, it is necessary to provide a more general expression for the Lorenz curve, involving the quantile function.

Given a positive random variable $X \sim F$, the quantile function $F^{-1}(x) := \inf\{y \in \mathbb{R} : F(y) \geq x\}$ can be understood as the equivalent of the ordered entries $x_{(i)}$ in Equation (1.5).

Observing this, Pietra in 1915 [9] and Gastwirth in 1971 [10] independently defined the continuous version of the original Lorenz curve as

$$L(u) = \frac{\int_0^u F^{-1}(x)\mathrm{d}x}{\int_0^1 F^{-1}(x)\mathrm{d}x}, \tag{1.6}$$

where $F^{-1}(x)$ is the quantile function of a positive-valued random variable X with finite mean and cumulative distribution function $F(y) = P(X \leq y)$. Naturally Equation (1.6) can be interpreted as the limit of Equation (1.5) when the size n of the population goes to infinity. A more precise statement involved a Glivenko–Cantelli-type result and was proven by Goldie [11] in which the almost sure uniform convergence of (1.5) to (1.6) is proven.

Being the quantile function $F^{-1}(x)$ increasing, its integral must be increasing and convex. Additionally, by construction, the Lorenz curve has an upper-bound in $L(1) = 1$. The Lorenz curve can then be understood as a functional that associates to a positive random variable X a suitable increasing convex function in the interval $[0,1]$, which recovers the original distribution function $F(x)$ up to a constant. Note in fact that $F^{-1}(u) = L'(u)\mu$.

In this thesis, mainly in Chapters 4 and 5, we exploit an alternative interpretation of the Lorenz curve, which puts aside its socio-economic nature and rather focuses on its pure geometric structure.

Being increasing and convex, the Lorenz curve is uniformly bounded from above by the identity map. The identity map, being trivially convex, is a Lorenz curve as well, sometimes called the *Perfect Equality* line in contrast to the *Perfect Inequality* line, corresponding to $L(x) = 0$, $\forall x \in [0,1)$, and $L(1) = 1$. Deriving Equation (1.6) with $L(x) = x$, it can be seen that the identity map is the Lorenz curve associated to a degenerate random variable distributed according to a Dirac delta function. From this boundary case, by taking random variables with distribution functions that are more and more spread in the positive half line, we obtain a distortion of the identity map, whose length increases in the just cited spread.

The Lorenz curve can then be understood as a way to generate increasing convex functions whose gradient has a simple characterization in terms of quantile function of positive L^1−integrable random variables.

We now conclude this short introduction on the Lorenz curve by underlying a very interesting and useful connection between the Lorenz curve and the partial order over the set of positive real valued vectors that goes under the name of *majorization*.

Consider a dataset consisting of two arrays $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n)$ of positive data points we define the partial order called majorization in the following way:

Definition 1.1. Take two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We say that $\mathbf{x}$ majorizes $\mathbf{y}$, in symbols $\mathbf{x} \succ \mathbf{y}$, if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{1.7}$$

and

$$\sum_{i=1}^{k} x_{[i]} \geq \sum_{i=1}^{k} y_{[i]}, \text{ for all } k = \{1, ..., n-1\}, \tag{1.8}$$

where $x_{[1]}, ..., x_{[n]}$ are the coordinates of the vector $\mathbf{x}$ sorted in descending order, so that $x_{[1]} \geq x_{[2]} \geq ... \geq x_{[n]}$. If the condition $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ is not satisfied, we speak of weak majorization, $\mathbf{x} \overset{w}{\succ} \mathbf{y}$.

Majorization has been extensively studied in the 20th century in the works of Muirhead [12], Schur [13], Dalton [14] and Hardy, Littlewood, Pólya [15] and the monograph from Marshall and Olkin [16]. Majorization provides a mathematical framework to study the dispersion of the components of positive vectors. In order to understand what kind of notion of dispersion is measured by majorization, we state the following important result due to Hardy, Littlewood, and Pólya [17]:

Theorem 1.1. *Let x and y be two real-valued positive vectors of size n, then if x majorizes y, namely $x \succ y$, then there exists an $n \times n$ doubly stochastic matrix P such that*

$$y = Px. \tag{1.9}$$

Recall that a double stochastic matrix P is a matrix whose rows and columns sum up to one. According to Theorem 1.1, if $x \succ y$, then each component of y can be written as a weighted average of the elements of x, where the weights are the row elements of P, namely $y_i = \sum_{j=1}^{n} x_i p_{i,j}$. Therefore, if vector y is majorized by vector x then y can been understood as a smoothed version of x with smoothing operator the double stochastic matrix P. Geometrically, since every double stochastic matrix can be obtained as a convex combination of some permutation matrix [16], implies that if $x \succ y$ then y belongs to the convex hull spanned by x. In general, the larger the convex hull spanned by a vector the more disperse its components are.

Another important notion strictly related to majorization is the so-called Schur-convex function. By definition, a Schur-convex function $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ of a real, positive, vector x is a function that preserves the majorization ordering, namely if $x \succ y$, then $\phi(x) \geq \phi(y)$[1].

[1] ϕ is said to be Schur-concave if $\phi(x) \leq \phi(y)$

1

Surprisingly, most of the variability measures used as summary statistics to describe dispersion of vectors are Schur-convex functions. Some examples are the standard deviation, the Shannon entropy, the mean absolute deviation, the arithmetic and geometric mean, but also the Gini mean deviation and the Gini index, see [16] for more examples.

Back to the Lorenz curves, using Equation (1.5), the Lorenz curves $L_y(i/n)$, $L_x(i/n)$ associated to each array can be determined. In particular, if the graphs of these curves do not intersect and $L_y(i/n) \geq L_x(i/n)$, then, when denoting by $\hat{x}$ and $\hat{y}$ the original arrays standardized by their mean, $\hat{x}_i = \frac{nx_i}{\sum_i^n x_i}$, with the convention that $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$, the following relation holds:

$$\sum_{i=1}^{k} \hat{y}_{(i)} \geq \sum_{i=1}^{k} \hat{x}_{(i)}, \quad \forall k = 1, ..., n-1, \tag{1.10}$$

and trivially

$$\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} \hat{x}_i. \tag{1.11}$$

The set of conditions described in (1.10) and (1.11) is easily proven to be equivalent to those appearing in the definition of majorization. Therefore, the following statement connecting majorization and the ordering of graphs of Lorenz curves holds.

Proposition 1.1. *Let $x, y \in \mathbb{R}_n^+$ and let $L_x(u)$, $L_y(u)$ be their Lorenz curves, if $x \succ y$, then $L_x(u) \leq L_y(u)$ $\forall u \in [0,1)$.*

The proof is quite trivial and can be found in [18]. The result stated in Proposition 1.1 shows that the order induced by looking at the behaviour of Lorenz curves is weaker than majorization and allows to compare vectors when the condition (1.11) is not met, making the comparison between population with different total wealth possible, or as common in risk management when the total losses of the different portfolios differ.

Finally, we briefly mention that the results on majorization and Lorenz curves that so far have been stated in terms of real vectors can be extended to continuous random variables. Extension useful when dealing with financial models.

By using Gastwirth's or Pietra's representation of the Lorenz curve, the notion of Lorenz order between random variables can be defined. Namely, let X and Y be two positive real-valued random variables with finite mean with Lorenz curves $L_X(u)$, $L_Y(u)$ then we say that X is larger than Y in the Lorenz sense $X \geq_L Y$, if and only if $L_X(u) \leq L_Y(u)$ $\forall u \in [0,1]$. In particular, the Lorenz order is strictly related to the so-called convex order [19], a well-known stochastic order which ranks random variables in terms of their appeal to a risk adverse individual [20], assumption often used in optimal portfolio choice as the Nobel price winner Markowitz showed [21].

With the concept of Lorenz curve and its related partial order in mind we are now ready to provide a second characterization of the Gini index, which was first discovered by Pietra in 1915 [9], i.e.

$$G = 1 - 2\int_0^1 L(p)\mathrm{d}p. \tag{1.12}$$

Note that after some algebraic manipulation we can re-write Equation (1.12) as

$$G = \frac{\int_0^1 |p - L(p)| \mathrm{d}p}{\int_0^1 |p| \mathrm{d}p}. \tag{1.13}$$

The expression in Equation (1.13) can be interpreted as the normalized L^1-distance between the perfect equality line and the Lorenz curve of the data. With this interpretation and the results regarding the Lorenz curve at hand, it is straightforward to conclude that the Gini index properly summarizes the information on the socio-economic variability embedded in the Lorenz curve measuring how far is the society from a state in which each individual owns the same amount of wealth. Moreover, by construction, the Gini index is consistent with the Lorenz order providing a necessary condition for two datasets to be ordered in the Lorenz sense. All these properties makes the Gini index a great measure for socio-economic variability.

Clearly, only one number, as the Gini index provides, is not sufficient to describe the entire behaviour of the Lorenz curve. For this reason many other variability measures have been developed in the literature. An example is given by the *distance indices* D_P, [22], which are obtained by generalizing the distance in Equation (1.13) to any L^p-distance:

$$D_p = \frac{||p - L(p)||_p}{||p||_p}. \tag{1.14}$$

Each distance index studies a different feature of the Lorenz curve and, as positive byproduct, a feature of the underlying data. For example by taking $p = \infty$ the Pietra Index [22] is obtained. Other examples of indices build over the Lorenz curve are developed in the literature, for example indices based on the length of the Lorenz curve via the Amato index [23], its curvature [22], or its self-symmetry [24].

On a more general note, hundreds of summary measures for the study of the Lorenz curve have been developed over the years. These measures are referred to as *socio-economic variability indices*, concentration indices or inequality measures[2], and more information can be found in [16, 18, 22].

Recognizing the possibility of applying the Lorenz curve to the study of a problem thus allows us to use a large variety of indices and measures that have been developed for more than 100 years, and which are likely to provide new information and insights on the problem under scrutiny.

In this thesis we will apply exactly this line of reasoning, all the different topics we will deal with are indeed united under a large portmanteau: they all show a *Lorenzian* structure, and they can thus be studied accordingly.

1.2. A GUIDE TO THE THESIS

The content of this thesis can be collected into two parts. The first one is a direct application of the original interpretation of the Lorenz curve as a tool to analyze the socio-economic variability of a dataset. In Chapters 2 and 3 we use the Gini index to study loss distributions and to draw conclusions over their tail risk.

[2] Not to be confused with concentration measures for the tail probabilities.

In the second group, i.e. Chapters 4 and 5, one finds the works in which the focus is more on the geometrical interpretation of the Lorenz curve.

Following Gini's argument on the nature of variability, in Chapter 2 we recognize that the variability usually observed in financial losses is of the socio-economic type, in other words there is no true value for losses to be measured with an error, but each different loss carries specific information, as it has manifested itself because of some deeper and complex mechanism that governs financial markets.

In quantitative risk management, the tail losses average, also known as Expected Shortfall ES_α, is a widespread measure, used on a daily basis. To account for Gini's critique, we thus propose a Lorenz-based approach to study the variability of the losses in a financial portfolio. In particular, we define the truncated version of the Gini index, as a function of the Value-at-Risk VaR_α and call it *Concentration Profile* $G(\alpha)$, for $\alpha \in [0,1]$. As a consequence of this procedure for a fixed α, $G(\alpha)$ corresponds to the Gini index of a new random variable X_α with support $[VaR_\alpha, c)$, with $c \leq \infty$, and expectation $\mu_{X_\alpha} = ES_\alpha$.

We hence argue that for fixed α, this measure provides a good degree of reliability of the Expected Shortfall once a Value-at-Risk α-level has been specified. In fact, by recalling the definition of Gini index, one knows that if $G(\alpha) = 0$ then the ES_α is the only possible realizable outcome for the loss distribution above the VaR_α, making the Expected Shortfall a reliable representation of tail risk, i.e. tail losses are more concentrated around ES_α. On the opposite side, values of $G(\alpha)$ closer to 1 signals a high variably of the truncated loss distribution, making the Expected Shortfall a less precise approximation for the losses being the structure of the losses a random variable with some infinite moments.

Furthermore, we prove that this new measure is able to characterize the loss distribution function up to a constant, opening up to the possibility of using the Concentration Profile as a tool to compare different loss distributions consistently with the Lorenz order. In particular, starting from the Concentration Profile, we propose a 2-dimensional map, the *Concentration map* that assesses the riskiness of portfolios according to their loss variability (and the risk aversion of the analyst).

Finally we study the behaviour of the Expected Shortfall when weighted by the corresponding truncated Gini. We show how this measure, called *Concentration adjusted Expected Shortfall*, exhibits different limits according to the parametric family the loss distribution belongs to. We argue that this type of measure can help to identify and distinguish classes of parametric distributions and can be a valuable tool for model selection.

Being the building block of the Concentration profile, the Gini index must be studied carefully, in particular the properties of its estimators. In Chapter 3 we analyze the asymptotic behaviour of two types of Gini index estimators under the assumptions of fat-tailed data with infinite variance–a common situation in the real financial world. We compare the asymptotic performance of the maximum likelihood estimator for the Gini index, when a Generalized Pareto distribution structure is assumed for the tails, against a naive plug-in estimator. In particular, we obtain the α-stable limiting distribution for the plug-in estimator and show its asymmetry when the underlying stochastic environment exhibits infinite variance. This reflects into a potential pre-asymptotic bias of the estimator.

This result is relevant in risk management because, being risk measures usually computed with historical data (and usually via non-parametric estimators), if the data exhibits heavy-tailed behaviour with infinite variance then the estimation bias could lead to an underestimation of risk.

Chapter 4 deals with systemic risk. Mathematically, this type of risk is often studied by modelling the dependence structure among the components of a financial portfolio. As first approximation the dependence is assumed to be monotonic and captured by the correlation matrix (in its more general definition). Being usually high-dimensional, one of the main problems when dealing with correlation matrices is to find the right way to compare them, so that a proper notion of risk is possible.

By noting that the spectra of correlation matrices for portfolios of the same size are positive real vectors summing up to the same value, we propose majorization as a systemic risk ordering. We then make use of Schur-convex functions theory to develop functions of the correlation matrix that are consistent with the majorization order and provide summary values of the overall risk embedded in the market.

As an application, we provide evidence of the presence of majorization in financial markets by back-testing its presence over a period of almost 20 years for components of the Industrial Dow Jones. We verify that during times of financial turbulence the number of successful majorization relations spikes. As expected, correlation matrices belonging to financial distress times majorize those of less turbulent periods.

Finally, we suggest that a new paradigm for risk evaluation could be built by studying the properties of the Directed Acyclic Graph that is naturally associated to the partial order of majorization. This allows for the study of systemic risk using tools from complex networks, such as centrality measures, communities and flow diagrams.

In Chapter 5, we leverage on the notion of Lorenz curve as a convex distortion to generate bivariate non-strict Archimedean copulas. Copulas are an important tool in risk management since they allow for a simple representation of the dependence among financial risks. Deriving new models that are easy to study is then always a plus.

To generate Archimedean copulas the main ingredient is the so-called generator. The generator is a one-place decreasing convex function with prescribed boundaries. We observe that to any Lorenz curve it is possible to pair a particular dual curve–the *mirrored Lorenz*–which represents a proper generator. We therefore take the mirrored Lorenz curve as a Archimedean generator and we use it to derive the new class of *Lorenz copulas*, which we study in detail, starting from the properties of the underlying univariate random variable X generating the Lorenz curve. For example, we establish a connection between univariate stochastic orders and multivariate orders based on copulas, and we demonstrate how asymptotic tail dependence in Lorenz copulas is only obtained when the Lorenz curve of a log-normal distribution is used.

All the chapters composing this thesis derive from papers that have been published or submitted to peer-reviewed journals. However, we want to underline that the versions presented here have been slightly modified to unify notation, where possible, and to account for minor revisions that are presented in the form of *errata corrige* footnotes.

We wish to conclude this introduction with a remark over the spirit that led me to the creation of the works composing this thesis. Using Corrado Gini's own words [1], we have approached risk management *"bearing in mind, in formulating [my] statistical methods,*

first the nature of the phenomena to which they are applied in relation to the object of the research, and only secondarily their formal properties".

REFERENCES

[1] L. Ceriani and P. Verme, *The origins of the gini index: extracts from variabilità e mutabilità (1912) by corrado gini,* The Journal of Economic Inequality **10**, 421 (2012).

[2] S. Yitzhaki and E. Schechtman, *More than a dozen alternative ways of spelling gini,* in *The Gini Methodology* (Springer, 2013) pp. 11–31.

[3] C. Gini, *Variabilità e mutabilità,* Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1912).

[4] W. Hoeffding, *A class of statistics with asymptotically normal distribution,* in *Breakthroughs in Statistics* (Springer, 1992) pp. 308–334.

[5] C. Kleiber and S. Kotz, *Statistical size distributions in economics and actuarial sciences.*, Vol. 470 (John Wiley & Sons, 2003).

[6] M. O. Lorenz, *Methods of measuring the concentration of wealth,* Publications of the American statistical association **9**, 209 (1905).

[7] V. Pareto, *Manuale di economia politica*, Vol. 13 (Societa Editrice, 1906).

[8] S. Van Gelder, *This changes everything: Occupy Wall Street and the 99% movement* (Berrett-Koehler Publishers, 2011).

[9] G. Pietra, *Delle relazioni tra gli indici di variabilità* (1914).

[10] J. L. Gastwirth, *A general definition of the lorenz curve,* Econometrica: Journal of the Econometric Society , 1037 (1971).

[11] C. M. Goldie, *Convergence theorems for empirical lorenz curves and their inverses,* Advances in Applied Probability **9**, 765 (1977).

[12] R. F. Muirhead, *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters,* Proceedings of the Edinburgh Mathematical Society **21**, 144 (1902).

[13] I. Schur, *Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie,* Sitzungsberichte der Berliner Mathematischen Gesellschaft **22**, 51 (1923).

[14] H. Dalton, *The measurement of the inequality of incomes,* The Economic Journal **30**, 348 (1920).

[15] G. Polya, G. H. Hardy, and Littlewood, *Inequalities* (University Press, 1952).

[16] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*, Vol. 143 (Springer, 1979).

[17] G. H. Hardy, *Some simple inequalities satisfied by convex functions,* Messenger Math. **58**, 145 (1929).

[18] B. C. Arnold and J. M. Sarabia, *Majorization and the Lorenz order with applications in applied mathematics and economics* (Springer, 2018).

[19] M. Shaked and J. G. Shanthikumar, *Stochastic orders* (Springer Science & Business Media, 2007).

[20] A. B. Atkinson *et al.*, *On the measurement of inequality,* Journal of economic theory **2**, 244 (1970).

[21] H. Markowitz, *Portfolio selection,* The journal of finance **7**, 77 (1952).

[22] I. Eliazar, *A tour of inequality,* Annals of Physics **389**, 306 (2018).

[23] B. C. Arnold, *On the amato inequality index,* Statistics & Probability Letters **82**, 1504 (2012).

[24] C. Damgaard and J. Weiner, *Describing inequality in plant size or fecundity,* Ecology **81**, 1139 (2000).

2

GINI BASED RISK MEASURES: CONCENTRATION PROFILE

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.

Keywords: *Concentration measures; Value-at-Risk; Expected Shortfall; Concentration Profile; Gini index.*

Parts of this chapter have been published in Insurance: Mathematics and Economics **78**, 13–29 (2018) [1].

2.1. INTRODUCTION

This chapter 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 are available, with no particular reference to the source of risk.

Following a common convention in risk management [2], 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 [3], and the dependence among losses is not a major source of concern[1]; an example being the one-year credit loss distribution, commonly used by banks under the Basel framework [7, 8] with the so-called historical simulation approach [2].

Value-at-Risk (VaR) and Expected Shortfall (ES) represent two important risk measures in modern risk management [2, 6–8]. 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 representativeness of the ES in representing tail risk. Empirical studies [2, 6, 9, 10] 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 [10–12] to analyze the dispersion of risk in the tail, and we focus our attention on the Lorenz curve [13] and the corresponding Gini index [14], for which we 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, for example it allows to make more precise inference on possible tail behaviours with respect to other type of tools such as the quantile function or the sequence of Expected Shortfalls. 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.

[1] Several interesting dynamic approaches dealing with dependence and time evolution also exist, for example in operational risk [4, 5], or in market and credit risk (see [6] and references therein), but we do not deal with them here.

[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 by Shalit and Yitzhaki [15, 16].

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 [17], 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 study of the the Gini index of truncated distributions it is not new to statistical and socio-economic literature, [18] for examples studies the effect of a left truncation on the Gini index. However, our work focuses more on the applications of such a tools on financial data rather than the study of their functional properties. Additionally, while [18] aims at deriving distributions given a pre-specified form for the truncated Gini, we take the opposite route and from a given distribution of the losses we derive the expression of the truncated Gini index and use it to study data proprieties.

The chapter is structured as follows: in Section 2.2 we briefly review some basic concepts about concentration measures, while in Section 2.3 some common measures of risk used in risk management are analyzed and put in relation with these concentration measures; in Section 2.4, we introduce and study the Concentration Profile, and in Section 2.5 we describe the Concentration Map; in Section 2.6 some additional extensions based on the Concentration Adjusted Expected Shortfall are discussed, whereas in Section 2.7 empirical results on simulated and actual data are provided; finally Section 2.8 closes the chapter. 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.

2.2. BASIC CONCENTRATION QUANTITIES

2.2.1. THE LORENZ CURVE

Introduced by Max Lorenz in 1905 [13], the Lorenz curve is a pivotal tool in the study of economic inequality and the distribution of wealth in the society [10].
Consider a positive continuous random variable Y, belonging to the $\mathscr{L}^1$ 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 $F^{-1}(\alpha) = \inf\{y : F(y) \geq \alpha\}$ with $0 \leq \alpha \leq 1$. The Lorenz curve $L(x)$ is formally given by

$$L(x) = \frac{\int_0^x Q(\alpha)\mathrm{d}\alpha}{\int_0^1 Q(\alpha)\mathrm{d}\alpha}, \; 0 \leq x, \alpha \leq 1. \tag{2.1}$$

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

Mathematically, the Lorenz curve $L : [0,1] \to [0,1]$ defined in Equation (2.1) 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.

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

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}, reachable only as limiting case for continuous random variables, states that only one individual owns all the wealth in the society. A visual representation of a possible Lorenz curve is given in Figure 2.1, where we also provide L_{pe} and L_{pi}.

Given its strong relation with the quantile function Q, the Lorenz curve can recover the cumulative distribution of Y up to a constant [10]. However, despite the Lorenz curve is theoretically a one-to-one mapping with a given distribution, discriminate among distributions just looking at their Lorenz curves [19] it is not an easy task to perform *by hand*. For example, a curve like the one in Figure 2.1 may give an indication of how unequal a society is, but it does not provide an easy-to-recognize visual pattern to be used to identify by which underlying distribution such inequality is generated.

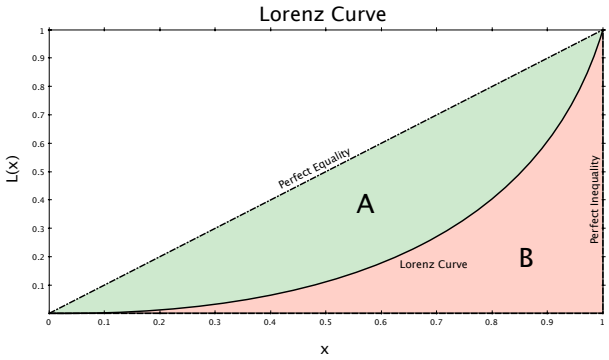


Figure 2.1: Graphical representation of a Lorenz curve, of the perfect equality and of the perfect inequality lines. We also show the geometric interpretation of the Gini index as the ratio of area A over A+B.

2.2.2. THE GINI INDEX

Given a Lorenz curve, and following [11], we define a general concentration (or inequality) measure as

$$D_p = \frac{\left\| L(x) - L_{pe}(x) \right\|_p}{\left\| L_{pi}(x) - L_{pe}(x) \right\|_p}, \tag{2.2}$$

where the denominator is for normalization purposes, so that

$$0 \le D_p \le 1 \quad \forall p \in [1, \infty]. \tag{2.3}$$

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 Equation (2.2) it is clear that, by varying the type of distance $\|\cdot\|_p$, we can define different indices: the Gini index [14] 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 [11], here we adopt the following:

$$G = \frac{\mathbb{E}(|\ Y_1 - Y_2\ |)}{2\mathbb{E}(Y)} = \frac{\mathbb{E}((Y_1 - Y_2)^+)}{\mathbb{E}(Y)} \in [0,1], \tag{2.4}$$

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 (2.4) 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 [15].

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 $\mathscr{L}^1$-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 [4, 6, 10, 20].

- The Gini index is a quasi-convex measure [21]. With a little abuse of notation, if $G(x)$ is the Gini index of a data set $X = (x_1, ..., x_n)$ from a distribution $F(x)$, and $G(y)$ is the Gini index associated with another distribution $H(y)$ and $Y = (y_1, ..., y_n)$, it can be shown that

  $$G(\lambda X + (1-\lambda)Y) \leq \max(G(X), G(Y)), \qquad \lambda \in [0,1].$$

  In words: the Gini index of a data set obtained as linear convex combination of two data sets, e.g. $(\lambda x_1 + (1-\lambda)y_1, ..., \lambda x_n + (1-\lambda)y_n)$, 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 [22], with important consequences for sub-additivity and risk diversification, and thus risk management in general, e.g. [23, 24] and references therein. In particular, by quasi-convexity we may handle distributions of risk that are not necessarily closed under convolution [9, 15, 16].

- 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 [10]. As [11] suggests, the Gini index can also be considered as an $\mathscr{L}^1$ alternative to the skewness coefficient, for measuring the asymmetry in the data, in particular to the right. As shown in [25], the numerator in Equation (2.4) moves in the same direction as the skewness coefficient, when the latter is defined (i.e. finite).

In Appendix A, Tables 2.3 and 2.4, 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 chapter.

2.3. 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 [10, 20].
- The loss distribution is usually fat-tailed [4, 6, 9, 10], 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 [7, 8], but also Solvency II[4] 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 chapter.

Definition 2.1. 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.

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

[3] In the case of operational risk, even the first moment, the mean, may be infinite [4, 5, 26], but we ignore this radical case here. Naturally, in such a situation, the Gini index itself would not be defined.
[4] Solvency II, European Union Directive 2009/138/CE (2009)

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 2.2. 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 2.3. Given a VaR_α, the Expected Shortfall ES_α, for a positive Y with c.d.f. $F(y)$, is given by

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

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

$$\mathbb{E}(Y \mid Y > u) = \int_u^{+\infty} y f_u(y) \mathrm{d}y, \qquad u > 0.$$

VaR_α and ES_α are two fundamental measures of risk in modern risk management [2, 6, 8]. It is well-known that, while the ES_α is a coherent risk measure (positive homogeneous, monotone, translation invariant and sub-additive), the VaR_α is not, unless we restrict our attention to elliptic loss distributions and co-monotonic portfolios [3, 9]. Therefore, in the recent years, ES_α appears to be preferred by regulators [8], even if both measures have been criticized by experts [6], for their incapacity of dealing with the dispersion of losses in the tails[5]. 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 *for a fixed* 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.

2.4. 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

[5] As far as ES_α is concerned, there are also issues related to back-testing [2, 27, 28].

inspection of the Lorenz curve does not identify any particular distribution [10]. 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 [12] 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.

2.4.1. MATHEMATICAL CONSTRUCTION

Let us introduce some notation.

Definition 2.4. 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

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

As $\alpha \to 1$, 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 $S_\alpha = \left[VaR_\alpha, y_F\right)$.

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

$$L_\alpha(x) = \frac{L(\alpha + (1-\alpha)x) - L(\alpha)}{1 - L(\alpha)}, \qquad 0 \le \alpha, x \le 1, \tag{2.5}$$

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

$$D_\alpha^T = \frac{\left\| L_\alpha(x) - L_{pe} \right\|_p}{\left\| L_{pi} - L_{pe} \right\|_p} \qquad 0 \le D_\alpha^T \le 1, \tag{2.6}$$

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

$$G(\alpha) = 1 - 2\int_0^1 L_\alpha(x)\mathrm{d}x = 1 - 2\int_0^1 \frac{L(\alpha + (1-\alpha)x) - L(\alpha)}{1 - L(\alpha)}\mathrm{d}x. \tag{2.7}$$

In terms of VaR_α and ES_α, Equation (2.7) 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 2.2, 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 $G(\alpha)$ allow for a higher probability of loss dispersion:

$$\mathbb{P}(|Y_1 - Y_2| > 2ES_\alpha | Y_1 \wedge Y_2 > VaR_\alpha) \leq G(\alpha) \leq 1. \tag{2.8}$$

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

Definition 2.5. 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 Figure 2.2 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 $\gamma = 1$).

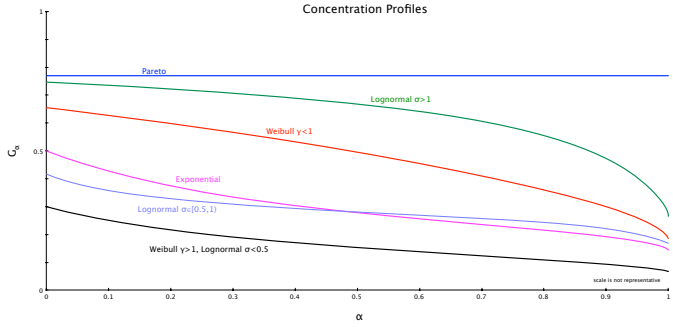


Figure 2.2: Theoretical Concentration Profiles for different distributions. Apart from the exponential case, the parameters of the distribution influence the Concentration Profile. For the Pareto distribution, the horizontal line is characteristic, but the height of the line depends on the shape parameter.

Figure 2.2 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.

2.4.2. 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 $F(y)$.

Theorem 2.1. *If $G(\alpha)$ is differentiable, then*

$$Q(\alpha) \propto \frac{(1-\alpha)G'(\alpha) - G(\alpha) + 1}{(1-\alpha)(1+G(\alpha))} \exp\{\int_0^\alpha \frac{(1-p)G'(p) - G(p) + 1}{(1-p)(1+G(p))} dp\}, \tag{2.9}$$

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

Theorem 2.1 whose proof is a simple application of the Fundamental Theorem of Calculus, complements the results obtained in [18] on the characterizing properties of the right-truncated Gini index. Examples of Concentration Profiles, obtained via Equation (2.7), are given in Table 2.1 for different distributions.

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	$G(\alpha) = \frac{\int_\alpha^1 (-1+2x-\alpha)e^{-\sqrt{2}\sigma\,\mathrm{erfc}^{-1}(2x)}\mathrm{d}x}{(1-\alpha)\int_\alpha^1 e^{-\sqrt{2}\sigma\,\mathrm{erfc}^{-1}(2x)}\mathrm{d}x}$
Exponential	$G(\alpha) = \frac{1}{2-2\log(1-\alpha)}$
Pareto	$G(\alpha) = \frac{1}{(2\rho-1)}$ for $\rho \in (1,+\infty)$

Table 2.1: Theoretical Concentration Profiles for some loss distributions, here $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function $\Gamma(a,b) = \int_b^{+\infty} e^{-x}x^{a-1}\mathrm{d}x$ and $\mathrm{erfc}^{-1}(x)$ is the inverse error function.

Another property of every CP is that it represents a characterizing tool for specific classes of distributions. As anticipated in Figure 2.2, 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* [9], which is a well-known distribution family used in extreme value theory [29], 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

$$H(z) = \begin{cases} 1-(1+\xi z)^{-1/\xi} & \xi \neq 0, \\ 1-e^{-z} & \xi = 0. \end{cases} \tag{2.10}$$

As discussed in [9], 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 Equation (2.10) 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 [9, 30], setting a high threshold defining

the tail, and 2) some non-restrictive technical conditions are met [29, 31, 32]. 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 [26] for more details) and not easily estimated [9]. 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 [9, 10, 29].

Theorem 2.2. *Consider the GPD, its Concentration Profile is given by the following formula:*

$$G(\alpha) = \begin{cases} \frac{\xi}{2-(1-\alpha)^{\xi}(-2+\xi)(-1+\xi)-\xi} & \xi \in (0,1), \\ \frac{1}{2-2\log(1-\alpha)} & \xi = 0. \end{cases} \tag{2.11}$$

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

See Appendix D for proof.
For every $\alpha \in [0,1)$, Equation (2.11) 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 Equation (2.11) tends to zero. For the special case of a purely fat-tailed distribution like the Pareto, we get the horizontal line in Figure 2.2. 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 2.1.

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 the fatter the tail, as also shown in Figure 2.3a, 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 Figure 2.2, 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 $\alpha \to 1$, $ES_\alpha \approx Var_\alpha$ [9, 27].

Let us consider in more detail the exponential CP, whose analytical formula is given in Table 2.1 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

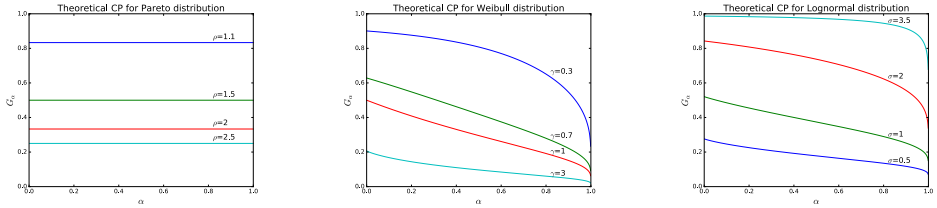
empirical $G(0)$ is far away from 0.5 cannot be of the exponential type. This is a useful heuristic we can derive from the CP.

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 $\gamma > 1$ 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 $\gamma > 1$ [10]. 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 2.7), 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 [6, 9]. 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 [10] 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 chapter 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 Figure 2.3, we show the theoretical CP of the Pareto, the Weibull and the lognormal distributions for different values of their relevant parameters.



(a) Pareto CP. (b) Weibull CP. (c) Lognormal CP.

Figure 2.3: Theoretical Concentration Profiles for the Pareto, the Weibull and the lognormal distributions, for different values of their relevant parameters.

2.5. 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*.

2.5.1. 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 2.6. 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 Figure 2.3a and Table 2.1. 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 2.7. 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 $\alpha \to 1$, i.e.,

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

From Theorem 2.2 we know that, for light-tailed distributions, the CP converges to zero as $\alpha \to 1$. 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 Figure 2.4, we show a graphical representation of r_1 and r_2.

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 $R(r_1, r_2) : [0,1] \times [0,1] \to [0,1]$ 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:

$$\frac{dR(r_1,r_2)}{dr_1} > 0 \qquad \text{and} \qquad \frac{dR(r_1,r_2)}{dr_2} < 0. \tag{2.12}$$

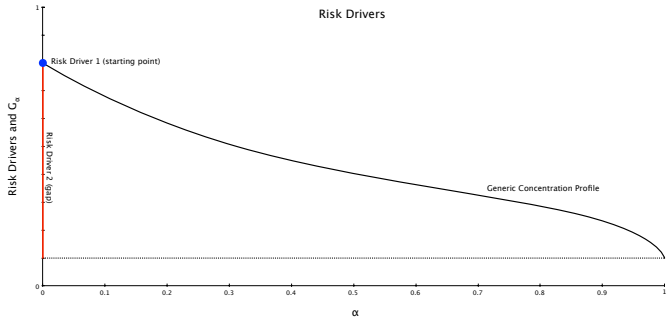


Figure 2.4: Risk drivers' visualization with respect to a generic Concentration Profile.

Let Γ be the set of functions $R(x,y) : [0,1] \times [0,1] \to [0,1]$ that respect the constraints in Equation (2.12). If we additionally impose 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 additional constraint, we have $r_1 \geq r_2$.

Looking for a suitable risk function, we rely on well-known results from utility and risk theory, e.g. [17]. For example, we can choose a Cobb-Douglas risk function,

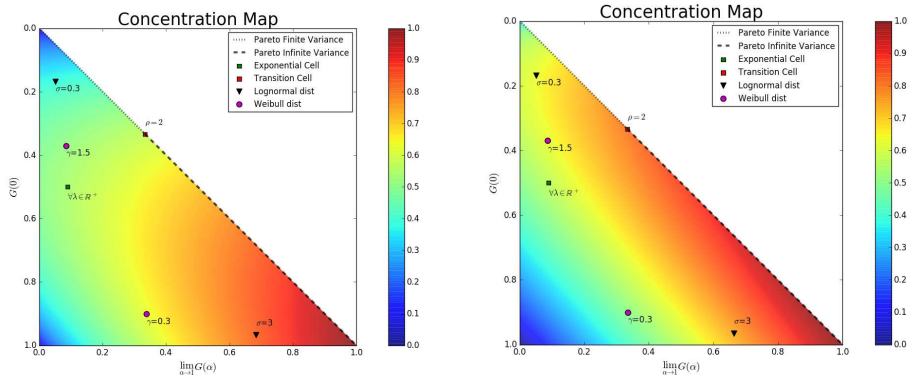
$$R(r_1,r_2) = r_1^a(1-r_2)^b \qquad \text{with } a,b \in \mathbb{R}^+. \tag{2.13}$$

This function allows a risk manager to decide which risk driver is relevant, by acting on the parameters a and b. In particular the higher a the more weight is given to the first risk driver, while the lower b the more weight is given to the second risk driver. Heuristically, the choice on how to assign weights can be given according to the following scheme. High a is global risk is the issue, low b is tail risk is the issue. 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 [17, 33]. Other functions may be used as well.

Figures 2.5a and 2.5b 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 $\lim_{\alpha\to 1} G(\alpha)$ 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 Figures 2.5a and 2.5b, 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 $Pareto(\rho \approx 1)$ corresponds to the point $(1,1)$, while a $Pareto(\rho \to +\infty)$ tends to $(0,0)$. This is because the CP associated to the Pareto



(a) $a = 0.5$, $b = 0.5$.

(b) $a = 0.2$, $b = 0.8$.

Figure 2.5: 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 $\lim_{\alpha\to 1} G(\alpha)$.

distribution is constant, so that the risk driver $r_2 = 0$ (since $G(0) = \lim_{\alpha\to 1} G(\alpha)$), 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 [17]. For example, the points placed on the hypotenuse, from $(0.\overline{33}, 0.\overline{33})$ to $(1,1)$ – what we call the transition cells in Figures 2.5a and 2.5b, 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.

2.6. 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 $\alpha \to 1$, 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[6], 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 2.8. 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:

$$\lim_{\alpha\to 1} CAES_\alpha = \infty, \tag{2.14}$$

$$\lim_{\alpha\to 1} CAES_\alpha = c, \tag{2.15}$$

and

$$\lim_{\alpha\to 1} CAES_\alpha = 0, \tag{2.16}$$

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

The case in Equation (2.14) 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 Equations (2.15) and (2.16), 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 $\alpha \to 1$. The situation of Equation (2.15) occurs for example for the exponential distribution, while that of Equation (2.16) manifests itself for Weibulls with $\gamma > 1$ and lognormals with $\sigma < 0.3$.

An extra remark is to be made for the lognormal distribution. Figure 2.6b 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.

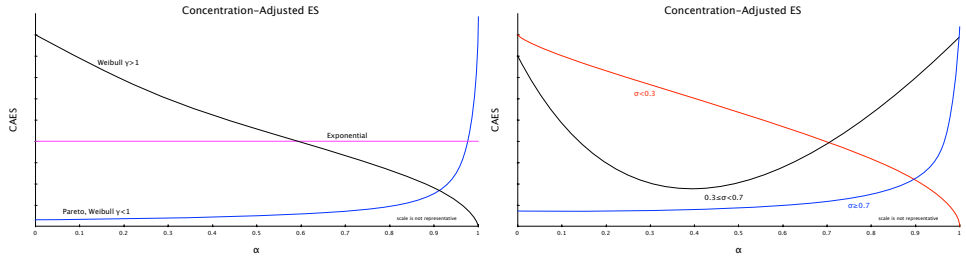
Figure 2.6a presents plots of the theoretical behavior of the $CAES_\alpha$ for the Pareto, the Weibull and the exponential, while Figure 2.6b deals with the lognormal.

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.

2.7. 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

[6] 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 [19].



(a) Pareto, Weibull and exponential.

(b) Lognormal for different values of σ.

Figure 2.6: Theoretical behavior of the Concentration Adjusted Expected Shortfall $CAES_\alpha$ for different distributions.

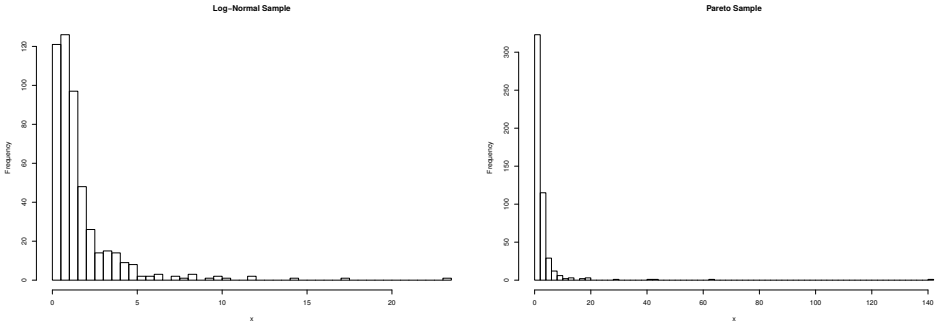


Figure 2.7: On the left, the histogram of 500 lognormally distributed data points with $\mu = 0$ and $\sigma = 1$. On the right, the histogram of 500 Pareto distributed data points with $\rho = 1.5$.

data sets as a basis for our analysis. Finally, in Subsection 2.7.3, 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.

2.7.1. 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 Figure 2.7 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 [12]. For example, one could use

$$\hat{G} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n\sum_{i=1}^{n} x_i}. \tag{2.17}$$

2

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 [30], we set k between 1% and 5% of the original ordered data points.

When estimating the Gini index, attention must be paid to consistent estimates. According to [34], the estimator in Equation (2.17) (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 2.4 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 [9]. 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 [9, 31]. Confidence intervals for the CP can be computed using resampling techniques like bootstrap, jackknife or k-jackknife [35, 36].

Figure 2.8 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 $\mathbb{E}(|X - \mathbb{E}(X)|)$ 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 [37, 38]. Totally compatible results are also obtained by using percentile bootstrap confidence intervals for the Paretian CP. Figure 2.8 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.

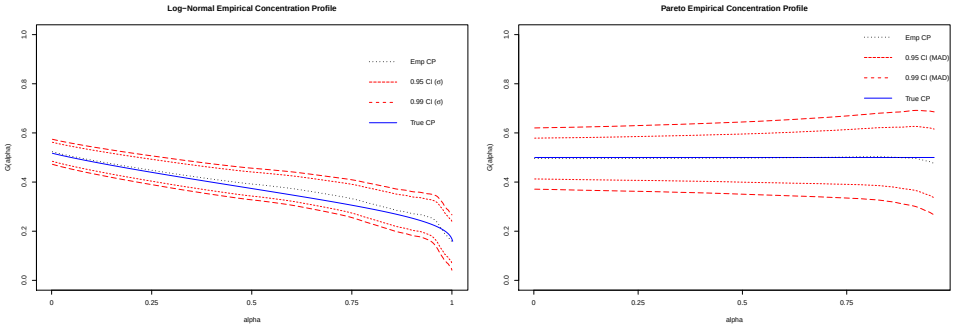


Figure 2.8: Empirical Concentration Profile (dotted black line) with 95% and 99% bootstrapped confidence intervals (red dotted lines) compared to the theoretical Concentration Profile (blue straight line). Confidence intervals are built using the standard deviation for the lognormal data and the mean absolute deviation for the Paretian ones.

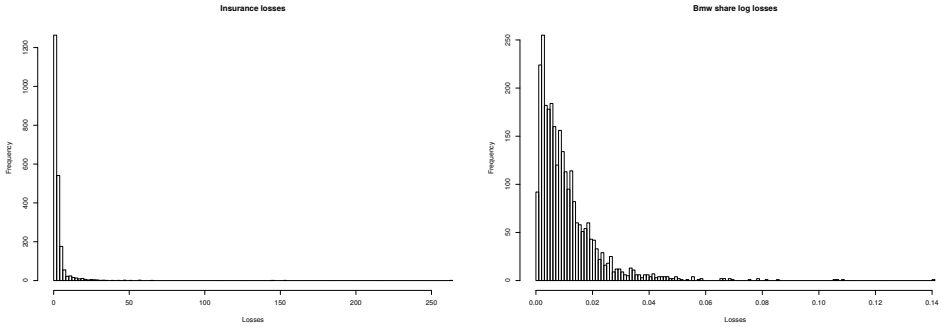
For the Paretian case, conversely, the coverage is around 75%. This seems unsatisfactory, but it can be justified by the fact that, for $\rho = 1.5$, the variance of the Pareto is not defined, and a sample of 500 observations may not be sufficient to correctly estimate ρ, given the large fluctuations in the samples [39][7]. 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 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.

2.7.2. REAL DATA EXAMPLE

We deal with two well-known data sets in the *evir* package of the statistical language R, used in, e.g., [6, 9, 29]. 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

[7] Note that in this case we used a sequence of marginal confidence intervals to estimate the coverage probability, despite this being standard practice confidence bands could also be used in order to improve accuracy, see for example [40]



(a) Danish insurance claims.

(b) BMW losses.

Figure 2.9: Histograms of the Danish insurance losses and of the BMW losses.

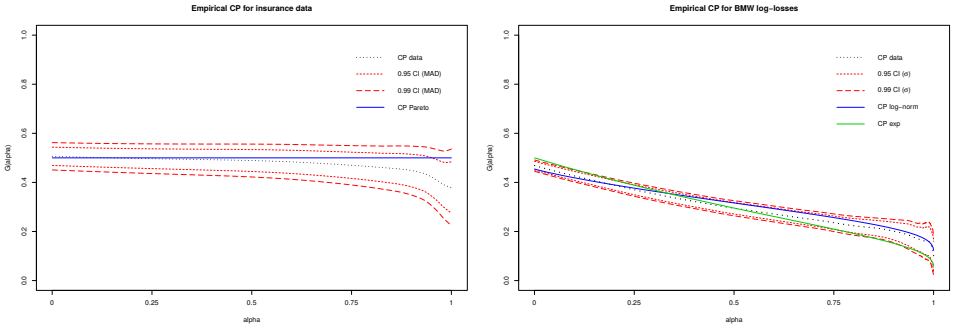
size). The second data set contains 6146 daily losses and returns on the BMW share price (name: BMW), but here we only consider the 2769 losses, which are known to be lognormally distributed. The empirical loss functions of the data are shown in Fig 2.9 as histograms.

Figure 2.10 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 [9]. For this reason, to not completely lose the dependence structure, a simple block-bootstrap approach [41] can be used[8], 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 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 [29], but it is better not to underestimate them when defining the CP for the lower thresholds.

In Figure 2.10a, the empirical CP of the fire insurance claim data is almost constant. Given the insights in Section 2.4, 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 Figure 2.11a.

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 2.4, we can exclude the Pareto distribution since the CP is not constant.

[8] More advanced approaches like the moving, the circular [41] or the stationary block bootstrap [42] could have been used, and we actually tried some of them, but our qualitative results about the shape of the CP do not vary.



(a) Danish insurance claims. (b) BMW losses.

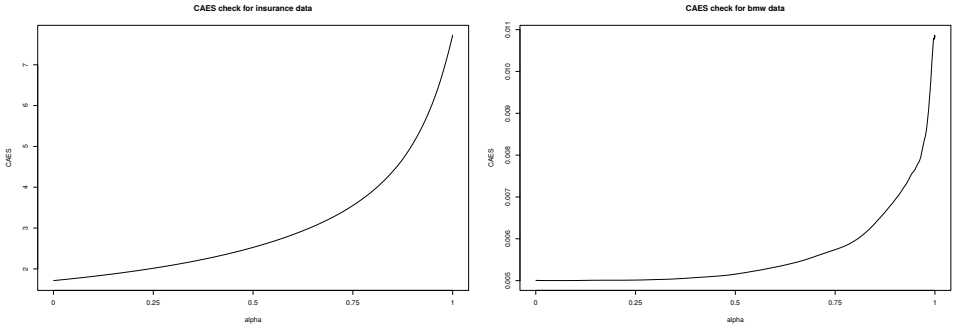
Figure 2.10: Empirical Concentration Profiles with 95% and 99% bootstrapped confidence intervals (1000 samples). For the Danish insurance data (left), confidence intervals are built using bootstrapping and the MAD as a robust measure of variability, while for the BMW data (right) we use block-bootstrapping with non-overlapping block of different sizes (size 200 in the figure) and the standard deviation.

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 $\sigma \to 1$. 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 Figure 2.6b. The results are shown in Figure 2.11b, 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 Figure 2.10b, 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.

Table 2.2 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.



(a) Danish insurance claims. (b) BMW losses.

Figure 2.11: Empirical Concentration Adjusted ES_α for the Danish insurance claims and the BMW share daily losses.

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.

	Danish Insurance losses					BMW log-losses				
α	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

Table 2.2: 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.

To conclude, in Figure 2.12 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 according to our definition of risk.

2.7.3. IDENTIFYING THRESHOLDS IN EXTREME VALUE THEORY

In the context of extreme value theory [30, 32, 43], a significant issue is the 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 [9]. 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 [19]. 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

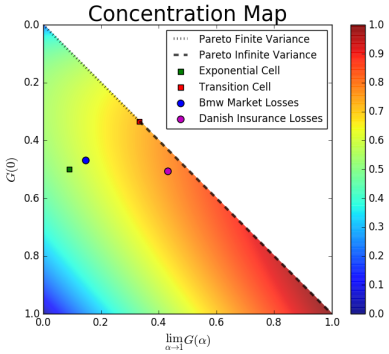
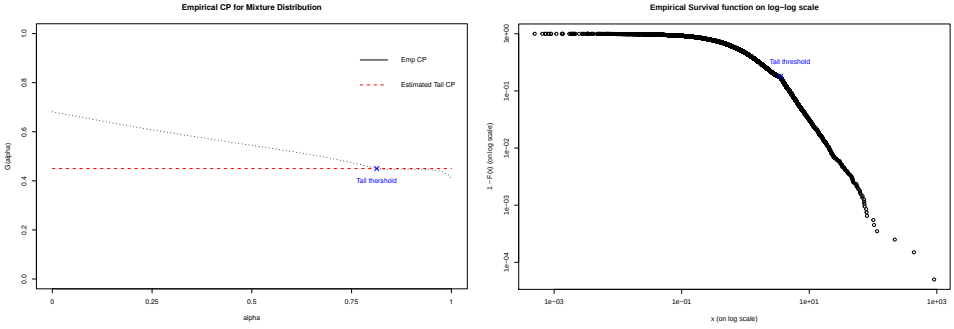


Figure 2.12: Risk evaluation of the Danish and BMW data sets via the Concentration Map. Cobb-Douglas parameters: $a = 0.3$, $b = 0.7$.

should be approximately constant.



(a) Empirical Concentration Profile of a sample(b) Zipf plot (log-log plot of the survival function) from the mixture distribution in Equation (2.18).for the same data.

Figure 2.13: 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.

An example performed on simulated data will clarify this. In Figure 2.13a, the empirical CP of a sample (1000 points) from the following mixture is plotted:

$$Y_{mix} = \beta Y + (1-\beta)Z, \tag{2.18}$$

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 Figure 2.13a).

To assess the quality of our findings we can compare our methodology with other techniques used in EVT to detect the tail threshold (see [9, 19] for a review). In Figure

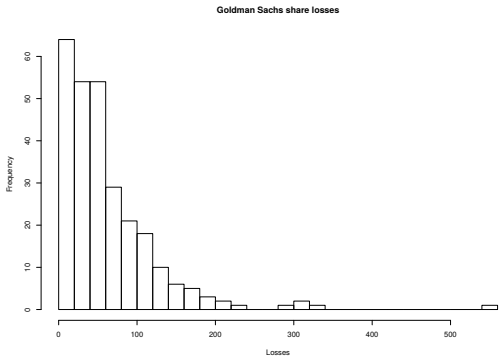


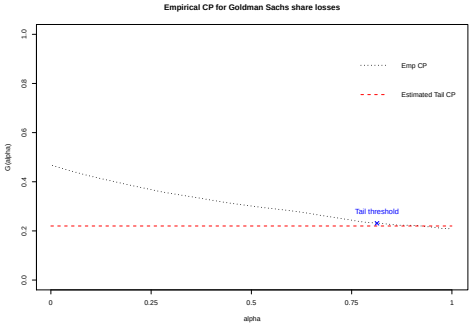
Figure 2.14: Histogram of losses for a portfolio on Goldman Sachs shares (272 observation in the period 2010-2013).

2.13b, we report the empirical survival function $S(x) = 1 - F(x)$ plotted in a log-log scale [19]. 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 [19].

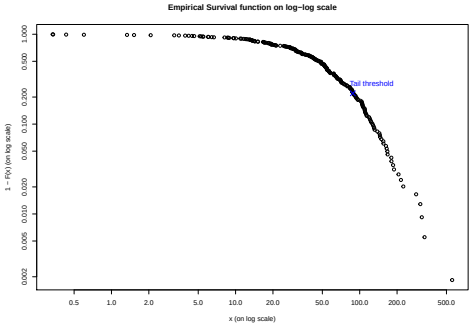
To provide a further real-world example, let us consider 272 non-equally spaced market losses in a portfolio related to Goldman Sachs shares in the period 2010-2013. The histogram of the loss distribution is shown in Figure 2.14, and its Concentration Profile is given in Figure 2.15a. Once again, in order to identify the Paretian threshold, we look for the point at which the concentration profile starts exhibiting an approximately horizontal behavior. In Figure 2.15a, 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 Figure 2.16a 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.

2.8. CONCLUSIONS

In this chapter 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

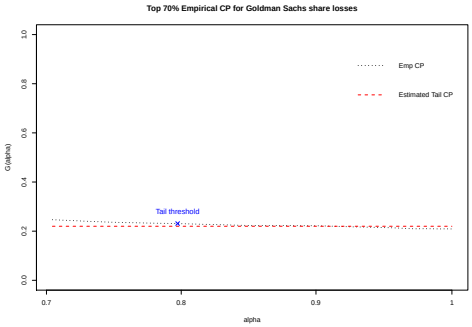


(a) Empirical Concentration Profile.

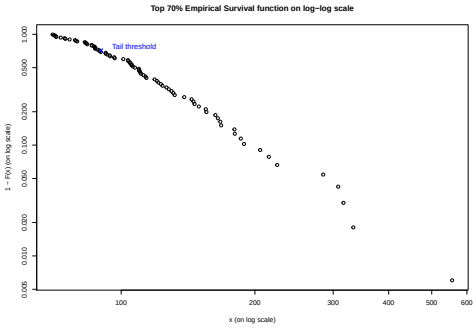


(b) Log-log plot of the survival function.

Figure 2.15: 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 Paretianity.



(a) Empirical Concentration Profile.



(b) Log-log plot of the survival function.

Figure 2.16: Zoomed-in versions of both the empirical Concentration Profile and the Zipf plot of Figure 2.15 to better visualize the right tail. The blue crosses indicate the suggested thresholds for Paretianity.

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 [44]. 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.

APPENDIX

A: GINI INDICES AND LORENZ CURVES OF SOME NOTABLE DISTRIBUTIONS

Table 2.3 contains some distributions often used for modeling losses [2, 10]. We collect them here, as they are also useful for our discussion.

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	$F(y) = \Phi(\frac{\ln(y)-\mu}{\sigma})$	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\ln(y)-\mu)^2}{\sigma^2}}$	$y \in \mathbb{R}^+$	$\sigma \in \mathbb{R}^+$
Exponential	$F(y) = 1 - e^{-\frac{y}{\lambda}}$	$f(y) = \lambda e^{-y\lambda}$	$y \in \mathbb{R}^+$	None
Pareto	$F(y) = 1 - (\frac{y}{y_0})^{-\rho}$	$f(y) = \frac{\rho y_0^\rho}{x^{\rho-1}} 1_{\{y \geq y_0\}}$	$y \in [y_0, +\infty)$	$\rho \in \mathbb{R}^+$

Table 2.3: C.d.f., p.d.f., support and shape parameter of some of the most used distributions in risk management.

For the distributions in Table 2.3, we collect in Table 2.4 the corresponding closed form formulas for their Lorenz curves and Gini indices.

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	$\frac{1}{2\rho-1}$	$L(x) = 1 - (1-x)^{\frac{\rho-1}{\rho}}$

Table 2.4: Gini index and Lorenz curve for the same distributions of Table 2.3 .

B: SAME GINI INDEX BUT DIFFERENT LORENZ CURVES: AN EXAMPLE

In Figure 2.17 two different Lorenz curves are shown. Their functional forms are

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

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 Figure 2.17 occupy the same area. This simple example shows that the Gini index alone

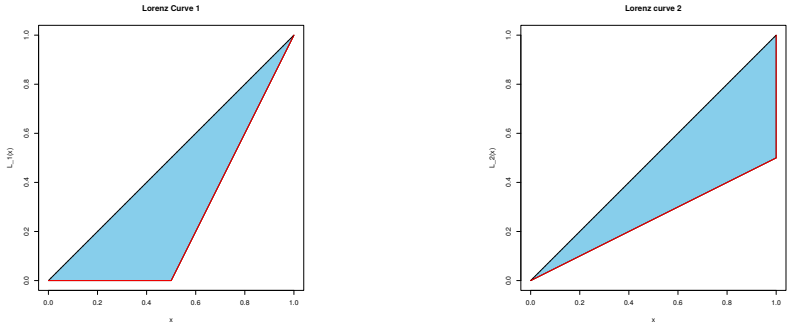


Figure 2.17: In the first graph Lorenz curve $L_1(x)$ is shown; in the second $L_2(x)$. The area shaded between L_{pe} and the two Lorenz curves is the same, giving the same Gini index, however the asymmetry is clearly different.

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 Figure 2.18.

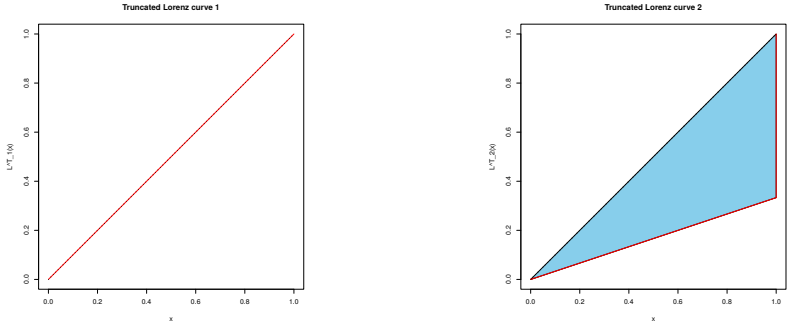


Figure 2.18: The rescaled versions of the Lorenz curves in Figure 2.17, where the upper 50% of the domain is considered. The shaded area is absent in the left side meaning that the Gini index of $L_1^{0.5}(x)$ is zero.

Using Equation (2.5) 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.

C: DERIVATION FOR TRUNCATED LORENZ CURVE

In order to derive the Lorenz curve for truncated random variables, i.e. Equation (2.5), recall that the p.d.f. of a random variable X left-truncated on $[u,+\infty)$ is given by:

$$f_u(x) = \frac{f(x)}{1-F(u)}. \tag{2.19}$$

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.

$$M(t) := \frac{\int_u^t \frac{xf(x)}{1-F(u)}\mathrm{d}x}{\int_u^{+\infty} \frac{xf(x)}{1-F(u)}\mathrm{d}x}, \tag{2.20}$$

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

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

The Lorenz curve is related to the Cumulative Mean Function via the change of variable $t = F^{-1}(x)$. Applying this to Equation (2.20), we get the Lorenz curve associated to the truncated random variable X_u on $[u,+\infty)$,

$$L^*_\alpha(x) = \frac{\int_\alpha^x F^{-1}(u)\mathrm{d}u}{\int_\alpha^1 F^{-1}(u)\mathrm{d}u}, \tag{2.21}$$

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 Equation (2.21), so that

$$L_\alpha(z) = L^*_\alpha(z(1-\alpha)+\alpha). \tag{2.22}$$

Notice that Equation (2.22) can be re-written as

$$L_\alpha(z) = L^*_\alpha(z(1-\alpha)+\alpha) = \frac{\int_\alpha^{z(1-\alpha)+\alpha} F^{-1}(u)\mathrm{d}u}{\int_\alpha^1 F^{-1}(u)\mathrm{d}u}. \tag{2.23}$$

Recalling that the Lorenz curve associated with a random variable with domain in $\mathbb{R}^+$ is given by Equation (2.1), we can adjust the right-hand side of Equation (2.23) so that

$$L_\alpha(z) = \frac{\int_\alpha^{z(1-\alpha)+\alpha} F^{-1}(u)\mathrm{d}u}{\int_\alpha^1 F^{-1}(u)\mathrm{d}u} = \frac{\int_0^{z(1-\alpha)+\alpha} F^{-1}(u)\mathrm{d}u - \int_0^\alpha F^{-1}(u)\mathrm{d}u}{\int_0^1 F^{-1}(u)\mathrm{d}u - \int_0^\alpha F^{-1}(u)\mathrm{d}u}. \tag{2.24}$$

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

$$L_\alpha(z) = \frac{L(z(1-\alpha)+\alpha) - L(\alpha)}{1-L(\alpha)}. \tag{2.25}$$

D: PROOF OF THEOREM 2.2

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

$$F^{-1}(v) = \frac{(1-v)^{-\xi}-1}{\xi}, \qquad 0 \le v \le 1. \tag{2.26}$$

From this, using Equation (2.1), we get

$$L(x) = \frac{\int_0^x \frac{(1-v)^{-\xi}-1}{\xi} \mathrm{d}v}{\int_0^1 \frac{(1-v)^{-\xi}-1}{\xi} \mathrm{d}v}. \tag{2.27}$$

Solving (2.27), we obtain

$$L(x) = \frac{1-(1-x)^{(1-\xi)}-x+\xi x}{\xi}, \tag{2.28}$$

which is the closed form solution for the Lorenz curve of the *GPD*.
We now obtain the truncated Lorenz curve by evaluating (2.28) in $\alpha+(1-\alpha)x$, that is

$$L_\alpha(x) = \frac{-(1-\alpha)^\xi((\alpha-1)(x-1))^{-\xi}+x(1-\alpha)^\xi\left(((\alpha-1)(x-1))^{-\xi}+\xi-1\right)+1}{\xi(1-\alpha)^\xi-(1-\alpha)^\xi+1}. \tag{2.29}$$

Using Equations (2.7) and (2.29), we observe

$$G(\alpha) = 1-2\int_0^1 \frac{-(1-\alpha)^\xi((\alpha-1)(x-1))^{-\xi}+x(1-\alpha)^\xi\left(((\alpha-1)(x-1))^{-\xi}+\xi-1\right)+1}{\xi(1-\alpha)^\xi-(1-\alpha)^\xi+1}\mathrm{d}x,$$

from which we derive

$$G(\alpha) = \frac{\xi}{2-(1-\alpha)^\xi(-2+\xi)(-1+\xi)-\xi}. \tag{2.30}$$

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

$$\frac{dG(\alpha)}{d\alpha} = \frac{1}{2(\alpha-1)(\log(1-\alpha)-1)^2}, \tag{2.31}$$

which is always negative.

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

$$\frac{dG(\alpha)}{d\alpha} = -\frac{(\xi-1)\xi^2(1-\alpha)^{\xi-1}}{(\xi-2)\left((\xi-1)(1-\alpha)^\xi+1\right)^2}, \tag{2.32}$$

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

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 (2.7) to obtain a workable form for the CP.

Recalling Table 2.4, 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 (2.5), so that after some algebra we get

$$L_\alpha(x) = \frac{-x + \log(1-\alpha) + (-1+x)\log((-1+x)(-1+\alpha))}{-1+\log(1-\alpha)}. \tag{2.33}$$

We can now exploit Equation (2.33) in formula (2.7), and we get

$$G(\alpha) = 1 - 2\int_0^1 \frac{-x + \log(1-\alpha) + (-1+x)\log((-1+x)(-1+\alpha))}{-1+\log(1-\alpha)}\mathrm{d}x. \tag{2.34}$$

The integral in (2.34) can be solved directly, thus getting the formula provided in Table 2.1.

REFERENCES

[1] A. Fontanari, P. Cirillo, and C. W. Oosterlee, *From concentration profiles to concentration maps. new tools for the study of loss distributions,* Insurance: Mathematics and Economics **78**, 13 (2018).

[2] J. Hull, *Risk management and financial institutions,+ Web Site*, Vol. 733 (John Wiley & Sons, 2012).

[3] A. Sironi and A. Resti, *Risk management and shareholders' value in banking: from risk measurement models to capital allocation policies*, Vol. 417 (John Wiley & Sons, 2007).

[4] V. Chavez-Demoulin, P. Embrechts, and M. Hofert, *An extreme value approach for modeling operational risk losses depending on covariates,* Journal of Risk and Insurance **83**, 735 (2016).

[5] M. Moscadelli, *The modelling of operational risk: experience with the analysis of the data collected by the basel committee,* Technical Report 517, Bank of Italy (2004).

[6] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition* (Princeton university press, 2015).

[7] B. for International Settlements, *Basel committee on banking supervision (bcbs),* International convergence of capital measurement and capital standards:: a revised-framework (2006).

[8] B. Committee *et al.*, *Basel iii: A global regulatory framework for more resilient banks and banking systems,* Basel Committee on Banking Supervision, Basel (2010).

[9] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events: for insurance and finance*, Vol. 33 (Springer Science & Business Media, 2013).

[10] C. Kleiber and S. Kotz, *Statistical size distributions in economics and actuarial sciences.*, Vol. 470 (John Wiley & Sons, 2003).

[11] I. I. Eliazar and I. M. Sokolov, *Measuring statistical evenness: A panoramic overview,* Physica A: Statistical Mechanics and its Applications **391**, 1323 (2012).

[12] S. Yitzhaki and E. Schechtman, *The Gini methodology: A primer on a statistical methodology*, Vol. 272 (Springer Science & Business Media, 2012).

[13] M. O. Lorenz, *Methods of measuring the concentration of wealth,* Publications of the American statistical association **9**, 209 (1905).

[14] C. Gini, *Variabilità e mutabilità (1912), reprinted in: Variabilità e mutabilità, e,* Pizetti and T. Salvemini, Memorie di Metodologica Statistica, Libreria Eredi Virgilio Veschi (1955).

[15] H. Shalit and S. Yitzhaki, *Mean-gini, portfolio theory, and the pricing of risky assets,* The journal of Finance **39**, 1449 (1984).

[16] H. Shalit and S. Yitzhaki, *The mean-gini efficient portfolio frontier,* Journal of Financial Research **28**, 59 (2005).

[17] W. Edwards, *Utility theories: Measurements and applications*, Vol. 3 (Springer Science & Business Media, 2013).

[18] N. U. Nair, P. Sankaran, and B. Vineshkumar, *Characterization of distributions by properties of truncated gini index and mean difference,* Metron **70**, 173 (2012).

[19] P. Cirillo, *Are your data really pareto distributed?* Physica A: Statistical Mechanics and its Applications **392**, 5947 (2013).

[20] F. Fiordelisi, M.-G. Soana, and P. Schwizer, *Reputational losses and operational risk in banking,* The European Journal of Finance **20**, 105 (2014).

[21] C. Blackorby and D. Donaldson, *Ethical indices for the measurement of poverty,* Econometrica (pre-1986) **48**, 1053 (1980).

[22] A. Cambini and L. Martein, *Generalized convexity and optimization: Theory and applications*, Vol. 616 (Springer Science & Business Media, 2008).

[23] M. Frittelli and M. Maggis, *Dual representation of quasi-convex conditional maps,* SIAM Journal on Financial Mathematics **2**, 357 (2011).

[24] D. Tian and L. Jiang, *Quasiconvex risk statistics with scenario analysis,* Mathematics and Financial Economics **9**, 111 (2015).

[25] D. Chotikapanich, *Modeling income distributions and Lorenz curves*, Vol. 5 (Springer Science & Business Media, 2008).

[26] P. Cirillo and N. N. Taleb, *Expected shortfall estimation for apparently infinite-mean models of operational risk,* Quantitative Finance **16**, 1485 (2016).

[27] V. Chavez-Demoulin, P. Embrechts, and S. Sardy, *Extreme-quantile tracking for financial time series,* Journal of Econometrics **181**, 44 (2014).

[28] Z. Du and J. C. Escanciano, *Backtesting expected shortfall: accounting for tail risk,* Management Science **63**, 940 (2016).

[29] L. De Haan and A. Ferreira, *Extreme value theory: an introduction* (Springer Science & Business Media, 2007).

[30] J. Pickands III *et al.*, *Statistical inference using extreme order statistics,* the Annals of Statistics **3**, 119 (1975).

[31] J. Beirlant and Y. Goegebeur, *Local polynomial maximum likelihood estimation for pareto-type distributions,* Journal of Multivariate Analysis **89**, 97 (2004).

[32] B. Gnedenko, *Sur la distribution limite du terme maximum d'une serie aleatoire,* Annals of mathematics , 423 (1943).

[33] E. Barucci, C. Fontana, *et al.*, *Financial markets theory* (Springer, 2003).

[34] J. L. Gastwirth, *The estimation of the lorenz curve and gini index,* The review of economics and statistics , 306 (1972).

[35] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap* (CRC press, 1994).

[36] B. Efron, *The jackknife, the bootstrap, and other resampling plans*, Vol. 38 (Siam, 1982).

[37] H. Konno and H. Yamazaki, *Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market,* Management science **37**, 519 (1991).

[38] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust statistics: theory and methods (with R)* (John Wiley & Sons, 2019).

[39] A. Fontanari, N. N. Taleb, and P. Cirillo, *Gini estimation under infinite variance,* Physica A: Statistical Mechanics and its Applications **502**, 256 (2018).

[40] W. J. Hall and J. A. Wellner, *Confidence bands for a survival curve from censored data,* Biometrika **67**, 133 (1980).

[41] A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application*, Vol. 1 (Cambridge university press, 1997).

[42] D. N. Politis and J. P. Romano, *The stationary bootstrap,* Journal of the American Statistical association **89**, 1303 (1994).

[43] R. A. Fisher and L. H. C. Tippett, *Limiting forms of the frequency distribution of the largest or smallest member of a sample,* in *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 24 (Cambridge University Press, 1928) pp. 180–190.

[44] G. Pietra, *Delle relazioni tra gli indici di variabilit a. part ii,* Atti del reale Istituto Veneto di Scienze, Lettere ed Arti, LXXIV (1915).

3

GINI ESTIMATION UNDER INFINITE VARIANCE

We study the problems related to the estimation of the Gini index in presence of a fat-tailed data generating process, i.e. one in the stable distribution class with finite mean but infinite variance (i.e. with tail index $\alpha \in (1,2)$). We show that, in such a case, the Gini coefficient cannot be reliably estimated using conventional nonparametric methods, because of a downward bias that emerges under fat tails. This has important implications for the ongoing discussion about economic inequality.

We start by discussing how the nonparametric estimator of the Gini index undergoes a phase transition in the symmetry structure of its asymptotic distribution, as the data distribution shifts from the domain of attraction of a light-tailed distribution to that of a fat-tailed one, especially in the case of infinite variance. We also show how the nonparametric Gini index bias increases with lower values of α. We then prove that maximum likelihood estimation outperforms nonparametric methods, requiring a much smaller sample size to reach efficiency. Finally, for fat-tailed data, we provide a simple correction mechanism to the small sample bias of the nonparametric estimator based on the distance between the mode and the mean of its asymptotic distribution.

Keywords: *Gini index; inequality measure; size distribution; extremes; α-stable distribution.*

Parts of this chapter have been published in Physica A: Statistical Mechanics and its Applications **502**, 256–269 (2018) [1].

3.1. INTRODUCTION

Wealth inequality studies represent a field of economics, statistics and econophysics exposed to fat-tailed data generating processes, often with infinite variance [2, 3]. This is not at all surprising if we recall that the prototype of fat-tailed distributions, the Pareto, has been proposed for the first time to model household incomes [4]. However, the fat-tailedness of data can be problematic in the context of wealth studies, as the property of efficiency (and, partially, consistency) does not necessarily hold for many estimators of inequality and concentration [3, 5].

The scope of this work is to show how fat tails affect the estimation of one of the most celebrated measures of economic inequality, the Gini index [6–8], often used (and abused) in the econophysical and economic literature as the main tool for describing the distribution and the concentration of wealth around the world [2, 9, 10].

The literature concerning the estimation of the Gini index is wide and comprehensive (e.g. [6, 8] for a review), however, strangely enough, almost no attention has been paid to its behavior in presence of fat tails, and this is curious if we consider that: 1) fat tails are ubiquitous in the empirical distributions of income and wealth [3, 10], and 2) the Gini index itself can be seen as a measure of variability and fat-tailedness [11–14].

The usual method for the estimation of the Gini index is nonparametric: one computes the index from the empirical distribution of the available data using Equation (3.5) below. However, such estimator is known to be suffer of finite sample bias [15] whose direction depends on the underlying true unknown distribution. In this cheaper by studying the asymptotic distribution of the non-parametric estimator for the Gini index under infinite variance assumptions we try to cast some light on the finite sampling distribution of such estimator and its pre-asymptotic behaviour. Additionally, we show how the maximum likelihood approach, despite the risk of model misspecification, needs much fewer observations to reach efficiency when compared to a nonparametric one.[1]

By fat-tailed data we indicate those data generated by a non-negative random variable X with cumulative distribution function (c.d.f.) $F(x)$, which is regularly-varying of order α [17], that is, for $\bar{F}(x) := 1 - F(x)$, one has

$$\lim_{x\to\infty} x^{\alpha}\bar{F}(x) \sim L(x), \tag{3.1}$$

where $L(x)$ is a slowly-varying function such that $\lim_{x\to\infty} \frac{L(cx)}{L(x)} = 1$ with $c > 0$, and where $\alpha > 0$ is called the tail exponent.

Regularly-varying distributions define a large class of random variables whose properties have been extensively studied in the context of extreme value theory [5, 18], when dealing with the probabilistic behavior of maxima and minima. As pointed out in [19], regularly-varying and fat-tailed are indeed synonyms. It is known that, if $X_1, ..., X_n$ are i.i.d. random variables with common c.d.f. $F(x)$ in the regularly-varying class, as defined in Equation (3.1), then their data generating process falls into the maximum domain of attraction of a Fréchet distribution with parameter ρ, in symbols $X \in MDA(\Phi(\rho))$[18].

[1] A similar bias also affects the nonparametric measurement of quantile contributions, i.e. those of the type "the top 1% owns x% of the total wealth" [16]. This chapter extends the problem to the more widespread Gini coefficient, and goes deeper by making links with the limit Theorems.

This means that, for the partial maximum $M_n = \max(X_1, ..., X_n)$, one has

$$P\left(a_n^{-1}(M_n - b_n) \leq x\right) \xrightarrow{d} \Phi_\rho(x) = e^{-x^{-\rho}}, \qquad \rho > 0, \tag{3.2}$$

with $a_n > 0$ and $b_n \in \mathbb{R}$ two normalizing constants. Clearly, the connection between the regularly-varying coefficient α and the Fréchet distribution parameter ρ is given by: $\alpha = \frac{1}{\rho}$ [5].

The Fréchet distribution is one of the limiting distributions for maxima in extreme value theory, together with the Gumbel and the Weibull; it represents the fat-tailed and unbounded limiting case [18]. The relationship between regularly-varying random variables and the Fréchet class thus allows us to deal with a very large family of random variables (and empirical data), also showing how the Gini index is highly influenced by maxima, i.e. extreme wealth, as intuition clearly suggests [3, 14], especially under infinite variance. Again, this recommends some caution when discussing economic inequality under fat tails.

It is worth remembering that the existence (finiteness) of the moments for a fat-tailed random variable X depends on the tail exponent α, in fact

$$\begin{aligned} &\mathbb{E}(X^\delta) < \infty \text{ if } \delta \leq \alpha, \\ &\mathbb{E}(X^\delta) = \infty \text{ if } \delta > \alpha. \end{aligned} \tag{3.3}$$

In this chapter, we restrict our focus on data generating processes with finite mean and infinite variance, therefore, according to Equation (3.3), on the class of regularly-varying distributions with tail index $\alpha \in (1,2)$.

Table 3.1 and Figure 3.1 present numerically and graphically our story, already suggesting its conclusion, on the basis of artificial observations sampled from a Pareto distribution (Equation (3.14) below) with tail parameter α equal to 1.1.

Table 3.1 compares the nonparametric Gini index of Equation (3.5) with the maximum likelihood (ML) tail-based one of Section 3.3. For the different sample sizes in Table 3.1, we have generated 10^8 samples, averaging the estimators via Monte Carlo. As the first column shows, the convergence of the nonparametric estimator to the true Gini value ($g = 0.8333$) is extremely slow and monotonically increasing; this suggests an issue not only in the tail structure of the distribution of the nonparametric estimator but also in its symmetry.

Figure 3.1 provides some evidence that the limiting distribution of the nonparametric Gini index loses its properties of normality and symmetry [20], shifting towards a skewed and fatter-tailed limit, when data are characterized by an infinite variance. As we prove in Section 3.2, when the data generating process is in the domain of attraction of a fat-tailed distribution, the asymptotic distribution of the Gini index becomes a skewed-to-the-right α-stable law.

This change of behavior is responsible for the steep increase in the sampling variability of the non-parametric Gini estimator as well as to its slower speed on convergence being Berry–Esseen type of results losing their validity [5]. Additionally, because of the right-skew of the asymptotic distribution it may be reasonable to assume that it is more likely to observe lower-than-the-average values for the non-parametric estimator in its pre-asymptotic behaviour as empirical evidence of Table 3.1 shows.

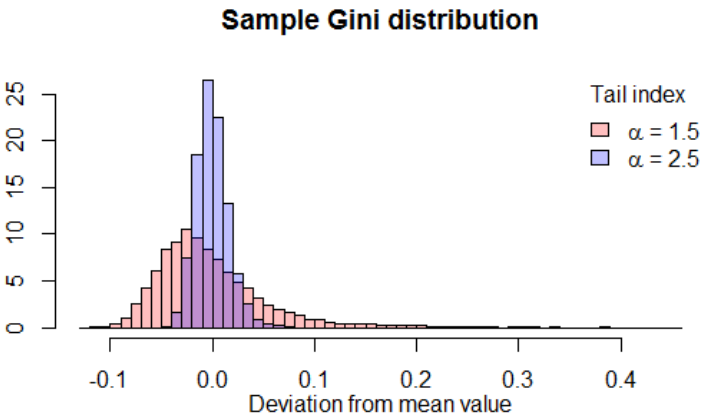


Figure 3.1: Histograms for the Gini nonparametric estimators for two Paretian (type I) distributions with different tail indices, with finite and infinite variance (plots have been centered to ease comparison). Sample size: 10^3. Number of samples: 10^2 for each distribution.

The knowledge of the new limit allows us to propose a correction for the nonparametric estimator, improving its quality, and thus reducing the risk of badly estimating wealth inequality, with all the possible consequences in terms of economic and social policies [3, 10, 21].

Table 3.1: Comparison of the Nonparametric (NonPar) and the Maximum Likelihood (ML) Gini estimators, using Paretian data with tail $\alpha = 1.1$ (finite mean, infinite variance) and different sample sizes. Number of Monte Carlo simulations: 10^8.

n	Nonpar		ML		Error Ratio
(number of obs.)	*Mean*	*Bias*	*Mean*	*Bias*	
10^3	*0.711*	*-0.122*	*0.8333*	*0*	1.4
10^4	*0.750*	*-0.083*	*0.8333*	*0*	3
10^5	*0.775*	*-0.058*	*0.8333*	*0*	6.6
10^6	0.790	-0.043	0.8333	0	156
10^7	0.802	-0.031	0.8333	0	10^5+

The chapter is organized as follows. In Section 3.2 we derive the asymptotic distribution of the sample Gini index when data possess an infinite variance. In Section 3.3 we deal with the maximum likelihood estimator, while in Section 3.4 we provide an illustration with Paretian observations. In Section 3.5 we propose a simple correction based on the mode-mean distance of the asymptotic distribution of the nonparametric estimator, to take care of its small-sample bias. Section 3.6 closes the chapter. A technical Appendix contains the longer proofs of the main results in the work.

3.2. ASYMPTOTICS OF THE NONPARAMETRIC ESTIMATOR UNDER INFINITE VARIANCE

We derive the asymptotic distribution for the nonparametric estimator of the Gini index when the data generating process is fat-tailed with finite mean but infinite variance.

The so-called stochastic representation of the Gini g is

$$g = \frac{1}{2}\frac{\mathbb{E}\left(|X' - X"|\right)}{\mu} \in [0,1], \tag{3.4}$$

where X' and $X"$ are i.i.d. copies of a random variable X with c.d.f. $F(x) \in [c,\infty)$, $c > 0$, and with finite mean $\mathbb{E}(X) = \mu$. The quantity $\mathbb{E}\left(|X' - X"|\right)$ is known as the "Gini Mean Difference" (GMD) [8]. For later convenience we also define $g = \frac{\theta}{\mu}$ with $\theta = \frac{\mathbb{E}(|X'-X"|)}{2}$.

The Gini index of a random variable X is thus the mean expected deviation between any two independent realizations of X, scaled by twice the mean [22].

The most common nonparametric estimator of the Gini index for a sample $X_1, ..., X_n$ is defined as

$$G^{NP}(X_n) = \frac{\sum_{1\le i<j\le n}|X_i - X_j|}{(n-1)\sum_{i=1}^n X_i}, \tag{3.5}$$

which can also be expressed as

$$G^{NP}(X_n) = \frac{\sum_{i=1}^n (2(\frac{i-1}{n-1}) - 1)X_{(i)}}{\sum_{i=1}^n X_{(i)}} = \frac{\frac{1}{n}\sum_{i=1}^n Z_{(i)}}{\frac{1}{n}\sum_{i=1}^n X_i}, \tag{3.6}$$

where $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are the ordered statistics of $X_1, ..., X_n$, such that: $X_{(1)} < X_{(2)} < ... < X_{(n)}$ and $Z_{(i)} = 2\left(\frac{i-1}{n-1} - 1\right)X_{(i)}$. The asymptotic normality of the estimator in Equation (3.6) under the hypothesis of finite variance for the data generating process is known [3, 8, 23]. The result directly follows from the properties of the U-statistics and the L-estimators involved in Equation (3.6)

A standard methodology to prove the limiting distribution of the estimator in Equation (3.6), and more in general of a linear combination of order statistics, is to show that in the limit, for $n \to \infty$, the sequence of order statistics can be approximated by a sequence of i.i.d random variables [24, 25]. However, this usually requires some sort of L^2 integrability of the data generating process, something we are not assuming here.

Lemma 3.1, whose proof is to be found in the Appendix at the end of the chapter, shows how to deal with the case of sequences of order statistics generated by fat-tailed L^1-only integrable random variables.

Lemma 3.1. *Consider the following sequence* $R_n = \frac{1}{n}\sum_{i=1}^n (\frac{i}{n} - U_{(i)})F^{-1}(U_{(i)})$ *where* $U_{(i)}$ *are the order statistics of a uniformly distributed i.i.d random sample. Assume that* $F^{-1}(U) \in L^1$. *Then the following results hold:*

$$R_n \xrightarrow{L^1} 0, \tag{3.7}$$

and

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} R_n \xrightarrow{L^1} 0, \tag{3.8}$$

with $\alpha \in (1,2)$ *and* $L_0(n)$ *a slowly-varying function.*

3.2.1. A QUICK RECAP ON α-STABLE RANDOM VARIABLES

We here introduce some notation for α-stable distributions, as we need them to study the asymptotic limit of the Gini index.

A random variable X follows an α-stable distribution, in symbols $X \sim S(\alpha, \beta, \gamma, \delta)$, if its characteristic function is

$$\mathbb{E}(e^{itX}) = \begin{cases} e^{-\gamma^\alpha |t|^\alpha (1 - i\beta \operatorname{sign}(t)) \tan(\frac{\pi\alpha}{2}) + i\delta t} & \alpha \neq 1 \\ e^{-\gamma|t|(1 + i\beta \frac{2}{\pi} \operatorname{sign}(t)) \ln|t| + i\delta t} & \alpha = 1 \end{cases},$$

where $\alpha \in (0,2)$ governs the tail, $\beta \in [-1,1]$ is the skewness, $\gamma \in \mathbb{R}^+$ is the scale parameter, and $\delta \in \mathbb{R}$ is the location one. This is known as the $S1$ parametrization of α-stable distributions [26, 27].

It is interesting to notice that there is a correspondence between the α parameter of an α-stable random variable, and the α of a regularly-varying random variable as per Equation (3.1): as shown in [20, 26], a regularly-varying random variable of order α is α-stable, with the same tail coefficient. This is why we do not make any distinction in the use of the α here. Since we aim at dealing with distributions characterized by finite mean but infinite variance, we restrict our attention to $\alpha \in (1,2)$, as the two α's coincide.

Recall that, for $\alpha \in (1,2]$, the expected value of an α-stable random variable X is equal to the location parameter δ, i.e. $\mathbb{E}(X) = \delta$. For more details, we refer to [26, 27].

The standardized α-stable random variable is expressed as

$$S_{\alpha,\beta} \sim S(\alpha, \beta, 1, 0). \tag{3.9}$$

α-stable distributions are a subclass of infinitely divisible distributions. Thanks to their closure under convolution, they can be used to describe the limiting behavior of (rescaled) partials sums, $S_n = \sum_{i=1}^n X_i$, in the General Central Limit Theorem (GCLT) setting [20]. For $\alpha = 2$ we obtain the normal distribution as a special case, which is the limit distribution for the classical CLTs, under the hypothesis of finite variance.

In what follows we indicate that a random variable is in the domain of attraction of an α-stable distribution, by writing $X \in DA(S_{\alpha,\beta})$. Just observe that this condition for the limit of partial sums is equivalent to the one given in Equation (3.2) for the limit of partial maxima [5, 20].

3.2.2. THE α-STABLE ASYMPTOTIC LIMIT OF THE GINI INDEX

Consider a sample $X_1, ..., X_n$ of i.i.d. random variables with common continuous c.d.f. $F(x)$ in the regularly-varying class, as defined in Equation (3.1), with tail index $\alpha \in (1,2)$. This corresponds to considering a sample whose data generating process is in the domain of attraction of a Fréchet distribution with $\rho \in (\frac{1}{2}, 1)$, given that $\rho = \frac{1}{\alpha}$.

To study the asymptotic distribution of the Gini index estimator, as presented in Equation (3.6), when the data generating process is characterized by an infinite variance, we can make use of the following two theorems: Theorem 3.2 deals with the limiting distribution of the Gini Mean Difference (the numerator in Equation (3.6)), while Theorem 3.3 extends the result to the complete Gini index. For both theorems the proofs are to be found in the Appendix at the end of the chapter.

Theorem 3.2. *Consider a sequence $(X_i)_{1\leq i\leq n}$ of i.i.d random variables with common distribution $F(x)$ on $[c,+\infty)$ with $c>0$, such that $X\sim F(x)$ is in the domain of attraction of an α-stable random variable, $X\in DA(S_\alpha)$, with $\alpha\in(1,2)$. Then the sample Gini mean difference (GMD) $\frac{\sum_{i=1}^n Z_{(i)}}{n}$ satisfies the following limit in distribution:*

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(\frac{1}{n}\sum_{i=1}^{n} Z_{(i)}-\theta\right)\xrightarrow{d} S_{\alpha,1}, \tag{3.10}$$

where $Z_i=(2F(X_i)-1)X_i$, $\mathbb{E}(Z_i)=\theta$, $L_0(n)$ is a slowly-varying function such that Equation (3.38) holds (see the Appendix), and $S_{\alpha,1}$ is a right-skewed standardized α-stable random variable defined as in Equation (3.9).

Moreover the statistic $\frac{1}{n}\sum_{i=1}^n Z_{(i)}$ is an asymptotically consistent estimator for the GMD, i.e. $\frac{1}{n}\sum_{i=1}^n Z_{(i)}\xrightarrow{P}\theta$.

Please observe that Theorem 3.2 could be restated in terms of the maximum domain of attraction $MDA(\Phi(\rho))$ as defined in Equation (3.2).

Theorem 3.3. *Given the same assumptions of Theorem 3.2, the estimated Gini index $G^{NP}(X_n)=\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_i}$ satisfies the following limit in distribution*

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(G^{NP}(X_n)-\frac{\theta}{\mu}\right)\xrightarrow{d} Q, \tag{3.11}$$

where $\mathbb{E}(Z_i)=\theta$, $\mathbb{E}(X_i)=\mu$, $L_0(n)$ is the same slowly-varying function defined in Theorem 3.2 and Q is a right-skewed α-stable random variable $S(\alpha,1,\frac{1}{\mu},0)$.

Furthermore the statistic $\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_i}$ is an asymptotically consistent estimator for the Gini index, i.e. $\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_i}\xrightarrow{P}\frac{\theta}{\mu}=g$.

In the case of fat tails with $\alpha\in(1,2)$, Theorem 3.3 tells us that the asymptotic distribution of the Gini estimator is always right-skewed notwithstanding the distribution of the underlying data generating process. Therefore heavily fat-tailed data not only induce a fatter-tailed limit for the Gini estimator, but they also change the shape of the limit law, which definitely moves away from the usual symmetric Gaussian. As a consequence, the Gini estimator, whose asymptotic consistency is still guaranteed [24], will approach its true value more slowly, and from below. Some evidence of this was already given in Table 3.1.

3.3. THE MAXIMUM LIKELIHOOD ESTIMATOR

Theorem 3.3 indicates that the usual nonparametric estimator for the Gini index is not the best option when dealing with infinite-variance distributions, due to the skewness and the fatness of its asymptotic limit. A way out is to look for estimators that still preserve their asymptotic normality under fat tails, but this is not possible with nonparametric methods, as they all fall into the α-stable Central Limit Theorem case [5, 20]. The solution is thus to use parametric techniques.

Theorem 3.4 shows how, once a parametric family for the data generating process has been identified, it is possible to estimate the Gini index via MLE. The resulting estimator is not only asymptotically normal but also asymptotically efficient.

In Theorem 3.4 we deal with random variables X whose distribution belongs to the large and flexible exponential family [28], i.e. a set of distributions whose density can be represented as

$$f_\theta(x) = h(x)e^{(\eta(\theta)T(x)-A(\theta))}, \tag{3.12}$$

with $\theta \in \mathbb{R}$, and where $T(x)$, $\eta(\theta)$, $h(x)$, $A(\theta)$ are known functions.

Theorem 3.4. *Let $X \sim F_\theta$ such that F_θ is a distribution belonging to the exponential family of Equation (3.12). Then the Gini index obtained by plugging-in the maximum likelihood estimator of θ, $G^{ML}(X_n)_\theta$, is asymptotically normal and efficient. Namely:*

$$\sqrt{n}(G^{ML}(X_n)_\theta - g_\theta) \xrightarrow{d} N(0, g_\theta'^2 I^{-1}(\theta)), \tag{3.13}$$

where $g'_\theta = \frac{dg_\theta}{d\theta}$ and $I(\theta)$ is the Fisher Information.

Proof. The result follows easily from the asymptotic efficiency of the maximum likelihood estimators of the exponential family, and the invariance principle of MLE. In particular, the validity of the invariance principle for the Gini index is granted by the continuity and the monotonicity of g_θ with respect to θ. The asymptotic variance is then obtained by application of the delta-method [28]. □

3.4. A PARETIAN ILLUSTRATION

We provide an illustration of the obtained results using some artificial fat-tailed data. We choose a Pareto I [4], with density

$$f(x) = \alpha c^\alpha x^{-\alpha-1}, x \geq c. \tag{3.14}$$

It is easy to verify that the corresponding survival function $\bar{F}(x)$ belongs to the regularly-varying class with tail parameter α and slowly-varying function $L(x) = c^\alpha$. We can therefore apply the results of Section 3.2 to obtain the following corollaries.

Corollary 3.1. *Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables with Pareto distribution with tail parameter $\alpha \in (1,2)$. The nonparametric Gini estimator is characterized by the following limit:*

$$D_n^{NP} = G^{NP}(X_n) - g \sim S\left(\alpha, 1, \frac{C_\alpha^{\frac{1}{\alpha}}}{(1+\alpha)^{\frac{1}{\alpha}} n^{\frac{\alpha-1}{\alpha}}} \frac{(\alpha-1)}{\alpha}, 0\right). \tag{3.15}$$

Proof. Without loss of generality we can assume $c = 1$ in Equation (3.14). The results is a mere application of Theorem 3.3, remembering that a Pareto distribution is in the domain of attraction of α-stable random variables with slowly-varying function $L(x) = 1$. The sequence c_n to satisfy Equation (3.38) becomes $c_n = n^{\frac{1}{\alpha}} C_\alpha^{-\frac{1}{\alpha}}$, therefore we have $L_0(n) = C_\alpha^{-\frac{1}{\alpha}}$, which is independent of n. Additionally the mean of the distribution is also a function of α, that is $\mu = \frac{\alpha}{\alpha-1}$. □

Corollary 3.2. *Let the sample $X_1, ..., X_n$ be distributed as in Corollary 3.1, let G_θ^{ML} be the maximum likelihood estimator for the Gini index as defined in Theorem 3.4. Then the MLE Gini estimator, rescaled by its true mean g, has the following limit:*

$$D_n^{ML} = G_\alpha^{ML}(X_n) - g \sim N\left(0, \frac{4\alpha^2}{n(2\alpha-1)^4}\right), \tag{3.16}$$

where N indicates a Gaussian.

Proof. The functional form of the maximum likelihood estimator for the Gini index is known to be $G_\theta^{ML} = \frac{1}{2\alpha^{ML}-1}$ [3]. The result then follows from the fact that the Pareto distribution (with known minimum value x_m) belongs to an exponential family and therefore satisfies the regularity conditions necessary for the asymptotic normality and efficiency of the maximum likelihood estimator. Also notice that the Fisher information for a Pareto distribution is $\frac{1}{\alpha^2}$. □

Now that we have worked out both asymptotic distributions, we can compare the quality of the convergence for both the MLE and the nonparametric case when dealing with Paretian data, which we use as the prototype for the more general class of fat-tailed observations.

In particular, we can approximate the distribution of the deviations of the estimator from the true value g of the Gini index for finite sample sizes, by using Equations (3.15) and (3.16).

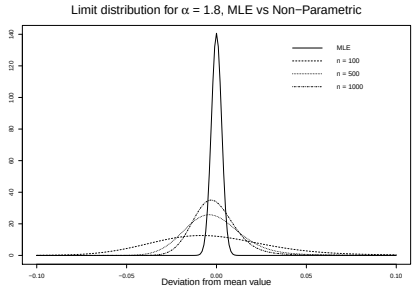
Figure 3.2 shows how the deviations around the mean of the two different types of estimators are distributed and how these distributions change as the number of observations increases. Additionally, in the Appendix we provide evidence of how the finite sample approximation given in Corollary 3.1 well fits the empirical distribution.

Fixing the number of observation in the MLE case and letting them vary in the nonparametric one we show how the impact of different tail indexes is on the consistency of the estimator. It is worth noticing that, as the tail index decreases towards 1 (the threshold value for a infinite mean), the mode of the distribution of the nonparametric estimator moves farther away from the mean of the distribution (centered on 0 by definition, given that we are dealing with deviations from the mean). Such a phenomenon is not present in the MLE case, thanks to the the normality of the limit for every value of the tail parameter.

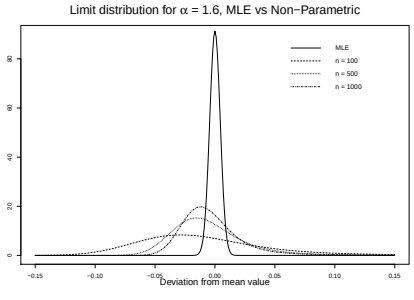
We can make our argument more rigorous by assessing the number of observations $\tilde{n}$ needed for the nonparametric estimator to be as good as the MLE one, under different tails. Let's consider the likelihood-ratio-type function

$$r(c,n) = \frac{P_S(|D_n^{NP}| > c)}{P_N(|D_{100}^{ML}| > c)}, \tag{3.17}$$

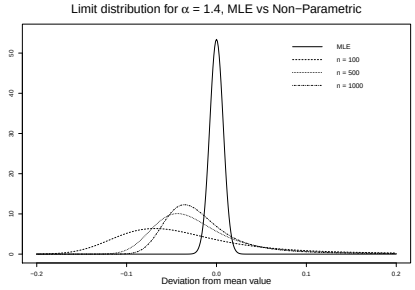
where $P_S(|D_n^{NP}| > c)$ and $P_N(|D_{100}^{ML}| > c)$ are the tail probabilities (α-stable and Gaussian respectively) of the centered estimators in the nonparametric and in the MLE cases to exceed the thresholds $\pm c$, as per Equations (3.16) and (3.15). In the nonparametric case the number of observations n is allowed to change, while in the MLE case it is fixed to 100. We then wish to find the value $\tilde{n}$ such that $r(c,\tilde{n}) = 1$ for fixed c.



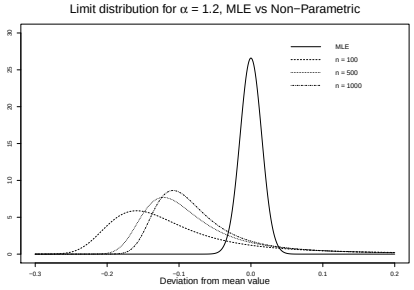
(a) $\alpha = 1.8$



(b) $\alpha = 1.6$



(c) $\alpha = 1.4$



(d) $\alpha = 1.2$

Figure 3.2: Comparisons between the maximum likelihood and the nonparametric asymptotic distributions for different values of the tail index α. The number of observations for MLE is fixed to $n = 100$. Note that, even if all distributions have mean zero, the mode of the distributions of the nonparametric estimator is different from zero, because of the skewness.

Table 3.2 displays the results for different thresholds c and tail parameters α. In particular, we can see how the MLE estimator outperforms the nonparametric one, which requires a much larger number of observations to obtain the same tail probability of the MLE with n fixed to 100. For example, we need at least 80×10^6 observations for the nonparametric estimator to obtain the same probability of exceeding the ± 0.02 threshold of the MLE one, when $\alpha = 1.2$.

One interesting thing to notice is that the number of observations needed to match the tail probabilities in Equation (3.17) does not vary uniformly with the threshold. This is expected, since as the threshold goes to infinity or to zero, the tail probabilities are the same for every number of n. Therefore, given the unimodality of the limit distributions, we expect that there will be a threshold maximizing the number of observations needed to match the tail probabilities, while for all the other levels the number of observations will be smaller.

Table 3.2: The number of observations $\tilde{n}$ needed for the nonparametric estimator to match the tail probabilities, for different threshold values c and different values of the tail index α, of the maximum likelihood estimator with fixed $n = 100$.

	Threshold c as per Equation (3.17):			
α	0.005	0.01	0.015	0.02
1.8	27×10^3	12×10^5	12×10^6	63×10^5
1.5	21×10^4	21×10^4	46×10^5	81×10^7
1.2	33×10^8	67×10^7	20×10^7	80×10^6

3.5. SMALL SAMPLE CORRECTION

Theorem 3.3 can be also used to provide a correction for the bias of the nonparametric estimator for small sample sizes. The key idea is to recognize that, for unimodal distributions, most observations come from around the mode. In symmetric distributions the mode and the mean coincide, thus most observations will be close to the mean value as well. For skewed distributions, conversely, this is not the case. In particular, for right-skewed continuous unimodal distributions the mode is lower than the mean. Therefore, given that the asymptotic distribution of the nonparametric Gini index is right-skewed assuming that such of behaviour is maintained in the pre-asymptotic we expect that the observed value of the Gini index estimator will be usually lower than the true one (placed at the mean level).

We can quantify this difference by looking at the distance between the mode and the mean, and once this distance is known, we can correct our Gini estimate by adding it back[2].

Formally, we want to derive a corrected nonparametric estimator $G^C(X_n)$ such that

$$G^C(X_n) = G^{NP}(X_n) + |m(G^{NP}(X_n)) - \mathbb{E}(G^{NP}(X_n))|, \tag{3.18}$$

where $|m(G^{NP}(X_n)) - \mathbb{E}(G^{NP}(X_n))|$ is the distance between the mode m and the mean of the distribution of the nonparametric Gini estimator $G^{NP}(X_n)$.

Performing the type of correction described in Equation (3.18) is equivalent to shifting the distribution of $G^{NP}(X_n)$ in order to place its mode on the true value of the Gini index.

Ideally, we would like to measure this mode-mean distance $|m(G^{NP}(X_n)) - \mathbb{E}(G^{NP}(X_n))|$ on the exact distribution of the Gini index to get the most accurate correction. However, the finite distribution is not always easily derivable and it requires assumptions on the parametric structure of the data generating process (actually, even in this situation, in most cases it is unknown for fat-tailed data [3]). Therefore we propose to use the limiting distribution for the nonparametric Gini obtained in Section 3.2 to approximate the finite sample distribution, and to estimate the mode-mean distance with it. This procedure allows for more freedom in the modeling assumptions and potentially decreases the number of parameters to be estimated, given that the limiting distribution only depends on the tail index of the data and the mean, which can be usually assumed to be a function of the tail index itself, as in the Paretian case where $\mu = \frac{\alpha}{\alpha-1}$.

[2] Another idea, which we have tested in writing the chapter, is to use the distance between the median and the mean; the performances are comparable.

By exploiting the location-scale property of α-stable distributions and Equation (3.11), we approximate the distribution of $G^{NP}(X_n)$ for finite samples by

$$G^{NP}(X_n) \sim S\left(\alpha, 1, \gamma(n), g\right), \tag{3.19}$$

where $\gamma(n) = \frac{1}{n^{\frac{\alpha-1}{\alpha}}} \frac{L_0(n)}{\mu}$ is the scale parameter of the limiting distribution.

As a consequence, thanks to the linearity of the mode for α-stable distributions, we have

$$|m(G^{NP}(X_n)) - \mathbb{E}(G^{NP}(X_n))| \approx |m(\alpha, \gamma(n)) + g - g| = |m(\alpha, \gamma(n))|,$$

where $m(\alpha, \gamma(n))$ is the mode function of an α-stable distribution with zero mean.

This means that, in order to obtain the correction term, the knowledge of the true Gini index is not necessary, view that $m(\alpha, \gamma(n))$ does not depend on g. We then estimate the correction term as

$$\hat{m}(\alpha, \gamma(n)) = \arg\max_x s(x), \tag{3.20}$$

where $s(x)$ is the numerical density of the associated α-stable distribution in Equation (3.19), but centered on 0. This comes from the fact that, for α-stable distributions, the mode is not available in closed form, but it can be easily computed numerically [26], using the unimodality of the law.

The corrected nonparametric estimator is thus

$$G^C(X_n) = G^{NP}(X_n) + \hat{m}(\alpha, \gamma(n)), \tag{3.21}$$

whose asymptotic distribution is

$$G^C(X_n) \sim S\left(\alpha, 1, \gamma(n), g + \hat{m}(\alpha, \gamma(n))\right). \tag{3.22}$$

Note that the correction term $\hat{m}(\alpha, \gamma(n))$ is a function of the tail index α and is connected to the sample size n by the scale parameter $\gamma(n)$ of the associated limiting distribution. It is important to point out that $\hat{m}(\alpha, \gamma(n))$ is decreasing in n, and that

$$\lim_{n\to\infty} \hat{m}(\alpha, \gamma(n)) \to 0.$$

This happens because, as n increases, the distribution described in Equation (3.19) becomes more and more centered around its mean value, shrinking to zero the distance between the mode and the mean. This ensures the asymptotic equivalence of the corrected estimator and the nonparametric one. Just observe that

$$\begin{aligned} \lim_{n\to\infty} |G(X_n)^C - G^{NP}(X_n)| &= \lim_{n\to\infty} |G^{NP}(X_n) + \hat{m}(\alpha, \gamma(n)) - G^{NP}(X_n)| \\ &= \lim_{n\to\infty} |\hat{m}(\alpha, \gamma(n))| \to 0. \end{aligned}$$

Naturally, thanks to the correction, $G^C(X_n)$ will always behave better in small samples. Please also consider that, from Equation (3.22), the distribution of the corrected estimator has now mean $g + \hat{m}(\alpha, \gamma(n))$, which converges to the true Gini g as $n \to \infty$.

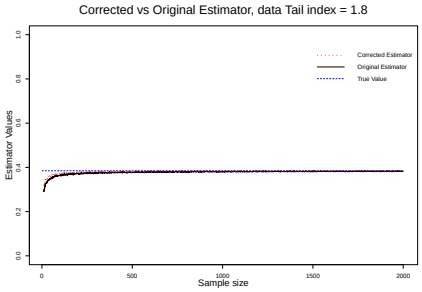
From a theoretical point of view, the quality of this correction depends on the distance between the exact distribution of $G^{NP}(X_n)$ and its α-stable limit; the closer the

two are to each other, the better the approximation. However, given that, in most cases, the exact distribution of $G^{NP}(X_n)$ is unknown, it is not possible to give more details.

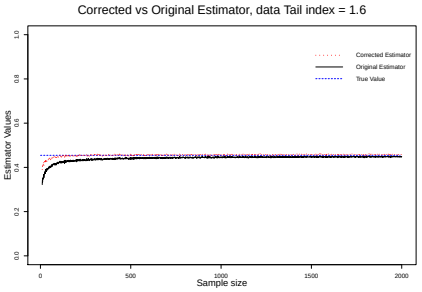
From what we have written so far, it is clear that the correction term depends on the tail index of the data, and possibly also on their mean. These parameters, if not assumed to be known a priori, must be estimated. Therefore the additional uncertainty due to the estimation will reflect also on the quality of the correction.

We conclude this Section by discussing the effect of the correction procedure with a simple example. In a Monte Carlo experiment, we simulate 1000 Paretian samples of increasing size, from $n = 10$ to $n = 2000$, and for each sample size we compute both the original nonparametric estimator $G^{NP}(X_n)$ and the corrected $G^{C}(X_n)$. We repeat the experiment for different α's. Figure 3.3 presents the results.

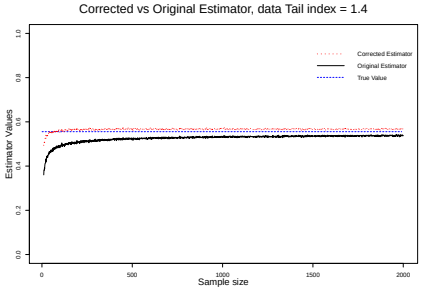
It is clear that the corrected estimators always perform better than the uncorrected ones in terms of absolute deviation from the true Gini value. In particular, our numerical experiment shows that for small sample sizes with $n \leq 1000$ the gain is quite remarkable for all the different values of $\alpha \in (1,2)$. However, as expected, the difference between the estimators decreases with the sample size, as the correction term decreases both in n and in the tail index α. Notice that, when the value of the tail index is equal to 2, we obtain the symmetric Gaussian distribution and the two estimators coincide, being the nonparametric estimator no longer biased, thanks to the finite variance.



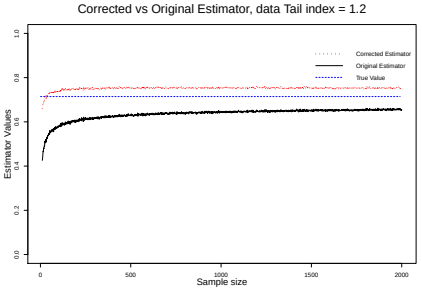
(a) $\alpha = 1.8$



(b) $\alpha = 1.6$



(c) $\alpha = 1.4$



(d) $\alpha = 1.2$

Figure 3.3: Comparisons between the corrected nonparametric estimator (in red, the one on top) and the usual nonparametric estimator (in black, the one below). For small sample sizes the corrected one clearly improves the quality of the estimation.

3.6. CONCLUSIONS

In this chapter we address the issue of the asymptotic behavior of the nonparametric estimator of the Gini index in presence of a distribution with infinite variance, an issue that has been curiously ignored by the literature.

In particular, we derive the asymptotic distribution of the nonparametric estimator of the Gini index under infinite second moment assumption of the data generating process. We show that such asymptotic distribution maintains the same tail index of the data generating process as-well-as becoming totally-right-skewed. This results casts light on the properties of such estimator in fat-tailed stochastic environment often encountered in practice. Additionally, we use such newly derived limit to build a bias-correction mechanics to deal with the well-known finite sample bias of the nonparametric Gini index estimator.

Finally, we compare our results with a fully parametric approach based on maximum likelihood estimators. As expected the MLE estimators has better efficiency properties but only provided that the true model is known. Therefore the trade of between efficiency and robustness must be taken into account when choosing which estimator to use.

TECHNICAL APPENDIX

PROOF OF LEMMA 3.1

Let $U = F(X)$ be the standard uniformly distributed integral probability transform of the random variable X. For the order statistics, we then have [25]: $X_{(i)} \overset{a.s.}{=} F^{-1}(U_{(i)})$. Hence

$$R_n = \frac{1}{n}\sum_{i=1}^{n}(i/n - U_{(i)})F^{-1}(U_{(i)}). \tag{3.23}$$

Now by definition of empirical c.d.f it follows that

$$R_n = \frac{1}{n}\sum_{i=1}^{n}(F_n(U_{(i)}) - U_{(i)})F^{-1}(U_{(i)}), \tag{3.24}$$

where $F_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1_{U_i \le u}$ is the empirical c.d.f of uniformly distributed random variables.

To show that $R_n \xrightarrow{L^1} 0$, we are going to impose an upper bound that goes to zero. First we notice that

$$\mathbb{E}|R_n| \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|(F_n(U_{(i)}) - U_{(i)})F^{-1}(U_{(i)})|. \tag{3.25}$$

To build a bound for the right-hand side (r.h.s) of (3.25), we can exploit the fact that, while $F^{-1}(U_{(i)})$ might be just L^1-integrable, $F_n(U_{(i)}) - U_{(i)}$ is L^∞ integrable, therefore we can use Hölder's inequality with $q = \infty$ and $p = 1$. It follows that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|(F_n(U_{(i)}) - U_{(i)})F^{-1}(U_{(i)})| \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)}) - U_{(i)})|\mathbb{E}|F^{-1}(U_{(i)})|. \tag{3.26}$$

Then, thanks to the Cauchy-Schwarz inequality, we get

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-U_{(i)})|\mathbb{E}|F^{-1}(U_{(i)})|$$
$$\leq\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-U_{(i)})|)^2\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}(F^{-1}(U_{(i)})))^2\right)^{\frac{1}{2}}. \tag{3.27}$$

Now, first recall that $\sum_{i=1}^{n}F^{-1}(U_{(i)})\overset{a.s.}{=}\sum_{i=1}^{n}F^{-1}(U_i)$ with U_i, $i=1,...,n$, being an i.i.d sequence, then notice that $\mathbb{E}(F^{-1}(U_i))=\mu$, so that the second term of Equation (3.27) becomes

$$\mu\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-U_{(i)})|)^2\right)^{\frac{1}{2}}. \tag{3.28}$$

The final step is to show that Equation (3.28) goes to zero as $n\to\infty$.

We know that F_n is the empirical c.d.f of uniform random variables. Using the triangular inequality the inner term of Equation (3.28) can be bounded as

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-U_{(i)})|)^2 \tag{3.29}$$
$$\leq\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-F(U_{(i)}))|)^2+\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}\sup_{U_{(i)}}|(F(U_{(i)})-U_{(i)})|)^2.$$

Since we are dealing with uniforms, we known that their cdf $F(x)$ is the identity map, therefore, $F(U)=u$, and the second term in the r.h.s of (3.29) vanishes.

We can then bound $\mathbb{E}(\sup_{U_{(i)}}|(F_n(U_{(i)})-F(U_{(i)})|)$ using the so called Vapnik-Chervonenkis (VC) inequality, a uniform bound for empirical processes [29–31], getting

$$\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-F(U_{(i)})|\leq\sqrt{\frac{\log(n+1)+\log(2)}{n}}. \tag{3.30}$$

Combining Equation (3.30) with Equation (3.28) we obtain

$$\mu\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}\sup_{U_{(i)}}|(F_n(U_{(i)})-U_{(i)})|)^2\right)^{\frac{1}{2}}\leq\mu\sqrt{\frac{\log(n+1)+\log(2)}{n}}, \tag{3.31}$$

which goes to zero as $n\to\infty$, thus proving the first claim.

For the second claim, it is sufficient to observe that the r.h.s of (3.31) still goes to zero when multiplied by $\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}$ if $\alpha\in(1,2)$.

PROOF OF THEOREM 3.2

The first part of the proof consists in showing that we can rewrite Equation (3.10) as a function of i.i.d random variables in place of order statistics, to be able to apply a Central Limit Theorem (CLT) argument.

Let's start by considering the sequence

$$\frac{1}{n}\sum_{i=1}^{n} Z_{(i)} = \frac{1}{n}\sum_{i=1}^{n}\left(2\frac{i-1}{n-1}-1\right)F^{-1}(U_{(i)}). \tag{3.32}$$

Using the integral probability transform $X \overset{d}{=} F^{-1}(U)$ with U standard uniform, and adding and removing $\frac{1}{n}\sum_{i=1}^{n}\left(2U_{(i)}-1\right)F^{-1}(U_{(i)})$, the r.h.s. in Equation (3.32) can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n} Z_{(i)} = \frac{1}{n}\sum_{i=1}^{n}(2U_{(i)}-1)F^{-1}(U_{(i)}) + \frac{1}{n}\sum_{i=1}^{n} 2\left(\frac{i-1}{n-1}-U_{(i)}\right)F^{-1}(U_{(i)}). \tag{3.33}$$

Then, by using the properties of order statistics [25] we obtain the following almost sure equivalence

$$\frac{1}{n}\sum_{i=1}^{n} Z_{(i)} \overset{a.s.}{=} \frac{1}{n}\sum_{i=1}^{n}(2U_i-1)F^{-1}(U_i) + \frac{1}{n}\sum_{i=1}^{n} 2\left(\frac{i-1}{n-1}-U_{(i)}\right)F^{-1}(U_{(i)}). \tag{3.34}$$

Note that the first term in the r.h.s of (3.34) is a function of i.i.d random variables as desired, while the second term is just a remainder, therefore

$$\frac{1}{n}\sum_{i=1}^{n} Z_{(i)} \overset{a.s.}{=} \frac{1}{n}\sum_{i=1}^{n} Z_i + R_n,$$

with $Z_i = (2U_i-1)F^{-1}(U_i)$ and $R_n = \frac{1}{n}\sum_{i=1}^{n}(2(\frac{i-1}{n-1}-U_{(i)}))F^{-1}(U_{(i)})$.

Given Equation (3.10) and exploiting the decomposition given in (3.34) we can rewrite our claim as

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(\frac{1}{n}\sum_{i=1}^{n} Z_{(i)}-\theta\right) = \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i-\theta\right) + \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}R_n. \tag{3.35}$$

From the second claim of the Lemma 3.1 and Slutsky Theorem, the convergence in Equation (3.10) can be proven by looking at the behavior of the sequence

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i-\theta\right), \tag{3.36}$$

where $Z_i = (2U_i-1)F^{-1}(U_i) = (2F(X_i)-1)X_i$. This reduces to proving that Z_i is in the fat tails domain of attraction.

Recall that by assumption $X \in DA(S_{\alpha,\beta})$ with $\alpha \in (1,2)$. This assumption enables us to use a particular type of CLT argument for the convergence of the sum of fat-tailed random variables. However, we first need to prove that $Z \in DA(S_{\alpha,\beta})$ as well, that is $P(|Z|>z) \sim L(z)z^{-\alpha}$, with $\alpha \in (1,2)$ and $L(z)$ slowly-varying.

Notice that

$$P(|\tilde{Z}|>z) \leq P(|Z|>z) \leq P(X>z),$$

where $\tilde{Z} = UX$, $U \perp X$ and $U \sim Unif(0,1)$. The first bound holds because of the positive dependence between X and $F(X)$ and it can be proven rigorously by noting that $2UX \leq$

$2F(X)X$ by the so-called re-arrangement inequality [32]. The upper bound conversely is trivial.

Using the properties of slowly-varying functions, we have $P(X > z) \sim L(z)z^{-\alpha}$. To show that $\tilde{Z} \in DA(S_{\alpha,\beta})$, we use Breiman's Theorem, which ensure the stability of the α-stable class under products, as long as the second random variable is not too fat-tailed [33].

To apply the Theorem we re-write $P(|\tilde{Z}| > z)$ as

$$P(|\tilde{Z}| > z) = P(\tilde{Z} > z) + P(-\tilde{Z} > z) = P(\tilde{U}X > z) + P(-\tilde{U}X > z),$$

where $\tilde{U}$ is a uniform $Unif(-1,1)$ with $\tilde{U} \perp X$.

We focus on $P(\tilde{U}X > z)$ since the procedure is the same for $P(-\tilde{U}X > z)$. We have

$$P(\tilde{U}X > z) = P(\tilde{U}X > z|\tilde{U} > 0)P(\tilde{U} > 0) + P(\tilde{U}X > z|\tilde{U} \leq 0)P(\tilde{U} \leq 0),$$

for $z \to +\infty$.

Now, we have that $P(\tilde{U}X > z|\tilde{U} \leq 0) \to 0$, while, by applying Breiman's Theorem, $P(\tilde{U}X > z|\tilde{U} > 0)$ becomes

$$P(\tilde{U}X > z|\tilde{U} > 0) \to E(\tilde{U}^{\alpha}|U > 0)P(X > z)P(U > 0).$$

Therefore

$$P(|\tilde{Z}| > z) \to \frac{1}{2}E(\tilde{U}^{\alpha}|U > 0)P(X > z) + \frac{1}{2}E((-\tilde{U})^{\alpha}|U \leq 0)P(X > z).$$

From this

$$\begin{aligned} P(|\tilde{Z}| > z) &\to \frac{1}{2}P(X > z)[E(\tilde{U})^{\alpha}|U > 0) + E((-\tilde{U}^{\alpha}|U \leq 0)] \\ &= \frac{1}{1+\alpha}P(X > z) \sim \frac{1}{1+\alpha}L(z)z^{-\alpha}. \end{aligned}$$

We can then conclude that, by the squeezing Theorem [20],

$$P(|Z| > z) \asymp L(z)z^{-\alpha},$$

as $z \to \infty$. Therefore $Z \in DA(S_{\alpha,\beta})$.

We are now ready to invoke the Generalized Central Limit Theorem (GCLT) [5] for the sequence Z_i, i.e.

$$nc_n^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mathbb{E}(Z_i)\right) \xrightarrow{d} S_{\alpha,\beta}. \tag{3.37}$$

with $\mathbb{E}(Z_i) = \theta$, $S_{\alpha,\beta}$ a standardized α-stable random variable, and where c_n is a sequence which must satisfy

$$\lim_{n\to\infty} \frac{nL(c_n)}{c_n^{\alpha}} = \frac{\Gamma(2-\alpha)|\cos(\frac{\pi\alpha}{2})|}{\alpha - 1} = C_{\alpha}. \tag{3.38}$$

Notice that c_n can be represented as $c_n = n^{\frac{1}{\alpha}}L_0(n)$, where $L_0(n)$ is another slowly-varying function possibly different from $L(n)$.

The skewness parameter β is such that

$$\frac{P(Z>z)}{P(|Z|>z)} \to \frac{1+\beta}{2}.$$

Recalling that, by construction, $Z \in [-c, +\infty)$, the above expression reduces to

$$\frac{P(Z>z)}{P(Z>z)+P(-Z>z)} \to \frac{P(Z>z)}{P(Z>z)} = 1 \to \frac{1+\beta}{2}, \tag{3.39}$$

therefore $\beta = 1$. This, combined with Equation (3.35), the result for the reminder R_n of Lemma 3.1 and Slutsky Theorem, allows us to conclude that the same weak limits holds for the ordered sequence of $Z_{(i)}$ in Equation (3.10) as well.

PROOF OF THEOREM 3.3

The first step of the proof is to show that the ratio $\frac{\sum_{i=1}^n Z_{(i)}}{\sum_{i=1}^n X_i}$, characterizing the Gini index, is equivalent in distribution to the ratio $\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i}$. In order to prove this, it is sufficient to apply the factorization in Equation (3.34) to Equation (3.11), getting

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i} - \frac{\theta}{\mu}\right) + \frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} R_n \frac{n}{\sum_{i=1}^n X_i}. \tag{3.40}$$

By Lemma 3.1 and the application of the continuous mapping and Slutsky Theorems, the second term in Equation (3.40) goes to zero at least in probability. Therefore to prove the claim it is sufficient to derive a weak limit for the following sequence

$$n^{\frac{\alpha-1}{\alpha}} \frac{1}{L_0(n)}\left(\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i} - \frac{\theta}{\mu}\right). \tag{3.41}$$

Expanding Equation (3.41) and recalling that $Z_i = (2F(X_i)-1)X_i$, we get

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \frac{n}{\sum_{i=1}^n X_i}\left(\frac{1}{n}\sum_{i=1}^n X_i\left(2F(X_i) - 1 - \frac{\theta}{\mu}\right)\right). \tag{3.42}$$

The term $\frac{n}{\sum_{i=1}^n X_i}$ in Equation (3.42) converges in probability to $\frac{1}{\mu}$ by an application of the continuous mapping Theorem, and the fact that we are dealing with positive random variables X. Hence it will contribute to the final limit via Slutsky Theorem.

We first start by focusing on the study of the limit law of the term

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \frac{1}{n}\sum_{i=1}^n X_i\left(2F(X_i) - 1 - \frac{\theta}{\mu}\right). \tag{3.43}$$

Set $\hat{Z}_i = X_i(2F(X_i) - 1 - \frac{\theta}{\mu})$ and note that $\mathbb{E}(\hat{Z}_i) = 0$, since $\mathbb{E}(Z_i) = \theta$ and $\mathbb{E}(X_i) = \mu$.

In order to apply a GCLT argument to characterize the limit distribution of the sequence $\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)} \frac{1}{n}\sum_{i=1}^n \hat{Z}_i$ we need to prove that $\hat{Z} \in DA(S_\alpha)$. If so then we can apply GCLT to

$$\frac{n^{\frac{\alpha-1}{\alpha}}}{L_0(n)}\left(\frac{\sum_{i=1}^n \hat{Z}_i}{n} - \mathbb{E}(\hat{Z}_i)\right). \tag{3.44}$$

Note that, since $\mathbb{E}(\hat{Z}_i) = 0$, Equation (3.44) equals Equation (3.43).

To prove that $\hat{Z} \in DA(S_{\alpha,\beta})$, remember that $\hat{Z}_i = X_i(2F(X_i)-1-\frac{\theta}{\mu})$ is just $Z_i = X_i(2F(X_i)-1)$ shifted by $\frac{\theta}{\mu}$. Therefore the same argument used in Theorem 3.2 for Z applies here to show that $\hat{Z} \in DA(S_{\alpha,\beta})$. In particular we can point out that $\hat{Z}$ and Z (therefore also X) share the same α and slowly-varying function $L(n)$.

Notice that by assumption $X \in [c,\infty)$ with $c > 0$ and we are dealing with continuous distributions, therefore $\hat{Z} \in [-c(1+\frac{\theta}{\mu}),\infty)$. As a consequence the left tail of $\hat{Z}$ does not contribute to changing the limit skewness parameter β, which remains equal to 1 (as for Z) by an application of Equation (3.39).

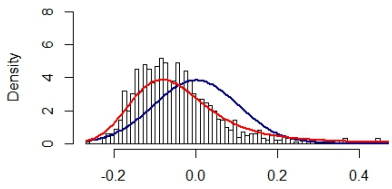
Therefore, by applying the GCLT we finally get

$$n^{\frac{\alpha-1}{\alpha}} \frac{1}{L_0(n)} \left(\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i} - \frac{\theta}{\mu}\right) \xrightarrow{d} \frac{1}{\mu} S(\alpha,1,1,0). \tag{3.45}$$

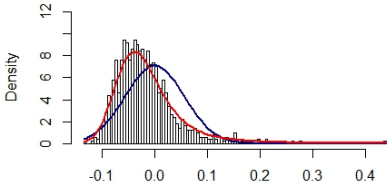
We conclude the proof by noting that, as proven in Equation (3.40), the weak limit of the Gini index is characterized by the ratio of $\frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n X_i}$ rather than the ordered one, and that an α-stable random variable is closed under scaling by a constant [27].

APPENDIX

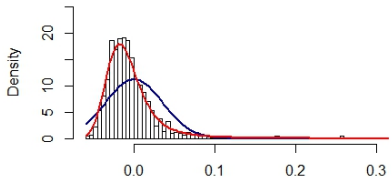
In this Section we provide some empirical evidence of how out asymptotic result approximates the finite sample distribution of $G^{NP}(X)$ for difference choice of sample size n and tail parameter α. All samples are taken from a Pareto random variable with pdf given by Equation (3.14) and redrawn 1000 times.



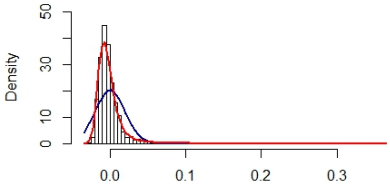
(a) $\alpha = 1.5$, $n = 100$



(b) $\alpha = 1.5$, $n = 1000$



(c) $\alpha = 1.5$, $n = 10000$



(d) $\alpha = 1.5$, $n = 100000$

Figure 3.4: In red the α-stable limiting approximation, in blue the Guassian CLT approximation.

REFERENCES

[1] A. Fontanari, N. N. Taleb, and P. Cirillo, *Gini estimation under infinite variance,* Physica A: Statistical Mechanics and its Applications **502**, 256 (2018).

[2] B. K. Chakrabarti, A. Chakraborti, S. R. Chakravarty, and A. Chatterjee, *Econophysics of income and wealth distributions* (Cambridge University Press, 2013).

[3] C. Kleiber and S. Kotz, *Statistical size distributions in economics and actuarial sciences.*, Vol. 470 (John Wiley & Sons, 2003).

[4] V. Pareto, *La courbe de la répartition de la richesse* (Librairie Droz, 1967).

[5] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events: for insurance and finance*, Vol. 33 (Springer Science & Business Media, 2013).

[6] I. Eliazar and M. H. Cohen, *On social inequality: Analyzing the rich–poor disparity,* Physica A: Statistical Mechanics and its Applications **401**, 148 (2014).

[7] C. Gini, *Variabilità e mutabilità (1912), reprinted in: Variabilità e mutabilità, e,* Pizetti and T. Salvemini, Memorie di Metodologica Statistica, Libreria Eredi Virgilio Veschi (1955).

[8] S. Yitzhaki and E. Schechtman, *The Gini methodology: A primer on a statistical methodology*, Vol. 272 (Springer Science & Business Media, 2012).

[9] D. Chotikapanich, *Modeling income distributions and Lorenz curves*, Vol. 5 (Springer Science & Business Media, 2008).

[10] T. Piketty, *About capital in the twenty-first century,* American Economic Review **105**, 48 (2015).

[11] I. Eliazar, *Inequality spectra,* Physica A: Statistical Mechanics and its Applications **469**, 824 (2017).

[12] I. Eliazar and I. M. Sokolov, *Maximization of statistical heterogeneity: From shannon's entropy to gini's index,* Physica A: Statistical Mechanics and its Applications **389**, 3023 (2010).

[13] I. I. Eliazar and I. M. Sokolov, *Gini characterization of extreme-value statistics,* Physica A: Statistical Mechanics and its Applications **389**, 4462 (2010).

[14] A. Fontanari, P. Cirillo, and C. W. Oosterlee, *From concentration profiles to concentration maps. new tools for the study of loss distributions,* Insurance: Mathematics and Economics **78**, 13 (2018).

[15] G. Deltas, *The small-sample bias of the gini coefficient: results and implications for empirical research,* Review of economics and statistics **85**, 226 (2003).

[16] N. N. Taleb and R. Douady, *On the super-additivity and estimation biases of quantile contributions,* Physica A: Statistical Mechanics and its Applications **429**, 252 (2015).

[17] A. H. Jessen and T. Mikosch, *Regularly varying functions,* Publications de L'institut Mathematique (2006).

[18] L. De Haan and A. Ferreira, *Extreme value theory: an introduction* (Springer Science & Business Media, 2007).

[19] P. Cirillo, *Are your data really pareto distributed?* Physica A: Statistical Mechanics and its Applications **392**, 5947 (2013).

[20] W. Feller, *An introduction to probability theory and its applications*, Vol. 2 (John Wiley & Sons, 2008).

[21] T. Piketty, *The economics of inequality* (Harvard University Press, 2015).

[22] I. I. Eliazar and I. M. Sokolov, *Measuring statistical evenness: A panoramic overview,* Physica A: Statistical Mechanics and its Applications **391**, 1323 (2012).

[23] W. Hoeffding, *A non-parametric test of independence,* The annals of mathematical statistics , 546 (1948).

[24] D. Li, M. B. Rao, and R. Tomkins, *The law of the iterated logarithm and central limit theorem for l-statistics,* Journal of multivariate analysis **78**, 191 (2001).

[25] H. A. David and H. N. Nagaraja, *Order statistics,* Encyclopedia of Statistical Sciences (2004).

[26] J. P. Nolan, *Parameterizations and modes of stable distributions,* Statistics & probability letters **38**, 187 (1998).

[27] G. Samoradnitsky, *Stable non-Gaussian random processes: stochastic models with infinite variance* (Routledge, 2017).

[28] J. Shao, *Mathematical statistics: exercises and solutions* (Springer Science & Business Media, 2006).

[29] A. DasGupta, *Probability for statistics and machine learning: fundamentals and advanced topics* (Springer Science & Business Media, 2011).

[30] A. W. Vaart and J. A. Wellner, *Weak convergence and empirical processes: with applications to statistics* (Springer, 1996).

[31] O. Bousquet, S. Boucheron, and G. Lugosi, *Introduction to statistical learning theory,* in *Summer School on Machine Learning* (Springer, 2003) pp. 169–207.

[32] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*, Vol. 143 (Springer, 1979).

[33] Y. Yang, S. Hu, and T. Wu, *The tail probability of the product of dependent random variables from max-domains of attraction,* Statistics & Probability Letters **81**, 1876 (2011).

4

QUANTUM MAJORIZATION FOR FINANCIAL CORRELATION MATRICES

We propose quantum majorization as a way of comparing and ranking correlation matrices, with the aim of assessing portfolio risk in a unified framework. Quantum majorization is a partial order in the space of correlation matrices, which are evaluated through their spectra.

We discuss the connections between quantum majorization and an important class of risk functionals, and we define two new risk measures able to capture interesting characteristics of portfolio risk.

Keywords: *Majorization; correlation matrix; Portfolio selection.*

Parts of this chapter are available online at the SSRN Electronic Journal [1] and are currently under review.

4.1. INTRODUCTION

Consider a portfolio $\mathscr{P}$ containing n assets. A common way to deal with the dependence structure of $\mathscr{P}$–and thus with portfolio risk–is to consider its $n \times n$ correlation matrix $\mathbf{C}$ [2].

We refer to the correlation matrix in the most general sense, as a real, symmetric, positive semidefinite, Hermitian matrix, whose entries are correlation coefficients according to some definition, all lying in the interval $[-1,1]$, with all ones on the main diagonal.

The approach here proposed is indeed applicable to all correlation matrices, and not only to the classical Pearson's correlation matrix: one can take into consideration matrices based on Spearman's ρ, Kendall τ, Gini correlation, etc. [3–6]. This allows to deal with more general forms of dependence, beyond the standard linear one of Pearson's correlation [7].

The main idea of this work is to find a way of comparing correlation matrices, and of extracting the portfolio risk information they contain, so that they can be ranked. The comparison can be over time, if one studies the evolution of the correlation matrix for a given portfolio, but it can also be cross-sectional, comparing different portfolios at the same time, for example to look for the one minimizing portfolio risk. The only requirement is that the size of the matrices is the same, i.e. only portfolios having the same number of assets are considered.

Our proposal is to use an ordering, developed in the field of quantum mechanics [8–10], called quantum majorization, to study the dynamics of the entropy of a quantum system, applying it to correlation matrices. To the best of our knowledge, this is the first time such an ordering is used in finance.

Quantum majorization is a partial order to rank matrices looking at their eigenvalues. The use of eigensystems to study multivariate dependence is not at all new [7], but we show how the spectrum of a correlation matrix can be used to capture relevant features of portfolio risk in a brand new way.

We introduce the $\mathscr{M}_\lambda$ class of risk functionals, which are isotonic to quantum majorization, and whose aim is to capture the (monotonic) dependence embedded into portfolio correlation matrices. An important property of such a class, stated in Proposition 4.2, is that under quantum majorization, i.e. when it is possible to rank correlation matrices according to the order introduced in Definition 4.3, all risk functionals in $\mathscr{M}_\lambda$ are comonotonic. The implication is that, if we are able to identify majorization, then the choice of the risk functional becomes secondary, as they will all behave in the same way, indicating an increase (or a decrease) of portfolio risk. It is when the ordering does not hold–as we shall see–that risk functionals may give inconsistent information.

With respect to single risk measures, quantum majorization thus provides a stronger characterization of risk and dependence among correlation matrices. We are therefore able to provide a unifying approach to the analysis of portfolio risk and correlation: we introduce several tools, we discuss their properties, and we show how to use them in practice. In doing so, we will avoid all unnecessary sophistication, giving space to financial interpretability and usability.

The chapter is organized as follows: Section 4.2 introduces the concept of quantum majorization for correlation matrices; Section 4.3 deals with the $\mathscr{M}_\lambda$ class of isotonic risk measures, analyzing its properties, and discussing its link with quantum majorization;

Section 4.4 introduces the empirical majorization matrix approach as a fully data-driven methodology to deal with portfolio risk; Section 4.5 is devoted to an application to empirical data related to the Industrial Dow Jones, in which we show how to use the tools introduced in the previous sections; finally Section 4.6 builds a connection between our approach and network analysis, which can open the path for future research.

4.2. THE QUANTUM MAJORIZATION OF CORRELATION MATRICES

The aim is to deal with portfolio risk, as represented by correlation matrices. We thus look for an order relation allowing us to compare and rank them. To guarantee financial applicability and interpretability, such an order should respect the following conditions:

C1 *Minimal Element*: According to the order, the least risky element among all correlation matrices should be the identity matrix;

C2 *Maximal Element(s)*: The riskiest elements among all correlation matrices should be the set of the matrices of rank 1;

C3 *Monotonocity*: The order should not increase in the rank of the correlation matrices;

C4 *Convexity*: The correlation matrix obtained as convex combination of two correlation matrices should not be riskier than the convex combination of the two original ones.

Conditions C1 and C2 fix the two extremes of the ordered set, i.e. the minimal and the maximal elements. C1 identifies the case of no correlation (the identity matrix) as the least risky one; while C2 finds the riskiest situation in a portfolio whose assets are all monotonic transformations of one of them (comonotonicity and countermonotonicity), therefore all correlation matrices of rank 1. From a portfolio risk prospective, we can think about the propagation of market shocks: if assets in our portfolio are completely uncorrelated, shocks on one of them do not propagate to any of the others; while in a comonotonic (countermonotonic) portfolio, shocks on one single asset affect all the others, possibly increasing the overall risk.

Notice that while the minimal element is necessarily unique (the identity), the maximal one corresponds to a set of correlation matrices, given that in a comonotonic portfolio all assets can represent the leading term.

Condition C3 implies that, given two portfolios $\mathscr{P}_1$ and $\mathscr{P}_2$, such that $\mathscr{P}_1$ has more monotonically dependent assets than $\mathscr{P}_2$ (hence its correlation matrix a smaller rank), then the risk associated to $\mathscr{P}_1$ should never be lower than that of $\mathscr{P}_2$.

Finally, condition C4 reflects the usual financial postulate that diversification decreases risk [11]. While the condition is stated for correlation matrices, the object of our analysis, in a sense we are requiring that, by taking convex combinations of portfolios– not necessarily uncorrelated, the overall risk should decrease.

Studying the available literature, a powerful partial order, compatible with our conditions C1-C4, and thus useful to account for portfolio risk, was introduced in the field

of quantum mechanics by Alberti and Uhlmann [8], and further analyzed by Ando [12]. This order, here called quantum majorization, was developed to study entropy increasing dynamics on density matrices, and it relies on the spectra (the vectors containing the eigenvalues) of the matrices under scrutiny, which are required to be Hermitian and with equal trace.

Being real symmetric, a correlation matrix is a special type of Hermitian matrix, and its trace is equal to the number of assets the matrix represents: n. In other words, the sum of the eigenvalues of a correlation matrix–equal to its trace–is nothing but the number of elements it deals with. Therefore, if we consider two different $n \times n$ correlation matrices, we know for sure that their spectra will sum to the same value n. This apparently simple property justifies the use of quantum majorization on correlation matrices, and not on variance-covariance ones.

It is important to stress that, in introducing their order, Alberti and Uhlmann [8] did not consider the C1-C4 conditions stated above, which are just relevant to us in terms of portfolio risk[1].

The conditions above can be easily restated in terms of the spectra of correlation matrices, a trick that will prove essential to deal with quantum majorization as per Theorem 4.1 below.

As stressed later, this "eigenrepresentation" of the axiomatic conditions also bridges towards the spectral study of random matrices [15, 16], and classical multivariate analysis à la Wilks [7], further supporting the approach we are proposing.

For instance, for a portfolio containing n uncorrelated assets, C1 requires the spectrum of the correlation matrix to be the n-dimensional vector of ones $\boldsymbol{\lambda} = [1, ..., 1]$. An $n \times n$ identity matrix (the minimal element according to C1) has indeed a single eigenvalue equal to 1, with an algebraic multiplicity of n. On the opposite side, C2 requires the spectrum of the correlation matrix of a comonotonic/countermonotonic portfolio to be equal to $\boldsymbol{\lambda} = [n, 0, ..., 0]$, which is the case of an $n \times n$ matrix of rank 1.

Before formally introducing quantum majorization for correlation matrices, we first need some basic results from majorization theory [17], an important field of linear algebra and order theory.

Introduced by the works of Polya, Hardly, Littlewood, Dalton, Muirhead and Schur [18, 19], majorization is a way to define a partial order in the space of vectors in $\mathbb{R}^n$, ranking them in terms of their intrinsic variability, i.e. how scattered they are with respect to their common vector of averages. Given a vector $\mathbf{x} = [x_1, x_2, ..., x_n] \in \mathbb{R}^n$, such that $\sum_{i=1}^n x_i = d$, the vector of averages is defined as

$$\bar{\mathbf{x}} = \left[\frac{\sum_{i=1}^n x_i}{n}, ..., \frac{\sum_{i=1}^n x_i}{n} \right] = \left[\frac{d}{n}, ..., \frac{d}{n} \right],$$

that is the n-dimensional vector whose entries correspond to the average of $\mathbf{x}$.

[1] A successful alternative to rank correlation matrices with conditions similar to C1-C4 has been developed by Giovagnoli and Romanazzi [13], using a special type of G-majorization [14]. However we believe that their approach, despite being theoretically fascinating, it is difficult to use in practice, especially for risk management purposes. Moreover, one can prove that the order we consider here is richer, as it orders at least as many elements as Giovagnoli and Romanazzi's one.

Definition 4.1. Take two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We say that $\mathbf{x}$ majorizes $\mathbf{y}$, in symbols $\mathbf{x} \succ \mathbf{y}$, if

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

and

$$\sum_{i=1}^{k} x_{[i]} \geq \sum_{i=1}^{k} y_{[i]}, \text{ for all } k = \{1, ..., n-1\}, \tag{4.1}$$

where $x_{[1]}, ..., x_{[n]}$ are the coordinates of the vector $\mathbf{x}$ sorted in descending order, so that $x_{[1]} \geq x_{[2]} \geq ... \geq x_{[n]}$. If the condition $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ is not satisfied, we speak of weak majorization, $\mathbf{x} \overset{w}{\succ} \mathbf{y}$.

Being a partial order [17],it may not be possible to verify the majorization between $\mathbf{x}$ and $\mathbf{y}$, as per Definition 4.1. In these cases, we write $\mathbf{x} \not\succ \mathbf{y}$. The fact that it is a partial order, however, guarantees that majorization respects the transitivity property: if $\mathbf{x} \succ \mathbf{y}$ and $\mathbf{y} \succ \mathbf{z}$, then $\mathbf{x} \succ \mathbf{z}$.

Strictly related to majorization, and fundamental for our work, is the class of Schur-convex or isotonic functions. These functions have indeed the property of preserving majorization when applied.

Definition 4.2. Let ϕ be a real valued function defined on $\mathbb{R}^n$, ϕ is Schur-convex if, whenever $\mathbf{x} \succ \mathbf{y}$, then $\phi(\mathbf{x}) \geq \phi(\mathbf{y})$. When the inequality is strict, we speak of strictly Schur-convex function.

If $\mathbf{x} \succ \mathbf{y}$ and $\phi(\mathbf{x}) \leq \phi(\mathbf{y})$ we call ϕ Schur-concave, such that $-\phi$ is Schur-convex.

Schur-convex functions can thus be seen as summary measures of the variability of a vector, when variability is defined in terms of majorization. Interestingly, several quantities commonly used in statistics and science to represent variability are Schur-convex: the variance, the coefficient of variation, the entropy, the arithmetic and the geometric means, the mean absolute deviation and inequality indices like the Gini and the Pietra [18]. These common measures of variability are therefore nothing more than functions naturally related to the concept of majorization, as we shall also see in the multivariate framework.

Let $\mathscr{C}$ be the class of correlation matrices of a given size, say $n \times n$. The following definition introduces what we will call quantum majorization.

Definition 4.3. Consider two correlation matrices $\mathbf{C}_1, \mathbf{C}_2 \in \mathscr{C}$, and denote their spectra by $\boldsymbol{\lambda}(\mathbf{C}_1)$ and $\boldsymbol{\lambda}(\mathbf{C}_2)$ respectively. We say that $\mathbf{C}_1$ quantum majorizes $\mathbf{C}_2$, i.e. $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$, if the spectrum of $\mathbf{C}_1$ majorizes the spectrum of $\mathbf{C}_2$, as per Definition 4.1, that is $\boldsymbol{\lambda}(\mathbf{C}_1) \succ \boldsymbol{\lambda}(\mathbf{C}_2)$.

It is important to remark that Definition 4.3 induces a partial order in the spectra of the correlation matrices, but only an equivalence relation on the space of correlation matrices themselves. In fact, if $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$ and $\mathbf{C}_2 \overset{\lambda}{\succ} \mathbf{C}_1$, we can conclude that $\boldsymbol{\lambda}(\mathbf{C}_1) = \boldsymbol{\lambda}(\mathbf{C}_2)$, that is to say that $\mathbf{C}_1$ and $\mathbf{C}_2$ are similar. This, in terms of risk, implies that $\mathbf{C}_1$ and $\mathbf{C}_2$ have the same riskiness. As an example, just consider all the rank 1 matrices involved in condition C2.

Theorem 4.1 below shows that the majorization[2] of Definition 4.3 satisfies conditions C1-C4, and it is therefore suitable to build comparisons among portfolio correlation matrices, providing a bridge between quantum mechanics and finance.

Theorem 4.1. *The partial order of Definition 4.3 is consistent with the requirements of minimal and maximal element, monotonicity and convexity of conditions C1-C4.*

Proof. The first two requirements are easily checked. Their proof follows from the definition of quantum majorization as majorization of the eigenvalues of the correlation matrix $\mathbf{C}$, and by noticing that all correlation matrices of size $n \times n$ have trace equal to $n = \sum_{i=1}^{n} \lambda_i(\mathbf{C})$.

Since every correlation matrix is positive semi-definite, we have that its eigenvalues are nonnegative, so that $\boldsymbol{\lambda}(\mathbf{C}) \in \mathbb{R}_+^n$. Therefore, using a result for vector majorization (Proposition C1 page 192 in [19]), we have that

$$\boldsymbol{\lambda}_{\max}(\mathbf{C}) \overset{\lambda}{\succ} \boldsymbol{\lambda}(\mathbf{C}) \overset{\lambda}{\succ} \boldsymbol{\lambda}_{\min}(\mathbf{C}),$$

where $\boldsymbol{\lambda}_{\max}(\mathbf{C}) = [n, 0, ..., 0]$ and $\boldsymbol{\lambda}_{\min}(\mathbf{C}) = [1, 1, ..., 1]$.

To verify C1, notice that the only diagonalizable matrix with spectrum equal to $\boldsymbol{\lambda}_{\min}(\mathbf{C})$ is the identity matrix. For C2, observe that the only matrices with spectrum equal to $\boldsymbol{\lambda}_{\max}(\mathbf{C})$ are those of rank 1.

To prove C3 recall that, given two correlation matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ of equal size, if $Rank(\mathbf{C}_1) \geq Rank(\mathbf{C}_2)$ then $\mathbf{C}_1$ must have a smaller number of zeros in its spectrum. Therefore, by applying Definition 4.1, we have that $\sum_{i=1}^{Rank(\mathbf{C}_2)} \lambda_{[i]}(\mathbf{C}_2) > \sum_{i=1}^{Rank(\mathbf{C}_2)} \lambda_{[i]}(\mathbf{C}_1)$, meaning that $\boldsymbol{\lambda}(\mathbf{C}_1)$ can never majorize $\boldsymbol{\lambda}(\mathbf{C}_2)$, thus verifying C3.

To check C4, we want to verify that, given two correlation matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ of equal size, and a number $\alpha \in [0, 1]$, we have

$$\sum_{i=1}^{k} \lambda_{[i]}(\mathbf{C}) \leq \alpha \sum_{i=1}^{k} \lambda_{[i]}(\mathbf{C}_1) + (1-\alpha) \sum_{i=1}^{k} \lambda_{[i]}(\mathbf{C}_2),$$

for every $k = 1, ..., n$, where $\mathbf{C} = \alpha\mathbf{C}_1 + (1-\alpha)\mathbf{C}_2$.

By the Fan representation theorem [20], we have that $\max_{\mathbf{U}\mathbf{U}^T=\mathbf{I}_k} Tr(\mathbf{U}\mathbf{C}\mathbf{U}^T) = \sum_{i=1}^{k} \lambda_{[i]}(\mathbf{C})$, with $\mathbf{U}$ a $k \times n$ unitary matrix and $\mathbf{I}_k$ the $k \times k$ identity. By the linearity of the trace operator and the subadditivity of the max, we then have

$$\max_{\mathbf{U}\mathbf{U}^T=\mathbf{I}_k} Tr(\mathbf{U}\mathbf{C}\mathbf{U}^T) \leq \alpha \max_{\mathbf{U}\mathbf{U}^T=\mathbf{I}_k} Tr(\mathbf{U}\mathbf{C}_1\mathbf{U}^T) + (1-\alpha) \max_{\mathbf{U}\mathbf{U}^T=\mathbf{I}_k} Tr(\mathbf{U}\mathbf{C}_2\mathbf{U}^T).$$

By applying the Fan representation theorem on both sides of the inequality above the desired result is obtained. □

The following proposition underlines an important connection between quantum majorization and the correlation matrices on which it is verified, providing a strong argument in favor of the use of majorization in studying portfolio risk.

[2] From now on, to avoid excessive repetitions, we just speak of majorization for both vectors and matrices. Naturally, when dealing with matrices we will refer to Definition 4.3, while for vectors we will imply Definition 4.1.

Proposition 4.1. *Take two correlation matrices* $\mathbf{C}_1, \mathbf{C}_2 \in \mathscr{C}$. *If* $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$, *then*

$$\mathbf{C}_2 = \sum_{i=1}^{n} \alpha_i \mathbf{U}_i \mathbf{C}_1 \mathbf{U}_i^T \tag{4.2}$$

where $(\alpha_i)_{i=1}^n$ *is a vector of probabilities summing to 1 and* $(\mathbf{U}_i)_{i=1}^n$ *is a collection of unitary matrices.*

with n being the size of the correlation matrix. A general proof of the proposition can be found in [9, 12], not only for correlation matrices. What is relevant to us is that Equation (4.2) builds a connection between majorization and the related correlations. In fact, if $\mathbf{C}_1$ majorizes $\mathbf{C}_2$, then the latter can be expressed in terms of the former, expliciting the underlying dependence. In the context of portfolio risk, this tells us that quantum majorization unveils a deep and powerful link among correlation matrices, usually not visible by only looking at the traditional risk measures.

4.3. THE $\mathscr{M}_\lambda$ CLASS OF MONOTONIC PORTFOLIO RISK MEASURES

Now that we have identified the quantum majorization of Definition 4.3 as a proper (risk) order in the space of correlation matrices, we are ready to quantify the amount of portfolio risk embedded in a given correlation matrix. To tackle this problem, we introduce a special class of matrix risk functionals, the $\mathscr{M}_\lambda$ class.

Definition 4.4. Given the set of $n \times n$ correlation matrices $\mathscr{C}$, the class $\mathscr{M}_\lambda$ contains all the $\overset{\lambda}{\succ}$-monotone (isotonic) functionals that are real-valued matrix functions $\phi : \mathscr{C} \to \mathbb{R}$ such that, for $\mathbf{C}_1, \mathbf{C}_2 \in \mathscr{C}$, if $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$, then $\phi(\mathbf{C}_1) \geq \phi(\mathbf{C}_2)$.

While a complete characterization of the functionals in the $\mathscr{M}_\lambda$ class is not available at the moment, later in Lemma 4.5 we provide a sufficient condition for a matrix function to be in $\mathscr{M}_\lambda$.

The following proposition collects important properties of the $\mathscr{M}_\lambda$ class.

Proposition 4.2. *The class* $\mathscr{M}_\lambda$ *exhibits the following properties:*

1. Comonotonicity with respect to majorization*: When applied to correlation matrices that are ordered according to the quantum majorization, all functionals in* $\mathscr{M}_\lambda$ *are comonotonic, i.e. they move in the same direction in terms of risk. When majorization does not hold, this is no longer guaranteed;*

2. Closure under increasing functions*: The class* $\mathscr{M}_\lambda$ *is closed under increasing functions. If* $\{\phi_1, ..., \phi_k\} \in \mathscr{M}_\lambda$ *and* $h : \mathbb{R}^k \to \mathbb{R}$ *is an increasing function in its arguments, then* $h(\phi_1, ..., \phi_k)$ *belongs to* $\mathscr{M}_\lambda$*;*

3. Bounds*: Every function in* $\mathscr{M}_\lambda$ *is bounded from below when evaluated in the identity matrix, and from above when evaluated in any of the rank 1 correlation matrices.*

Proof. The proof of the first property easily derives from the definition of the $\mathscr{M}_\lambda$ class itself.

In order to prove the second statement notice that if $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$, then $[\phi_1(\mathbf{C}_1), ..., \phi_k(\mathbf{C}_1)] \geq [\phi_1(\mathbf{C}_2), ..., \phi_k(\mathbf{C}_2)]$, where $\geq$ denotes the standard product order. The result follows by h being increasing in all its arguments.

The last statement simply follows from conditions C1 and C2 and Theorem 4.1. □

The first property tells us that, given ϕ and ψ in $\mathcal{M}_\lambda$, if $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$, then $\phi(\mathbf{C}_1) \geq \phi(\mathbf{C}_2)$ and $\psi(\mathbf{C}_1) \geq \psi(\mathbf{C}_2)$. This means that if two correlation matrices can be ordered according to majorization, all functionals in $\mathcal{M}_\lambda$ will show the same behavior, assigning a higher portfolio risk to the majorizing correlation matrix ($\mathbf{C}_1$), and a lower risk to the majorized one ($\mathbf{C}_2$).

The practical implications of such a property are evident: if it is possible to observe majorization on the market, all the risk measures belonging to $\mathcal{M}_\lambda$ will move in the same direction, making every discussion about which measure is better less relevant. One can verify that, when majorization is lost, functionals in $\mathcal{M}_\lambda$ can show different behavior, moving in different directions: one measure may see increasing risks, while another may indicate a decrease. Therefore, the consistent and inconsistent behavior of risk functionals is a way of identifying majorization (and vice versa). Empirical investigations, as the one we offer in Section 4.5, suggest that financial data commonly show majorization, especially during periods of crisis, when the phenomenon appears to be quite strong.

The second property simply extends the same behavior that Schur-convex functions exhibit for real vectors [19], and it thus allows to build new risk measures starting from existing ones.

Finally, the last property can be used to introduce a trivial standardization for the functions in $\mathcal{M}_\lambda$, that is

$$\bar{\phi}(\mathbf{X}) = \frac{\phi(\mathbf{X}) - \phi(\mathbf{I})}{\phi(\mathbf{J}) - \phi(\mathbf{I})} \tag{4.3}$$

where the $\mathbf{J}$ matrix is the special rank 1 matrix with $J_{i,j} = 1$ for every i, j.

In the next Subsections, we discuss some examples of matrix functions belonging to $\mathcal{M}_\lambda$. Some of these measures represent new tools for the analysis of financial data, which–we hope–other researchers will further test and develop.

4.3.1. THE QUANTUM LORENZ CURVE AND THE INEQUALITY FUNCTIONALS

An important set of functionals belonging to the $\mathcal{M}_\lambda$ class is given by the matrix representation of some famous measures of economic inequality [21, 22].

Mimicking the approach used in the study of economic size distributions and inequality [23], the starting point is the definition of a multivariate version of the well-known Lorenz curve [24].

The (univariate) Lorenz function was introduced to study the distribution of wealth among individuals, but it later developed into a powerful tool of statistical analysis [6].

Given a vector of ordered positive quantities $x_{[1]} \geq x_{[2]} \geq ... \geq x_{[n]}$, say incomes [24], the classical Lorenz curve is defined as

$$L_k(\mathbf{x}) = \frac{\sum_{i=1}^{k} x_{[i]}}{\sum_{i=1}^{n} x_i} \quad \text{for } k = \{1, ..., n\}. \tag{4.4}$$

This function has several useful properties, for which we refer to [6, 25]. For our purposes, it is sufficient to notice that if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n_+$ and $L_k(\mathbf{x}) \geq L_k(\mathbf{y})$ for every $k \in \{1, ..., n\}$, then $\mathbf{x} \succ \mathbf{y}$ and vice versa. For this reason the majorization relation between two vectors $\mathbf{x}$ and $\mathbf{y}$ is sometimes interpreted as $\mathbf{x}$ being more unequal than $\mathbf{y}$.

Several multivariate extensions of the Lorenz curve have been proposed in the literature to deal with multivariate inequality, and for a review we suggest [6, 23]. Here we introduce a brand new one, suitable for portfolio risk management and compatible with the quantum majorization of Definition 4.3. For this reason, we call it quantum Lorenz curve.

Definition 4.5. Let $\mathbf{C}$ be an $n \times n$ correlation matrix, its quantum Lorenz curve is the matrix function $\mathbf{L} : \mathscr{C} \to \mathbb{R}^n$, such that

$$L_k(\mathbf{C}) := \frac{\max_{\mathbf{U}\mathbf{U}^T = \mathbf{I_k}} Tr(\mathbf{U}\mathbf{C}\mathbf{U}^T)}{Tr(\mathbf{C})} = \frac{\sum_{i=1}^k \lambda_{[i]}(\mathbf{C})}{\sum_{i=1}^n \lambda_i(\mathbf{C})} \quad \text{for } k = \{1, ..., n\}, \tag{4.5}$$

where $\mathbf{U}$ is a $k \times n$ unitary invariant matrix, Tr is the trace operator, and $\lambda_{[1]} \geq \lambda_{[2]} \geq ... \geq \lambda_{[n]}$ are the ordered eigenvalues of C.

The definition can be easily extended[3] to any $n \times m$ matrix $\mathbf{C}$ that admits a singular value decomposition $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$ by setting the numerator in (4.5) equal to:

$$\max_{\mathbf{U}\mathbf{U}^\mathbf{T} = \mathbf{V}\mathbf{V}^\mathbf{T} = \mathbf{I_k}} Tr(\mathbf{U}\mathbf{A}\mathbf{V})$$

The quantum Lorenz has a nice interpretation in terms of portfolio risk: for a given k, it represents the percentage of the total portfolio variance explained by the first k eigenvalues [26]. Alternatively, in the context of dimensionality reduction, $1 - L_k(\mathbf{C})$ provides an estimate of the percentage of information that is lost by taking a rank k approximation of the matrix $\mathbf{C}$ [7].

Proposition 4.3. *A quantum Lorenz curve* $\mathbf{L}(\mathbf{C})$ *has the following properties:*

1. Unitary invariance*:* $\mathbf{L}(\mathbf{U}\mathbf{C}\mathbf{U}^T) = \mathbf{L}(\mathbf{C})$ *with* $\mathbf{U}$ *being a unitary matrix;*
2. Convexity and sub-additivity*:* $L_k(\alpha\mathbf{C}_1 + (1-\alpha)\mathbf{C}_2) \leq \alpha L_k(\mathbf{C}_1) + (1-\alpha)L_k(\mathbf{C}_2)$ *and* $L_k(\mathbf{C}_1 + \mathbf{C}_2) \leq L_k(\mathbf{C}_1) + L_k(\mathbf{C}_2)$ *for every* $k = \{1, ..., n\}$*;*
3. Perfect equality*:* $L_k(\mathbf{C}) = \frac{k}{n}$ *for every* $k = \{1, ..., n\}$ *if and only if* $\mathbf{C}$ *is the identity matrix.*
4. Perfect inequality*:* $L_k(\mathbf{C}) = 1$ *for every* $k = \{1, ..., n\}$ *if and only if* $\mathbf{C}$ *is a rank 1 matrix.*

Proof. Property 1 simply follows from basic properties of the trace operator: linearity and invariance to cyclic permutations.

Property 2 follows from the sub-additivity of the max and the fact that, for $n \times n$ correlation matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, $Tr(\mathbf{C}_1) = Tr(\mathbf{C}_2)$.

[3] This would be particularly useful in the study of economic inequality, together with many other properties of the quantum Lorenz, which nevertheless go beyond the scope of this chapter.

To prove the "if part" of Property 3, just recall that, for $\mathbf{U}$ a $k \times n$ unitary matrix, we have that $Tr(\mathbf{UIU}) = Tr(\mathbf{I_k}) = k$. The "only if" part once again follows from the Fan representation theorem [20], and it mimics the same reasoning used in the proof of Theorem 4.1.

Property 4 can be proven in a similar way, after noticing that the set of rank 1 matrices is writable as $\{\mathbf{C} \in \mathscr{C} : \mathbf{C} = \boldsymbol{\rho}\boldsymbol{\rho}^T, \boldsymbol{\rho} \in \{-1,1\}^n\}$. Therefore $Tr(\boldsymbol{\rho}\boldsymbol{\rho}^T) = n$, since $\rho_i \rho_i = 1$ for every $\boldsymbol{\rho} \in \{-1,1\}^n$, and $Tr(\mathbf{U}(\boldsymbol{\rho}\boldsymbol{\rho}^T)\mathbf{U^T}) = n$ for every $k \times n$ matrix $\mathbf{U}$. □

The following proposition identifies a close connection between quantum majorization and the quantum Lorenz, further justifying the choice of the name.

Proposition 4.4. *Given two correlation matrices $\mathbf{C}_1, \mathbf{C}_2$, we have that $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_2$ if and only if $\mathbf{L}(\mathbf{C}_1) \geq \mathbf{L}(\mathbf{C}_2)$, with $\geq$ being the standard product order for vectors.*

Proof. The proof is trivial if one notices that the spectral representation of the quantum Lorenz curve implies Definition 4.3. □

On the basis of the quantum Lorenz, we can finally introduce a class of inequality (risk) functionals.

Definition 4.6. Given a correlation matrix $\mathbf{C}$ and a real valued function $\psi : \mathbb{R}^n \to \mathbb{R}$ increasing in all its arguments, we define an inequality functional as

$$\mathscr{D}_\psi(\mathbf{C}) = \psi(\mathbf{L}(\mathbf{C})), \tag{4.6}$$

where $\mathbf{L}(\mathbf{C})$ is the quantum Lorenz curve associated to $\mathbf{C}$.

A consequence of Propositions 4.4 and 4.2 is that the set of inequality functionals belongs to the $\mathscr{M}_\lambda$ class. The following lemma gives us a way of constructing these functionals.

Lemma 4.5. *Any real-valued matrix functional built via a Schur-convex function ϕ applied to the spectrum of a correlation matrix $\mathbf{C}$, i.e. $\phi(\boldsymbol{\lambda}(\mathbf{C}))$, is an inequality functional in $\mathscr{M}_\lambda$.*

Proof. Consider $\Delta_k := L_k(\mathbf{C}) - L_{k-1}(\mathbf{C})$. The definition of the quantum Lorenz tells that $\Delta_k = \lambda_{[k]}/n$.

Given the properties of Schur-convex functions [19], the application of an increasing function to $L(\mathbf{C})$ is equivalent to the application of a Schur-convex function to Δ_k. Now, set $D_\phi(\mathbf{C}) = \phi(\boldsymbol{\lambda}(\mathbf{C}))$, where ϕ is a Schur-convex function mapping $\boldsymbol{\lambda}(\mathbf{C})$ onto the reals. The use of Equation (4.5) concludes the proof. □

Lemma 4.5 generalizes a result by Alberti and Uhlmann [8], who proved the isotonicity to quantum majorization only for symmetric convex functions of the spectra of Hermitian matrices, clearly a subset of the larger Schur-convex class [19].

Let us now consider some examples of inequality functionals obtained via Definition 4.6. These functionals generalize some of the most famous socio-economic inequality indices [25, 27].

Example 4.1 (Matrix distance indices)**.** Let $\psi = ||.||_p$ be the L^p vector distance between the quantum Lorenz curve $\mathbf{L}(\mathbf{C})$ of the correlation matrix $\mathbf{C}$ and its lower bound $\mathbf{L}(\mathbf{I})$. Then

$$\mathscr{D}_{||.||_p}(C) = ||\mathbf{L}(\mathbf{C}) - \mathbf{L}(\mathbf{I})||_p \tag{4.7}$$

is the matrix extension of the univariate distance indices of [25].

For $p = 1$, we obtain the (correlation) matrix version of the Gini index [28], i.e.

$$\mathscr{D}_{||.||_1}(\mathbf{C}) = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{k}|\lambda_{[i]} - k|. \tag{4.8}$$

For $p = \infty$, we get the matrix version of the Pietra index, or [29]

$$\mathscr{D}_{||.||_\infty}(\mathbf{C}) = \max_k |L_k(\mathbf{C}) - L_k(\mathbf{I})|. \tag{4.9}$$

Example 4.2 (Ratio indices)**.** If $\psi = \pi_k$, where π_k is the projection operator onto the k-th coordinate, we have

$$\mathscr{D}_{\pi_k}(\mathbf{C}) = L_k(\mathbf{C}). \tag{4.10}$$

This represents the matrix extension of the top-α% inequality index used in the social sciences to measure the proportion of the total wealth owned by the top α% richest individuals [21] in the economy. A particular case of ratio index, called absorption ratio, has already been used as portfolio risk measure for correlation matrices in [30, 31]. It is obtained by setting $k = \lfloor 0.05 \times n \rfloor$.

4.3.2. ENTROPY-BASED FUNCTIONALS

Following the seminal work of Alberti and Uhlmann [8], a large class of real-valued matrix functions for Hermitian matrices, called entropy-like functionals, was introduced in the physical literature to study the behaviour of Gibbs densities [9, 10].

Since correlation matrices are trivially Hermitian, these functionals can be also applied in our portfolio risk framework, and they naturally fall in the $\mathscr{M}_\lambda$ class.

Definition 4.7 (Entropy-like functionals)**.** Let $\mathbf{H}$ be an Hermitian matrix and $f : \mathbb{R} \to \mathbb{R}$ a convex function, the class of entropy-like matrix functional S_f is defined as

$$S_f(\mathbf{H}) = Tr(f(\mathbf{H})), \tag{4.11}$$

where $f(\mathbf{H})$ is the usual notation for matrix functions, i.e.

$$f(\mathbf{H}) := \mathbf{U}\begin{bmatrix} f(\lambda_1(\mathbf{H})) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f(\lambda_n(\mathbf{H})) \end{bmatrix}\mathbf{U}^T,$$

with $\mathbf{U}$ being a unitary matrix that diagonalizes $\mathbf{H}$.

It should be clear that, for a Schur-convex function ϕ, one has $\phi(\boldsymbol{\lambda}(\mathbf{H})) = \sum_{i=1}^{n} f(\lambda_i(\mathbf{H}))$. As a consequence, Lemma 4.5 ensures that S_f is monotone with respect to quantum majorization.

An useful property of entropy-like functionals is their convexity, that is

$$S_f(\alpha \mathbf{C_1} + (1-\alpha)\mathbf{C_2}) \leq \alpha S_f(\mathbf{C_1}) + (1-\alpha) S_f(\mathbf{C_2}),$$

where $\alpha \in [0,1]$ and $\mathbf{C_1}$, $\mathbf{C_2}$ are $n \times n$ correlation matrices. This further justifies their possible use in evaluating portfolio risk.

Example 4.3 (Von Neumann entropy)**.** By taking $f(\lambda_i) = -\lambda_i \log(\lambda_i)$ in (4.11), we obtain the Von Neumann entropy [32] of a correlation matrix,

$$S_{VN}(\mathbf{H}) := \sum_{i=1}^{n} \lambda_i(\mathbf{H}) \log \lambda_i(\mathbf{H}). \tag{4.12}$$

The function f is trivially concave, therefore it is sufficient to define S_{-f} to preserve the $\overset{\lambda}{\succ}$ order.

Example 4.4 (Renyi α-entropies and effective rank)**.** This class of entropies à la Renyi [8, 23] is obtained as a generalization of S_f, by taking the function $g(S_f) := \frac{1}{1-\alpha} \log S_f$ and by setting $f(x) = x^\alpha$, $\alpha \geq 0$. One has

$$S^\alpha(\mathbf{H}) := \frac{1}{1-\alpha} \log Tr(\mathbf{H}^\alpha) = \frac{1}{1-\alpha} \log \sum_{i=1}^{n} \lambda_i(\mathbf{H})^\alpha, \tag{4.13}$$

where $\mathbf{H}^\alpha$ is the matrix power function.

This class is once again concave, hence one defines $-S^\alpha$ to ensure convexity. To verify the $\overset{\lambda}{\succ}$ monotonicity, just observe that g is an increasing function and apply Proposition 4.2, Property 2, with $k = 1$.

By defining $g(S_f) := e^{-S_f}$ with $f(\lambda_i) = \lambda_i \log(\lambda_i)$, one obtains another entropy-based index commonly called effective rank [33], that is

$$ER(\mathbf{H}) := e^{-Tr(f(\mathbf{H}))} = e^{-\sum_{i=1}^{n} \lambda_i(\mathbf{H}) \log \lambda_i(\mathbf{H})}$$

This measure was developed as an attempt to propose a real-valued proxy for the rank of a correlation matrix. Clearly $-ER$ belongs to the $\mathscr{M}_\lambda$ class.

4.3.3. OTHER QUANTUM MAJORIZATION PRESERVING FUNCTIONALS

Many other functionals related to quantum majorization can be identified and proposed. Here we only cite two additional possibilities: determinant-based measures and matrix norms.

Determinant-based measures originate from the works of Wilks [26], and use the determinant of the covariance matrix as a summary tool for the dispersion of an n-dimensional random vector. More recently, the determinant of the correlation matrix has also been taken into consideration [34]. It is not difficult to show that all these measures belong to the $\mathscr{M}_\lambda$ class.

Matrix norms were introduced by Von Neumann [32, 35] as the set of all norms $||.||$ such that, given a matrix $\mathbf{C}$, one has $||\mathbf{C}|| = ||\mathbf{UCU}^T||$, with $\mathbf{U}$ being a unitary matrix. A notable example is the Frobenius norm $||\mathbf{C}||_F = \sqrt{\sum_{i=1}^{n} \lambda_i^2(\mathbf{C})}$ [35, 36], often used in portfolio optimization and management [37].

Mirsky [38] showed that there is a one-to-one correspondence between matrix norms and a special subclass of Schur-convex functions–symmetric gauge functions–on the spectrum of the matrix. Thanks to this connection, Lemma 4.5 guarantees that matrix norms belong to the $\mathscr{M}_\lambda$ class.

4.4. THE QUANTUM MAJORIZATION MATRIX

Let $\mathbf{C}_t$ be the $n \times n$ correlation matrix of an portfolio $\mathscr{P}$ with n assets, which is observed over time[4], for $t = 1, ..., T$.

Our aim is to introduce a concise way of representing the majorization among all the $\mathbf{C}_t$ matrices as time passes. In fact, if during a certain period the correlation matrix $\mathbf{C}_t$ majorizes the previous one, $\mathbf{C}_{t-1}$, we can read this as an relative[5] increase of portfolio risk (Section 4.5 shows that this actually happens). And, in general, if $\mathbf{C}_t \overset{\lambda}{\succ} \mathbf{C}_{t-s}$, for some s, then the portfolio at time t is riskier than what it was at time $t-s$. Further, if $\mathbf{C}_t \overset{\lambda}{\succ} \mathbf{C}_{t-1} \overset{\lambda}{\succ} ... \overset{\lambda}{\succ} \mathbf{C}_{t-s}$, with no interruption in the majorization series, risk has increased for s periods, and so on. Naturally, on the opposite side, if $\mathbf{C}_t \overset{\lambda}{\prec} \mathbf{C}_{t-1}$ risk is decreasing. It is important to notice that, however, if $\mathbf{C}_t \overset{\lambda}{\not\succ} \mathbf{C}_{t-1}$, it is not possible to state that risk has either increased or decreased with confidence.

Following the results of Section 4.3, one could be inclined to choose a measure $\phi \in \mathscr{M}_\lambda$ to evaluate risk on every correlation matrix $\mathbf{C}_t$, for $t = 1, ..., T$, thus creating a sequence $(\phi(\mathbf{C}_t))_{t=1}^T$ to assess portfolio risk over time. However, given Propositions 4.1 and 4.2, it seems more natural to deal with majorization directly, given that: 1) it represents a strong connection among correlations; and 2) when it manifests itself, all functionals in $\mathscr{M}_\lambda$ are comonotonic and give the same type of information about changes in portfolio risk.

The tool we propose is the *quantum majorization matrix*, that is a matrix of indicators, where each element points out the majorization relation between two different correlation matrices.

Definition 4.8. Consider a collection of correlation matrices $\{\mathbf{C}_t\}_{t=1}^T$ for a portfolio $\mathscr{P}$. Let $\mathbf{A}$ be a $T \times T$ matrix such that

$$A_{i,j} = \begin{cases} 1 & \text{if } \mathbf{C}_i \overset{\lambda}{\succ} \mathbf{C}_j \\ 0 & \text{otherwise} \end{cases}, \tag{4.14}$$

for $i, j = 0, 1, ..., T$.

The matrix $\mathbf{A}$ is called quantum majorization matrix.

By definition the quantum majorization matrix keeps track of all the $\overset{\lambda}{\succ}$ ordering relations in the collection $\{\mathbf{C}_t\}_{t=1}^T$. Recalling Proposition 4.1, $\mathbf{A}$ thus helps in noticing the presence of a non-trivial dependence structure in the portfolio.

Fix a row i in $\mathbf{A}$. The majorized (or dominated) set $\mathscr{D}_i := \{\mathbf{C}_k : A_{i,k} = 1, \forall k \in \{1, ..., T\}\}$ corresponds to the set of the all correlation matrices that are majorized by $\mathbf{C}_i$, being less

[4] In view of Section 4.5 we will compare correlation matrices over time, but–as said–the quantum majorization approach can naturally be applied cross-sectionally, to compare and rank portfolios in terms of risk.

[5] I this case we use to world relative since the quantum majorization represents an ordering relation between objects and not an absolute reference. In other words one won't be able to asses risk using quantum majorization over a single correlation matrix. That could still be achieved by using functions in the $\mathscr{M}_\lambda$ class.

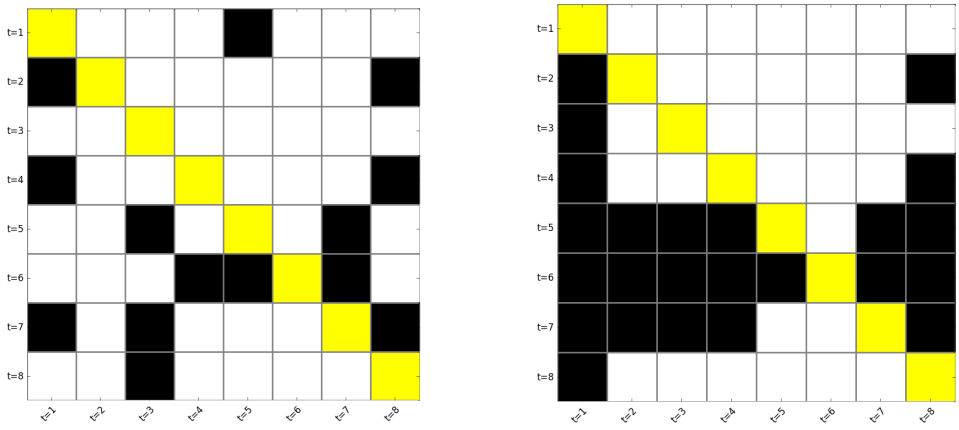
risky than $\mathbf{C}_i$. The majorizing (or dominating) set $\mathscr{ND}_i := \{\mathbf{C}_k : A_{k,i} = 1, \forall k \in \{1,...,T\}\}$, conversely, contains all the correlation matrices that majorize $\mathbf{C}_i$, and are thus riskier. As a consequence, $\mathscr{S}_i := \mathscr{D}_i \cap \mathscr{ND}_i$ is the set of all the correlation matrices that are similar to $\mathbf{C}_i$, while $\mathscr{N}_i := \mathscr{D}_i^c \cap \mathscr{ND}_i^c$ is the set of all the correlation matrices that have no ordering relation with $\mathbf{C}_i$.

From now on, without any loss of generality[6], we impose not to have similar correlation matrices. Therefore, the set $\mathscr{S}_i$ only contains $\mathbf{C}_i$, and the total number of majorizations in $\mathbf{A}$ lies in the interval $\left[0, \frac{T(T+1)}{2}\right]$. Also observe that

$$\sum_{k=1}^{T} (A_{i,k} + A_{k,i}) - 1 \leq T. \tag{4.15}$$

4

Figure 4.1 gives two graphical representations of possible 8×8 quantum majorization matrices. To represent the different $A_{i,j}$'s in the matrices we choose black for 1, and white for 0. Yellow is used to indicate the trivial diagonal of each correlation matrix majorizing and being majorized by itself.



(a) Case 1 (b) Case 2

Figure 4.1: Two examples of 8×8 quantum majorization matrices. Black squares represent ones, white squares zeros, and the yellow diagonal shows the trivial situation of a correlation matrix majorizing itself.

Consider Subfigure 4.1a: cell (t_2, t_1) is black, indicating that $A_{2,1} = 1$. This means that the correlation matrix $\mathbf{C}_2$ majorizes $\mathbf{C}_1$, i.e. $\mathbf{C}_2 \overset{\lambda}{\succ} \mathbf{C}_1$: the embedded portfolio risk has thus increased at time t_2 relative to t_1. As expected cell (t_1, t_2) is white $(A_{1,2} = 0)$.

Let us now look at (t_5, t_1): it is white, while the symmetric cell (t_1, t_5) is black. In this case, we have $\mathbf{C}_1 \overset{\lambda}{\succ} \mathbf{C}_5$, i.e. $\mathbf{C}_5$ is majorized.

Take now (t_3, t_1): it is white, but also (t_1, t_3) is. Therefore, between $\mathbf{C}_1$ and $\mathbf{C}_3$ there is no majorization, $\mathbf{C}_1 \overset{\lambda}{\not\succ} \mathbf{C}_3$, and no unique comment on risk can be made.

[6] In applications it is not plausible to observe two correlation matrices with the same exact spectrum.

All in all, if one of the two symmetric cells is black, majorization takes place[7], if both cells are white, no majorization can be observed.

Consider now Subfigure 4.1b, and let us focus our attention on the correlation matrix $\mathbf{C}_6$, represented by row t_6. It is clear that this correlation matrix majorizes all the others, thus representing the up-to-date maximum in terms of risk (as time passes, the ranking could change). Matrix $\mathbf{C}_5$ represents the second riskiest portfolio (up-to-date), which majorizes all other portfolios apart from the one represented by $\mathbf{C}_6$, by which it is majorized.

Looking at patterns in a quantum majorization matrix we can therefore identify periods of higher and lower risk, and we can compare them. As we shall see in Section 4.5, the presence of large black areas in the plot of a quantum majorization matrix will indicate times of financial distress for our portfolio. When a portfolio is granular and representative of a market, majorization can be used to look for and analyze increases in systemic risk and the emergence of financial crises.

It is important to stress out that the risk valuation obtained by studying the quantum majorization matrix $\mathbf{A}$ either by visual inspection or by the methodologies explained in the next Sections are can only lead to a relative risk comparison between correlation matrices within the dataset. The external validity of our method can be claimed if additionally assumptions are made, for example under weak alpha-mixing conditions of the underlying price processes and/or when a sufficient amount of correlation matrices are included in the dataset. Finally, we suggest, as standard practice in non-parametric data analysis and fiance, to use re-calibration once a sufficient number of time has passed and new data are available.

4.4.1. TWO SIMPLE RISK MEASURES ON THE QUANTUM MAJORIZATION MATRIX

In applications, when T is large, it is common to obtain datasets with a very big number of correlation matrices. This makes it cumbersome to just explore the associated quantum majorization matrix graphically, as we did a few lines above[8]. A natural solution is to introduce some summary measures based on the majoritazion matrix $\mathbf{A}$.

A simple tool, which proves very useful in applications (see Section 4.5), is the function $\theta : \mathbf{C}_i \to \mathbb{R}$, defined as

$$\theta(\mathbf{C}_i) = \frac{1}{2} + \frac{\#\mathscr{D}_i}{2T} - \frac{\#\mathscr{N}\mathscr{D}_i}{2T}, \tag{4.16}$$

where # indicates the cardinality (i.e. the number of elements) of the set.

The counting measure in Equation (4.16) expresses the risk embedded in each correlation matrix as a function of its majorized and majorizing sets. Interestingly, $\theta(\mathbf{C}_i)$ is also order preserving with respect to quantum majorization, so that if $\mathbf{C_i} \overset{1}{\succ} \mathbf{C_j}$ then $\theta(\mathbf{C_i}) \geq \theta(\mathbf{C_j})$.

It is easy to observe that $\theta(\mathbf{C}_i) \in [0,1]$. The case $\theta(\mathbf{C}_i) = 0$ indicates that the correlation $\mathbf{C}_i$ is majorized by all the other matrices in the collection, while $\theta(\mathbf{C}_i) = 1$ tells us that $\mathbf{C}_i$

[7] Recall that we have excluded similarity among correlation matrices, otherwise we could have two corresponding black cells.

[8] Yet, as discussed later in Section 4.5, we always suggest to plot the quantum majorization matrix, to have a quick heuristic idea of portfolio risk.

is the up-to-date maximum in terms of risk.

Given its definition, the value $\frac{1}{2}$ represents for $\theta(\mathbf{C}_i)$ a threshold. If $\theta(\mathbf{C}_i) \leq \frac{1}{2}$, the correlation matrix $\mathbf{C}_i$ can be classified as non risky with respect to the great part of the remaining matrices, while the opposite happens for $\theta(\mathbf{C}_i) > \frac{1}{2}$.

From the quantum majorization matrix $\mathbf{A}$ we can also obtain a quantity that bridges towards Proposition 4.2, Property 1.

Recall that every entry $A_{i,j} = 1$ in $\mathbf{A}$ marks the presence on ordering relation between correlation matrices $\mathbf{C}_i$ and $\mathbf{C}_j$. Therefore, in order to obtain a numerical measure of the ordering relations inside a given dataset it is sufficient to count all the entries $A_{i,j} = 1$ inside the matrix and multiply this quantity by the constant $k = 2$. The role of k is to make sure that the outcome of the counting procedure is the total number of ordering relations regardless the direction. Finally, in order to obtain a comparable measure we scale it by the total possible number of ordering relations in $\mathbf{A}$ $T(T-1)$.

We have build the following quantity denoted by U:

$$U = \frac{2\sum\sum_{i\neq j} A_{i,j}}{T(T-1)} \tag{4.17}$$

U is an index telling us how much majorization takes place in a collection of correlation matrices. The lower the index the more often majorization occurs, therefore the more often all functionals in the $\mathscr{M}_\lambda$ class are comonotonic. The higher U the more we can expect the measures in $\mathscr{M}_\lambda$ to behave in a not necessarily coherent way.

If we come back to the two examples in Figure 4.1, we can compute $U = 0.5$ in the first case, and $U = 0.14$ in the second. This is consistent with what we have discussed before, given that the second case visually shows a higher level of majorization.

4.5. AN EXAMPLE ON ACTUAL DATA

Our dataset–available for download–consists of 4563 daily log-returns for 28 components of the Industrial Dow Jones (INDJ), between January 03 2000 and February 21 2018. Using 100-day moving windows, with a 90-day overlap, we have constructed a series of 447 correlation matrices, which we will compare and study using majorization. It is worth underlying that changing the size of the moving windows and of the overlaps does not affect our conclusions, until it is possible to play with a sufficient number of correlation matrices (≥ 50). The combination here shown is the one providing the best results from a graphical point of view. However, in the Appendix we provide some more additional examples with different overlaps as-well-as a numerical evidence of the presence of majorization structure in the data with respect to an i.i.d draw form a uniform random correlation matrix.

For each correlation matrix in the collection, we have computed the associated eigenvalues via a standard singular value decomposition.

For each 100-day moving window, Figure 4.2 shows the average log-returns of the INDJ index; superimposed we also provide the time series of the corresponding matrix Gini index ($\mathscr{D}_{||.||_1}$), computed on the different correlation matrices. The chosen measure seems definitely capable of identifying some major events in the observation period, something not always possible by just looking at the log-returns.

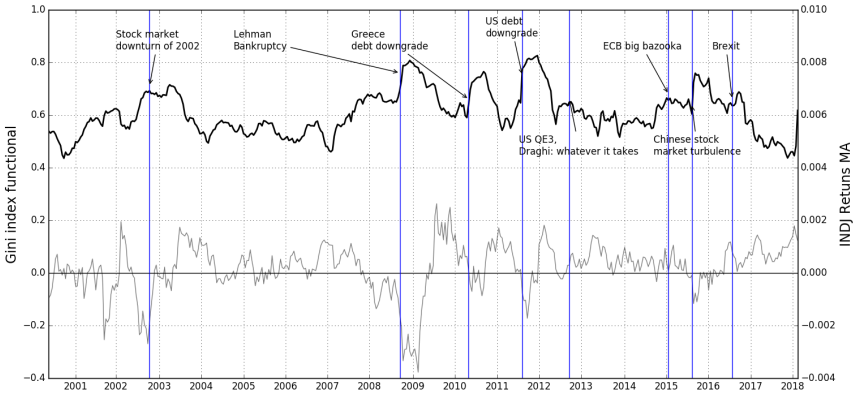


Figure 4.2: The average INDJ index (gray) in the period between January 03 2000 and February 21 2018, over rolling windows of 100 days, with a 90-day overlap. For the same windows we provide the matrix Gini index (black) computed on the corresponding correlation matrices. Major events in the observation period are indicated.

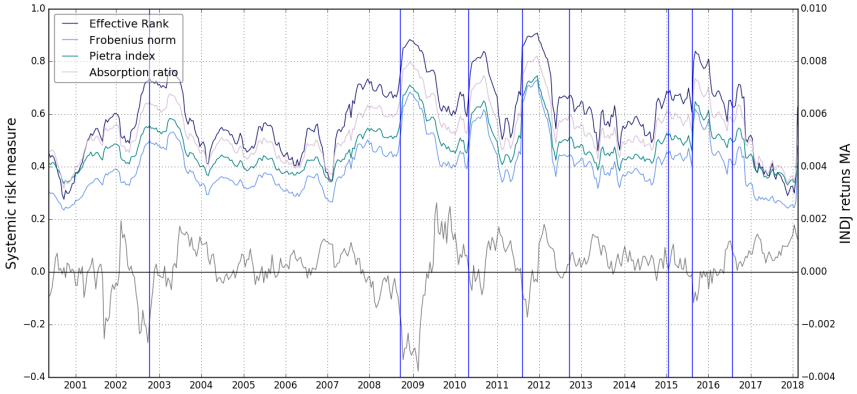


Figure 4.3: Time series of four $\overset{\lambda}{\succ}$ monotonic measures applied to the dataset.

The choice of the matrix Gini in Figure 4.2 is arbitrary, that is why in Figure 4.3 we also provide other possible measures of risk in $\mathscr{M}_\lambda$: the Frobenius norm, the effective rank, the absorption ratio and the Pietra index. Evidently, all these measures tend to behave similarly, suggesting the presence of several majorizations in the portfolio evolution, consistently with Proposition 4.2, Property 1.

A measure of the amount of majorization in the portfolio is provided by the U measure of Equation (4.17): the closer U to zero, the more often majorization occurs. The average U in our data is 0.15, with a standard deviation of 0.03; this indicates that quantum majorization is a rather common phenomenon in the INDJ portfolio. Therefore we can expect the different risk measures to behave similarly quite often, as per Figure 4.3.

Figure 4.4 shows the behavior of U over 100-day rolling windows. Interestingly, during the riskiest moments of the last twenty years, like the 2007-2009 crisis or the second wave of 2011-2012, the U index decreases down to 0.1, indicating an increase in majorization–and thus risk–on the market. Conversely, during the periods of good market conditions, like 2006-2007 majorization decreases, and U grows up to 0.3.

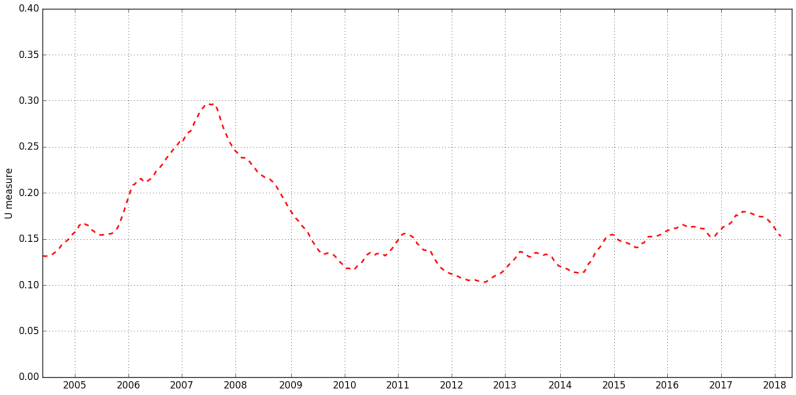


Figure 4.4: Time evolution of the U measure over a 100-day rolling window.

Let us now have a look at Figure 4.5, which presents the quantum majorization matrix $\mathbf{A}$ of the INDJ sample. An interesting fact, consistent with the behavior of U, can be observed: the correlation matrices of the 2007-2009 crisis period are the up-to-date maximal elements for the dataset. In fact, as clear from the red band in the picture, they majorize most of the other correlation matrices, with just a few exceptions of non-majorization ($\not\prec$). Conversely, the correlation matrices of the years 2005, 2006 or 2014 are majorized by almost all the other correlations, representing the least risky set.

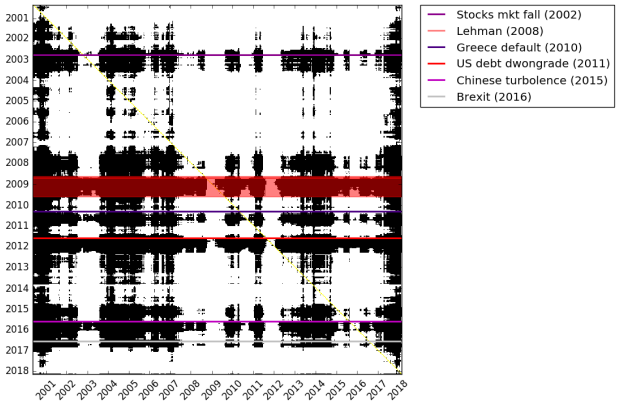


Figure 4.5: Quantum majorization matrix of the INDJ dataset. Black squares represent quantum majorizations between correlation matrices, the red lines indicate major market events during the observation period, while the red band underlines the riskiest period (in term of majorizations) in the dataset.

In Figure 4.6, we provide the behavior of another risk measure we have discussed in Subsection 4.4.1, the quantity $\theta(\mathbf{C}_i)$, together with the average log-returns of the INDJ portfolio. For every correlation matrix $\{\mathbf{C}_t\}_{t=1}^{447}$ the corresponding $\theta(\mathbf{C}_t)$ is computed. Looking at the graph, we can see that all periods of crisis (see again Figure 4.2) are characterized by riskier correlation matrices, i.e. $\theta(\mathbf{C}_t) > 0.5$, while less turbolent periods do show $\theta(\mathbf{C}_t) < 0.5$. Once again the maximal levels of majorization are observed in 2007-2009 and 2011-2012, where θ almost reaches 1.

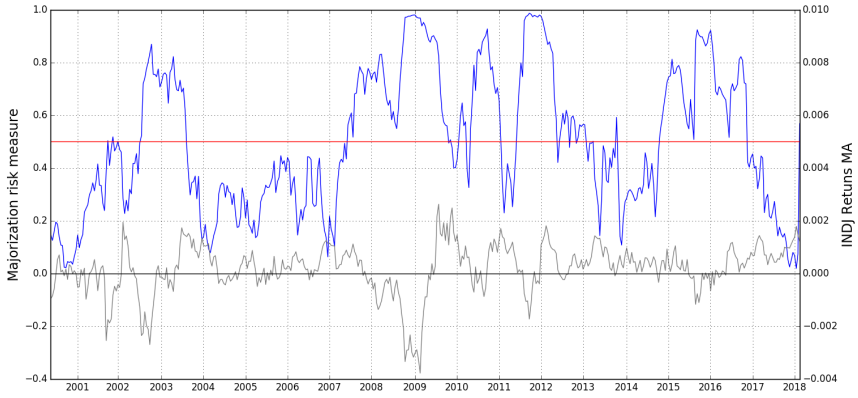


Figure 4.6: Time evolution of the risk measure $\theta(\mathbf{C}_i)$ for $\{\mathbf{C}_t\}_{t=1}^{T}$, $T = 447$, with its 0.5 threshold level. The INDJ average log-returns are also provided for readers' convenience.

On the basis of the quantum majorization matrix of Figure 4.5, we can also give the intuition of a simple alarm system for portfolio risk. The idea is to cluster the correlation matrices $\{\mathbf{C}_t\}_{t=1}^{447}$ according to the information embedded in the matrix $\mathbf{A}$, so to group together the correlations with similar majorization patters. For the collected matrices we can then compute the corresponding θ index, and verify whether particular values manifest themselves.

As common in cluster analysis [7], to group observations we need some notion of distance among the majorization patterns in $\mathbf{A}$, as induced by each $\mathbf{C_t}$. A standard way to deal with this type of problems is spectral embedding [39], which associates each $\mathbf{C_t}$–that for us is a data point in the space of correlation matrices–with a position in a multidimensional space, whose dimension and coordinates are defined by the left and right eigenvectors of $\mathbf{A}$.

Figure 4.7 shows the position of each correlation matrix in a 2-dimensional space, where the x-coordinates are given by the first entries of the left eigenvectors of $\mathbf{A}$, and the y-coordinates by the first entries of the right eigenvectors[9].

[9] Plots in higher dimensions do not add much information, that is why we restrict our attention to the 2d case.

The identification of clusters can then be performed using several different methods; here we rely on the standard k-means algorithm [7], with the goal of identifying 2 and 3 possible clusters (a larger number of clusters becomes difficult to interpret in this framework). The obtained groups are visible in the subplots of Figure 4.7, and their separation is definitely good. In the 2-cluster case of Figure 4.7, the red dots on the right identify the risky correlation matrices, those which tend to majorize the others, while the white dots represent the majorized correlations, associated to the less risky periods. In the 3-cluster situation, the yellow group appears, taking elements from the two original clusters. This group represents a situation of intermediate risk, possibly corresponding to a transition between more and less risky periods.

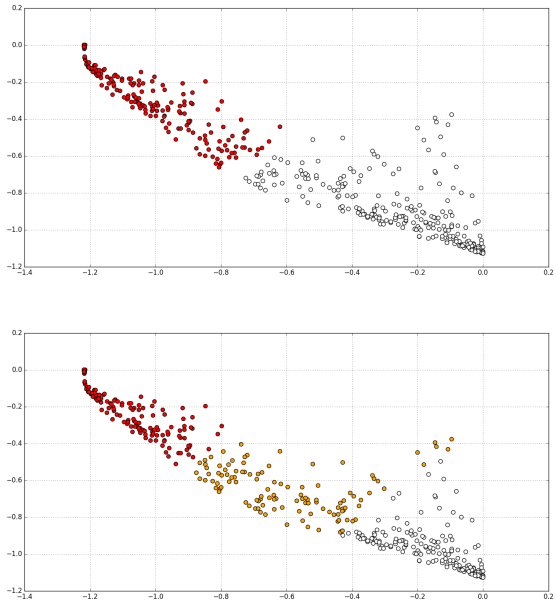
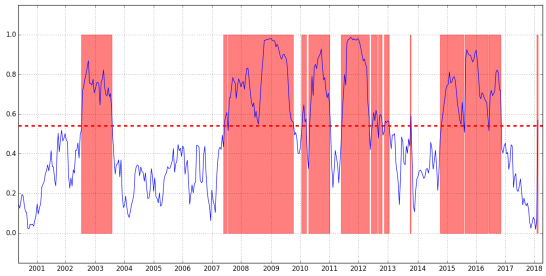


Figure 4.7: Projection of the INDJ dataset on a 2D Euclidean plane, on the basis of the eigendecomposition of **A**. The separation into 2 (top) and 3 (bottom) clusters via the k-means algorithm is evident.

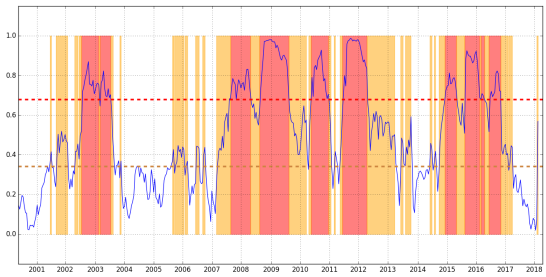
In Figure 4.8 the clusters of Figure 4.7 are superimposed to the time series of the $\theta(\mathbf{C}_t)$ risk index also given in Figure 4.6.

In Subfigure 4.8a, the red areas clearly indicate the majorizing correlation matrices (the red dots in Figure 4.7). The empirical threshold for $\theta(\mathbf{C}_t)$ that distinguishes majorizing from majorized matrices is 0.54. This value is close to the theoretical one (0.5) we expect from Equation (4.16). The periods of high portfolio risk are quite evident, and consistent with the recent economic history.

In the 3-cluster case of Subfigure 4.8b, the yellow areas show those correlation matrices that correspond to transitions between the majorizing/riskier (red) and the majorized/less risky (white) situations. In the INDJ dataset, a correlation matrix belongs to the yellow cluster if its θ falls in the interval $[0.34, 0.68]$.



(a) 2-cluster case, discriminating threshold at 0.54.



(b) 3-cluster case, discriminating thresholds at 0.34 and 0.68.

Figure 4.8

Subfigure 4.8b can thus represent a first example of alarm system based on quantum majorization. If we monitor the θ index over time, the closer (from below) the value to the 0.68 threshold, the higher the chance of entering into a risky period of strong quantum majorization. The lower θ, the smaller our portfolio risk.

As in all alarm systems [40], false alarms represent a problem, which requires serious treatment. For example, in Subfigure 4.8b, around year 2013, we see that θ oscillates a lot in the yellow area. This could possibly generate a number of false alarms regarding a possible increase in quantum majorization, which we should treat.

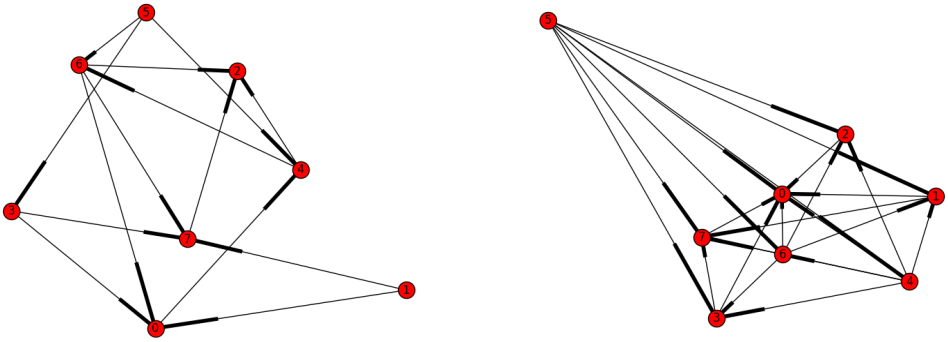
The development of a full alarm system based on quantum majorization goes beyond the scope of the present chapter, but it will surely be object of future research. A fascinating possibility could be to build a urn-based alarm system, similar to the one introduced in [41].

4.6. A NEW INSIGHT

In studying quantum majorization we have noticed an important connection with network analysis, something that, in our knowledge, has never been considered before. The fascinating consequence of this liaison is that it could facilitate the graphical exploration of portfolio risk.

Just notice that a quantum majorization matrix $\mathbf{A}$ can be seen as the adjacency matrix of a directed graph [42]. If the non-similarity assumption holds, then $\bar{\mathbf{A}} = \mathbf{A} - \mathbf{I}$ is the adjacency matrix of a directed acyclical graph.

We call the graph associated to $\mathbf{A}$ majorization graph, and the space of all $n \times n$ correlation matrices $\mathscr{C}$ is now the vertex space containing all the possible vertices the graph can have. The collection $\{\mathbf{C}_t\}_{t=1}^{T}$ represents the set of vertices that actually compose the graph in reality, and the set of directed edges among the vertices is given by the corresponding quantum majorization matrix. If $\mathbf{C_i} \overset{\lambda}{\succ} \mathbf{C_j}$, there is a directed edge from $\mathbf{C_i}$ to $\mathbf{C_j}$, while for $\mathbf{C_i} \prec^{\lambda} \mathbf{C_j}$ the directed edge goes from $\mathbf{C_j}$ to $\mathbf{C_i}$. In case of no majorization, $\mathbf{C_i} \not\succ \mathbf{C_j}$, no edge is present. This makes the majorization graph a non-trivial lattice[10] [42].



(a) Case 1

(b) Case 2

Figure 4.9: Examples of graphs associated to the quantum majorization matrices of Figure 4.1. Thick lines represent the direction of the edges between the nodes, and can ve read as arrows.

Figure 4.9 shows an example of graphs associated to the two quantum majorization matrices presented in Figure 4.1.

A majorization graph can be thought as a special case of directed inhomogeneous graph, in which the connection probability is different for each vertex, and it depends on the size of the permutraedon spanned by the spectrum of the correlation matrix associated to that vertex [43].

Thanks to the graph representation of majorization, one can immediately observe some interesting connections with what we have developed in this chapter. For example, the risk measure $\theta(\mathbf{C_i})$ corresponds to the difference between the out-degree and the in-degree centralities of the vertex associated to $\mathbf{C_i}$, while measure U is nothing but the density of the graph.

A majorization graph could also be used to build and calibrate a stochastic model to predict the dynamics of a correlation matrix, for instance by exploiting the approach of [42].

Another active field of research in networks theory is the detection of communities

[10] Note that, unlike most direct acyclical graphs in the literature[42], the flow to and from the vertices in the graph does not depend on their time index, but rather on their ordering.

[44]. In a majorization graph, communities can be understood as nodes corresponding to correlation matrices with a similar behaviour. Their individuation could therefore lead to results compatible with those discussed in Section 4.5, also extending them towards the development of a taxonomy of communities for crisis detection.

In conclusion, we believe that the link between quantum majorization and random graphs represents a fruitful line of future research, which we are willing to explore.

APPENDIX

In this Section we show how the structure of the majorization matrix **A** changes when we do not take overlapping intervals. In addition, we show evidence that the patterns observed in **A** are meaningful. In order to do that we sample uniformly at random from the space of correlation matrices sampling first the eigenvalues matrix from the space of diagonal matrices with entries chosen from a Dirichlet distribution and then multiplying it with the eigenvector matrices sampled from unitary group equipped with the Haar measure.

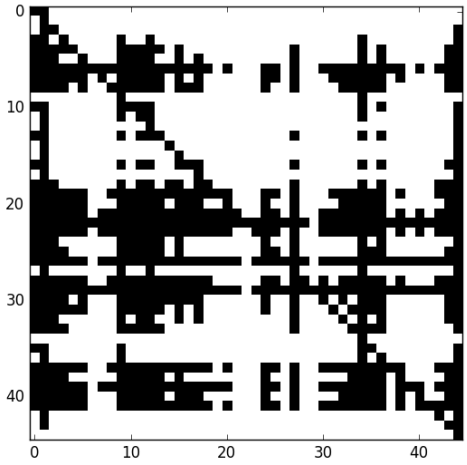


Figure 4.10: Quantum majorization matrix of the INDJ dataset. Date has been removed in order to ease the comparison with the artificial ones.

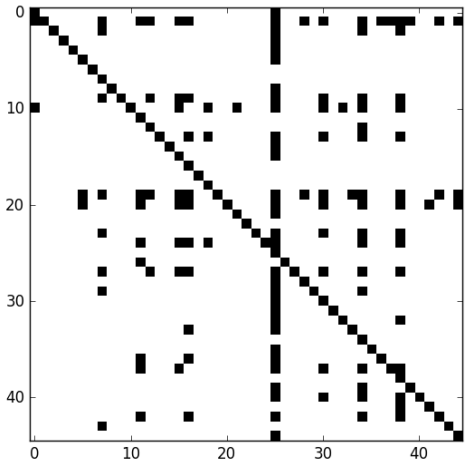


Figure 4.11: A quantum majorization matrix obtained by sampling i.i.d uniformly the space of correlation matrices with eigenvalues Dirichelt distributed.

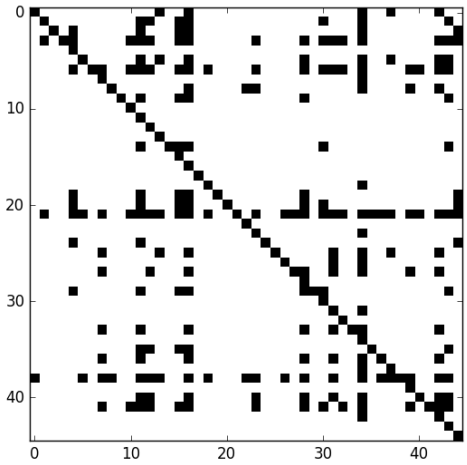


Figure 4.12: A quantum majorization matrix obtained by sampling i.i.d uniformly the space of correlation matrices with eigenvalues Dirichelt distributed.

As we can see from the above pictures it seems that when compared to an i.i.d draw from the space of correlation matrices the market majorization matrix exhibits a higher number of ordered correlation matrices as well as higher tendency to cluster.

REFERENCES

[1] A. Fontanari, I. Eliazar, P. Cirillo, and C. W. Oosterlee, *Portfolio risk and the quantum majorization of correlation matrices,* Available at SSRN 3309585 (2019).

[2] E. Barucci, *Financial markets theory : equilibrium, efficiency and information.* (Springer, 2003).

[3] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling extremal events: for insurance and finance*, Vol. 33 (Springer Science & Business Media, 2003).

[4] M. G. Kendall, *A new measure of rank correlation,* Biometrika **30**, 81 (1938).

[5] E. Schechtman and S. Yitzhaki, *On the proper bounds of the gini correlation,* Economics letters **63**, 133 (1999).

[6] S. Yitzhaki and E. Schechtman, *The Gini Methodology* (Springer, 2013).

[7] R. A. Johnson and D. M. Wichern, *Applied Multivariate Statistical Analysis 6th edition* (Pearson, 2007).

[8] P. M. Alberti and A. Uhlmann, *Stochasticity and partial order* (Deutscher Verlag der Wissenschaften, 1982).

[9] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, Cambridge, 2000).

[10] M. A. Nielsen, *Probability distributions consistent with a mixed state,* Physical Review A **62**, 052308 (2000).

[11] H. Bühlmann, *Mathematical Methods in Risk Theory*, Grundlehren der mathematischen Wissenschaften (Springer Berlin Heidelberg, 2005).

[12] T. Ando, *Majorizations and inequalities in matrix theory,* Linear algebra and its Applications **199**, 17 (1994).

[13] A. Giovagnoli and M. Romanazzi, *A group majorization ordeting for correlation matrices,* Linear Algebra and its Applications **127**, 139 (1990).

[14] A. Giovagnoli and H. P. Wynn, *G-majorization with applications to matrix orderings,* Linear algebra and its applications **67**, 111 (1985).

[15] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices* (Springer, 2009).

[16] L. A. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices* (American Mathematical Society, 2011).

[17] B. C. Arnold and J. M. Sarabia, *Majorization and the Lorenz order with applications in applied mathematics and economics* (Springer, 2018).

[18] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge university press, 1952).

[19] I. Olkin and A. W. Marshall, *Inequalities: Theory of Majorization and Its Applications*, Vol. 143 (Academic Press, 2016).

[20] K. Fan, *On a theorem of weyl concerning eigenvalues of linear transformations: Ii,* Proceedings of the National Academy of Sciences **36**, 31 (1950).

[21] I. Eliazar, *A tour of inequality,* Annals of Physics **389**, 306 (2017).

[22] A. Fontanari, P. Cirillo, and C. W. Oosterlee, *From concentration profiles to concentration maps. new tools for the study of loss distributions,* Insurance: Mathematics and Economics **78**, 13 (2018).

[23] C. Kleiber and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences* (Wiley, 2003).

[24] M. O. Lorenz, *Methods of measuring the concentration of wealth,* Publications of the American statistical association **9**, 209 (1905).

[25] I. Eliazar and I. M. Sokolov, *Measuring statistical evenness: A panoramic overview,* Physica A: Statistical Mechanics and its Applications **391**, 1323 (2012).

[26] S. S. Wilks, *Multidimensional statistical scatter,* Contributions to probability and statistics **2**, 486 (1960).

[27] A. B. Atkinson, *On the measurement of inequality,* Journal of economic theory **2**, 244 (1970).

[28] C. Gini, *Variabilità e mutabilità,* Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1912).

[29] G. Pietra, *Delle relazioni tra gli indici di variabilità* (Atti Del Reale Istituto Veneto Di Scienze, Lettere Ed Arti, Tomo LXXIV Parte II., 1915).

[30] M. Kritzman, Y. Li, S. Page, and R. Rigobon, *Principal components as a measure of systemic risk,* The Journal of Portfolio Management **37**, 112 (2011).

[31] M. Kritzman, *Risk disparity,* The Journal of Portfolio Management **40**, 40 (2013).

[32] L. van Hove, *Von neumann's contributions to quantum theory,* Bull. Amer. Math. Soc. **64**, 95 (1958).

[33] O. Roy and M. Vetterli, *The effective rank: A measure of effective dimensionality,* in *Signal Processing Conference, 2007 15th European* (IEEE, 2007) pp. 606–610.

[34] I. Eliazar, *How random is a random vector?* Annals of Physics **363**, 164 (2015).

[35] J. Von Neumann, *Some matrix inequalities and metrization of metric space,* Tomsk Univ. Rev **1**, 286 (1937).

[36] J. E. Gentle, *Numerical Linear Algebra for Applications in Statistics*, Statistics and Computing (Springer, 1998).

[37] F. J. Fabozzi, P. N. Kolm, D. A. Pachamanova, and S. M. Focardi, *Robust Portfolio Optimization and Management* (Wiley, 2007).

[38] L. Mirsky, *Symmetric gauge functions and unitarily invariant norms,* The quarterly journal of mathematics **11**, 50 (1960).

[39] S. Fortunato and D. Hric, *Community detection in networks: A user guide,* Physics Reports **659**, 1 (2016).

[40] G. Lindgren, *Model processes in nonlinear prediction with application to detection and alarm,* The Annals of Probability **8**, 775 (1980).

[41] P. Cirillo, J. Hüsler, and P. Muliere, *Alarm systems and catastrophes from a diverse point of view,* Methodology and Computing in Applied Probability **15**, 821 (2013).

[42] B. Karrer and M. E. Newman, *Random graph models for directed acyclic networks,* Physical Review E **80**, 046110 (2009).

[43] G. Dahl, *Majorization permutahedra and (0, 1)-matrices,* Linear Algebra and its Applications **432**, 3265 (2010).

[44] R. Van Der Hofstad, *Random graphs and complex networks*, Vol. 1 (Cambridge university press, 2016).

5

LORENZ-GENERATED ARCHIMEDAN COPULAS

An alternative generating mechanism for non-strict bivariate Archimedean copulas via the Lorenz curve of a non-negative random variable is proposed. Lorenz curves have been extensively studied in economics and statistics to characterize wealth inequality and tail risk. In this chapter, these curves are seen as integral transforms generating increasing convex functions in the unit square.

Many of the properties of these "Lorenz copulas", from tail dependence and stochastic ordering, to their Kendall distribution function and the size of the singular part, depend on simple features of the random variable associated to the generating Lorenz curve. For instance, by selecting random variables with lower bound at zero it is possible to create copulas with asymptotic upper tail dependence.

An "alchemy" of Lorenz curves that can be used as general framework to build multiparametric families of copulas is also discussed.

Keywords: *Archimedean copulas; Lorenz curves; stochastic ordering; tail dependence; Gini index.*

Parts of this chapter are available online at the SSRN Electronic Journal [1] and are currently under review.

5.1. INTRODUCTION

Borrowing tools from the flourishing literature on socio-economic inequality and statistical size distributions [2–4], this chapter introduces an alternative way to generate a class of non-strict bivariate Archimedean copulas [5], called Lorenz copulas, via the mirrored Lorenz curve [6]. Appealing characteristics of the novel approach are its flexibility and the possibility of characterizing the (upper tail) dependence structure of the related copulas, by studying some simple features of the Lorenz generator, and of its underlying size distribution.

The chapter is organized as follows: the next three subsections introduce the basic notation and the necessary tools; Section 5.2 is devoted to the study of Lorenz-generated non-strict Archimedean copulas; Section 5.3 contains some examples of copulas generated via the new approach; Section 5.4 discusses how to obtain new generators and how to develop multiparametric families of copulas; Section 5.5 closes the chapter. An Appendix contains technical and space-consuming details.

5.1.1. BIVARIATE ARCHIMEDEAN COPULAS: A QUICK REVIEW

First introduced by Sklar [7], copulas represent a convenient way of modeling multivariate phenomena [8], by disentangling the joint dependence structure from the marginal behavior. This is particularly true in applications, where the flexibility of copulas appears preferable to the direct fitting of multivariate distributions, which may be difficult to treat [5]. For the sake of completeness, not all statisticians agree with this view: for example Mikosch [9] maintains that the static separation of the dependence function from the marginal distributions gives a biased view of stochastic dependence.

In the bivariate framework, consider the two-variable function $C : [0,1]^2 \to [0,1]$. $C(u,v)$ is a two-dimensional copula [5] if:

1. $C(u,0) = C(0,v) = 0$ for every $u, v \in [0,1]$;
2. $C(u,1) = u$ and $C(1,v) = v$ for every $u, v \in [0,1]$;
3. $C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$ for every $u_1, v_1, u_2, v_2 \in [0,1]$, such that $u_1 \leq u_2$ and $v_1 \leq v_2$.

Sklar's theorem [7] shows that, if (X_1, X_2) is a random vector with joint distribution F and margins F_1 and F_2, then

$$F(x_1,x_2) = C(F_1(x_1), F_2(x_2)) = C(u,v), \qquad u,v \in [0,1].$$

In other words, it is possible to represent the bivariate distribution F of the random vector (X_1, X_2) in terms of the copula function C, and of two uniform margins obtained via the probability integral transform. A bivariate copula is thus nothing more than a bivariate distribution with uniform margins. If the marginals F_1 and F_2 are continuous then C is unique, otherwise it is uniquely determined on the Cartesian product of the support of the two marginals distributions.

It can be easily verified that, for every copula $C(u,v)$, one has

$$W(u,v) = \max(u+v-1, 0) \leq C(u,v) \leq \min(u,v) = M(u,v),$$

where W and M are known as the Fréchet-Hoeffding lower and upper bound respectively [5]. In the bivariate framework here considered, both W and M are proper copula functions.

A copula $C(u,v)$ is Archimedean if it is associative, $C(u,C(z,w)) = C(C(u,z),w)$ for every $u,z,w \in [0,1]$, and if its diagonal $\delta_C(x) := C(x,x)$ is such that $\delta_C(x) < x$ for all $x \in [0,1]$.

The associative nature of Archimedean copulas [10] allows for a very convenient representation in terms of a one-place function called the generator.

Theorem 5.1. *Let $C(u,v) \in [0,1]^2$ be an Archimedean copula, then there exists a function $\psi : [0,1] \to \mathbb{R}^+$ such that*

$$C(u,v) = \psi^{[-1]}(\psi(u) + \psi(v)) \quad u,v \in [0,1], \tag{5.1}$$

where $\psi^{[-1]}$ denotes the pseudo-inverse of ψ. In particular ψ is a decreasing convex function with $\psi(0) \leq \infty$ and $\psi(1) = 0$.

Note that the pseudo-inverse of a continuous, strictly decreasing function $f : [0,1] \to \mathbb{R}^+$ with $f(1) = 0$, is defined as: $f^{[-1]} : \mathbb{R}^+ \to [0,1]$ and such that

$$f^{[-1]}(x) = \begin{cases} f^{-1}(x) & 0 \leq x \leq f(0) \\ 0 & f(0) \leq x \leq \infty \end{cases}$$

where f^{-1} is the standard inverse of f. For a proof, see for example [5]. If one is familiar with non-Newtonian calculus [11], Equation (5.1) can be recognized as a ψ-arithmetic operation, a ψ-sum specifically [12]. This suggests that a bivariate Archimedean copula endows the interval $[0,1]$ with a semi-group structure.

To an Archimedean copula $C(u,v)$ it is always possible to associate a dual copula

$$\hat{C}(u,v) := u + v - C(u,v).$$

This dual copula is an S-norm [10], whose generator $\hat{\psi}$ is an increasing convex function with swapped boundary conditions with respect to those in Theorem 5.1.

In the literature, several functions ψ have been proposed over the years, from the well-known logarithmic and exponential generators behind the famous Clayton, Gumbel, Joe and Independence copulas [5], to those based on the inverse Laplace and the Williamson transforms [13]. In particular, this last class of generators provides a solution to the problem of finding d-monotonic functions, thus extending the Archimedean construction to an arbitrary number of dimensions.

Naturally, a large number of generators has given birth to a large number of copulas, and this richness (and flexibility) is one of the reasons of the popularity of the Archimedean family, in particular in applications [14].

From a theoretical point of view, an appealing feature of the Archimedean family is that the properties of a generator essentially determine the properties of the corresponding copula. For example, looking at the value $\psi(0)$, it is possible to distinguish between strict ($\psi(0) = \infty$) and non-strict ($\psi(0) < \infty$) copulas [5]. Non-strict copulas are the objects of interest of this chapter: their main peculiarity is that a subset of their domain has

zero probability mass (but positive Lebesgue measure), taking the name of zero set, or $Z(C)$.

Strictly related to $Z(C)$ is the zero curve $v = \kappa(u)$, that is the level curve separating the zero set from the part of the copula domain with positive mass. For a non-strict Archimedean copula the zero curve can be easily derived in terms of generator ψ, by setting $C(u,v) = 0$ in Equation (5.1), so that

$$\kappa(u) = \psi^{[-1]}(\psi(0) - \psi(u)), \qquad \forall u \in [0,1].$$

It is not difficult to verify that a non-strict Archimedean copula $C(u,v)$ is not able to model independence, i.e. $C(u,v) \neq uv \neq \Pi(u,v)$, either directly or as a limiting case. Nevertheless, non-strict copulas can be very useful when dealing with phenomena that exhibit upper tail dependence, or when one is interested in the dependence structure of random quantities that do not take on low quantiles simultaneously [15–17].

When dealing with bivariate copulas, and in particular with Archimedean copulas, a very important object of study is the Kendall distribution function $K(t)$ [18]. Such a function represents the bivariate equivalent of the univariate probability integral transform, and it is formally defined as

$$K(t) = P(C(U,V) \leq t), \qquad t \in [0,1],$$

where U and V are standard uniforms on $[0,1]$. For a fixed $t \in [0,1]$, the Kendall distribution function can be seen as the measure–also known as the $C-$measure [10]–of the set $\{(u,v) \in [0,1]^2 : C(u,v) \leq t\}$.

In an Archimedean copula, $K(t)$ can be obtained from the generator ψ, i.e.

$$K(t) = t - \frac{\psi(t)}{\psi'(t^+)}, \quad t \in [0,1], \tag{5.2}$$

with $\psi'(t^+)$ denoting the right derivative of ψ at t. In the Archimedean family, the function $K(t)$ uniquely determines the corresponding copula via its generator ψ.

The Kendall distribution function has many applications in copula theory. Just to list two of them that are useful here: 1) $K(t)$ induces a dependence ordering in the set of copulas, the so-called Kendall stochastic ordering [19]; and 2) $K(t)$ can be used to obtain association and dependence measures between random variables. For instance, the well-known Kendall's τ, measuring the concordance between two random variables, can be computed as

$$\tau = 3 - 4\int_0^1 K(t)\mathrm{d}t. \tag{5.3}$$

In this chapter τ will represent the main measure of (monotonic) dependence discussed. This is due to the fact that–as per Equation (5.3)–τ has a direct link with the generator ψ via the Kendall distribution function K, something that will be useful in Sections 5.2 and 5.3. Other measures like Spearman's ρ can naturally be computed, but their derivation is copula specific and it is not directly linked to ψ [5, 20], therefore they are less interesting in the framework of this chapter.

5.1.2. THE LORENZ CURVE

The Lorenz curve L of a non-negative random variable X, with finite expectation and cumulative distribution function F, is defined as

$$L(p) = \frac{\int_0^p F^{-1}(y)\mathrm{d}y}{\int_0^1 F^{-1}(y)\mathrm{d}y}, \quad p \in [0,1], \tag{5.4}$$

where F^{-1} is the quantile function of X [21]. A Lorenz curve completely characterizes the corresponding distribution function F up to a scale transformation [4].

Introduced by Max Lorenz in 1905 [6], the Lorenz curve is a well-known tool in the study of wealth and income inequality [3]. When the random variable X represents wealth in a given society, the curve $L(u)$ represents the percentage of wealth owned by the lower u% of the population. This makes the curve L essential to study economic size distributions, and to verify the so-called Pareto principle [2, 22].

Given a Lorenz curve it is then possible to construct a large number of inequality indices [3]. Among them, a famous one is the Gini G index [23], defined as (5.5).

$$G = 2\int_0^1 \big(p - L(p)\big)\,\mathrm{d}p. \tag{5.5}$$

Clearly $G \in [0,1]$. A Gini equal to 0 indicates a society in which everyone possesses the same amount of wealth, or alternatively the underlying random variable X is a Dirac delta function. For continuous random variables a Gini equal to 1 can be achieved only as a limit of random variables X_n representing more and more unequal distribution of wealth. The limiting random variable $\lim_{n\to\infty} X_n = X$ represents the maximum inequality, for discrete random variables it describes a situation in which one individual owns everything and all the others nothing. For the continuous case it signals the loss of the L^1 integrability condition. All other values represent intermediate situations: the higher the Gini, the more unequal the society.

As observed in [24], all inequality indices are nothing more than generalizations and improvements of some common measures of variability like the variance or the standard deviation. This justifies the rising interest for their application outside inequality studies, in fields like biostatistics and finance [25, 26].

The following proposition collects some useful properties of the Lorenz curve which will be needed later. For proofs, please refer to [27].

Proposition 5.1. *Let L be the Lorenz curve associated with the random variable $X \geq 0$, with $E[X] = \mu < \infty$ and distribution function F. The following holds:*

1. *$L(p)$ is a non-decreasing and convex function in $p \in [0,1]$, differentiable almost everywhere. If F is strictly increasing, then L is also strictly increasing. Moreover L satisfies the boundary conditions $L(0) = 0$ and $L(1) = 1$;*
2. *If L admits first derivative then $L'(p) = \frac{F^{-1}(p)}{\mu}$;*
3. *For all $p \in [0,1]$, the curve $L(p)$ is always bounded from above by $L_{PE}(p) = p$ and from below by $L_{PI}(p) = 0$ for all $p \in [0,1)$ and $L_{PI}(1) = 1$. The curves $L_{PE}(p)$ and $L_{PI}(p)$ are respectively called perfect equality and perfect inequality lines.*

4. *L can be seen as a distribution function. In particular, it represents the distribution function of the random variable $Y_L = L^{-1}(U)$ with U standard uniform on* $[0,1]$.

From Proposition 5.1 one can derive two important facts: 1) every non-decreasing convex function $g : [0,1] \to [0,1]$, such that $g(0) = 0$ and $g(1) = 1$ is a Lorenz curve [24] corresponding to some non-negative random variable X with finite expectation; 2) every Lorenz curve can be seen and used as a distortion function [20, 28], i.e. an integral transform generating increasing convex functions in the unit square. This second fact will prove useful later in the chapter.

5.1.3. ORDERS

Dependence orderings are multivariate stochastic orders defining posets[1] among copulas [29]. The following definition introduces three cases relevant for the chapter.

Definition 5.1. Let $C_1(u,v)$ and $C_2(u,v)$ be two copulas, with Kendall distribution functions $K_1(t)$ and $K_2(t)$ respectively. Denote by ${}_xC(v)$, the conditional probability $P(V \leq v | U \leq x)$, and by ${}_xC^{-1}$ its inverse. We have

1. Left tail decreasing order (LTD): $C_1 \succ^{\text{LTD}} C_2$, if ${}_{x'}C_1({}_xC_1^{-1}(u)) \leq {}_{x'}C_2({}_xC_2^{-1}(u))$, for $0 \leq x < x' \leq 1$.
2. Positive Kendall order (PK): $C_1 \succ^{\text{PK}} C_2$, if $K_1(t) \leq K_2(t)$ for every $t \in [0,1]$.
3. Positive quadrant order (PQD): $C_1 \succ^{\text{PQD}} C_2$, if $C_1(u,v) \geq C_2(u,v)$, for all $(u,v) \in [0,1]^2$.

If $C_1(u,v)$ and $C_2(u,v)$ are Archimedean, one has the following relevant implications

$$LTD \Rightarrow PK \Rightarrow PQD.$$

As proven in [15], for Archimedean copulas the conditions given in Definition 5.1 can be restated in terms of their generators, as in the following theorem.

Theorem 5.2. *Let $C_1(u,v)$ and $C_2(u,v)$ be two Archimedean copulas with generators ψ_1 and ψ_2 respectively. The following holds:*

1. *$C_1 \succ^{LTD} C_2$, if and only if $\psi_1(\psi_2^{[-1]}(y))$ is convex;*
2. *$C_1 \succ^{PK} C_2$, if and only if $\psi_1(\psi_2^{[-1]}(y))$ is star-shaped;*
3. *$C_1 \succ^{PQD} C_2$, if and only if $\psi_1(\psi_2^{[-1]}(y))$ is super-additive.*

As in the case of copulas, it is possible to define notions of stochastic ordering which make the set of non-negative random variables with finite expectations a poset [27]. The following definition introduces two useful stochastic orders related to the Lorenz curve.

Definition 5.2. Let X_1 and X_2 be two non-negative random variables with finite expectation, and let L_1, L_2 be their Lorenz curves. The Lorenz order and the star order are defined as follows:

[1] Recall that the word poset stands for *partially ordered set*

1. Lorenz order: $X_1 \succ^{\mathrm{L}} X_2$, if $L_1(p) \leq L_2(p)$ for all $p \in [0,1]$;
2. Star order: $X_1 \succ^{*} X_2$, when $L_1(L_2^{-1}(p))$ is convex in $p \in [0,1]$.

Observe that, if $X_1 \succ^{\mathrm{L}} X_2$, then $G_1 \geq G_2$, where G_1 and G_2 are the Gini indices of X_1 and X_2 respectively. Furthermore, notice that the star order condition in Definition 5.2 is not the usual one relying on quantile functions [29]. However, as shown in the Appendix, the two definitions are equivalent. Finally, it can be easily shown that the star order implies the Lorenz one [27].

5.2. LORENZ GENERATORS AND LORENZ COPULAS

Let L be a strictly increasing Lorenz function, associated to the strictly increasing distribution function F of a non-negative random variable X with finite mean. For $p \in [0,1]$ define the mirrored Lorenz curve as

$$\bar{L}(p) := L(1-p).$$

This new function is strictly decreasing and convex, with $\bar{L}(0) = 1$ and $\bar{L}(1) = 0$.

Recalling Theorem 5.1, it is evident that $\bar{L}(p)$ is a valid generator for a bivariate copula $C(u,v)$ of Archimedean type. Since $\bar{L}(0) < \infty$, the copula will be non-strict. Similarly, the original Lorenz curve $L(p)$ can be seen as a proper generator for the dual copula $\hat{C}(u,v)$.

The Gini index corresponding to $\bar{L}$ is given by $\bar{G} = 2\int_0^1 (1-p-\bar{L}(p))\mathrm{d}p$, and it is easy to verify that its value will coincide with that of the standard Gini G associated to L. For this reason, the notation G will be used to indicate both indices.

Definition 5.3. Let L and $\bar{L}$ be the standard and the mirrored Lorenz curves associated to a non-negative random variable X with finite mean. The corresponding bivariate non-strict Archimedean copula $C(u,v)$ is given by

$$C(u,v) = \bar{L}^{[-1]}(\bar{L}(u) + \bar{L}(v)) = 1 - L^{[-1]}(L(1-u) + L(1-v)), \qquad u,v \in [0,1]^2. \tag{5.6}$$

$C(u,v)$ is also referred to as the copula associated with X, or the copula generated by L and $\bar{L}$.

In the rest of the chapter, $C(u,v)$ will always represent a non-strict bivariate Lorenz-generated Archimedean copula, often just called Lorenz copula for brevity.

Every non-strict copula is characterized by the presence of a zero set and the relative zero curve. For a Lorenz copula, the zero curve $\kappa(p)$ is easily derived to be

$$\kappa(p) = 1 - L^{[-1]}(1 - L(1-p)), \quad p \in [0,1]. \tag{5.7}$$

Interestingly, the function $\kappa(p)$ is itself a mirrored Lorenz curve, given that it follows the composition rules described in [27]. This means that a particular Gini index, the zero Gini G_κ, can be computed from it. Such an index has an appealing interpretation in terms of the corresponding Lorenz copula: G_κ measures indeed how far $C(u,v)$ is from its Fréchet-Hoeffding bounds, i.e. from co- and countermonotonicity. $G_\kappa = 0$ tells that the copula coincides with the Fréchet-Hoeffding upper bound M, while $G_\kappa = 1$ indicates that $C(u,v)$ coincides with the Fréchet-Hoeffding lower bound W.

The following proposition clarifies the relation between G_κ and the Gini G associated with the Lorenz generator $\bar{L}$.

Proposition 5.2. *Let G_κ be the zero Gini associated to the zero set $Z(C)$ of the Lorenz copula $C(u,v)$, while G is the Gini associated to the Lorenz generator $\bar{L}$. Then*

$$G_\kappa \geq G.$$

Moreover, if X_1 and X_2 are two non-negative random variables and $X_1 \succ^L X_2$, one has

$$G^1_\kappa \leq G^2_\kappa,$$

where G^i_κ, $i = 1,2$, is the zero Gini of the copula associated with X_i.

Proof. To prove the first statement, notice that

$$G_\kappa - G \propto \int_0^1 \left(L^{[-1]}(1-L(1-p)) - 1 + L(1-p)\right) \mathrm{d}p. \tag{5.8}$$

In order for $G_\kappa \geq G$ to hold true it is sufficient to show that the right-hand side of Equation (5.8) is non-negative.

Setting $y = 1 - L(1-p)$, one gets

$$\int_0^1 \frac{1}{F^{-1}(1-y)} (L^{[-1]}(y) - y)\, dy \geq 0. \tag{5.9}$$

Since, by construction, $L^{[-1]}(y) \geq y$ for all $y \in [0,1]$, and view that F^{-1} is the quantile function of a non-negative random variable, the integrand in Equation (5.9) is positive for every $y \in [0,1]$. Hence $G_\kappa \geq G$.

To prove the second statement it is sufficient to show that

$$\int_0^1 L_1^{[-1]}(1-L_1(1-p))\mathrm{d}p \leq \int_0^1 L_2^{[-1]}(1-L_2(1-p))\mathrm{d}p, \tag{5.10}$$

where $L_i(p)$ is the Lorenz curve of X_i, $i = 1,2$, such that $L_1(u) \leq L_2(u)$ for every $p \in [0,1]$.

The function $1-L(1-p)$ is known as Leimkuhler curve in the inequality literature [30] and it induces a partial ordering, $\succ^{LK}$, in the space of non-negative random variables and size distributions [4]. According to such an order, $X_1 \succ^L X_2$ if and only if $X_1 \prec^{LK} X_2$. Now, observing that $L^{-1}(p)$ is an increasing concave function for $p \in [0,1]$, and that $L_1^{-1}(p) \leq L_2^{-1}(p)$ for all $p \in [0,1]$ thanks to the Leimkuhler order, then it is possible to conclude that (5.10) holds true. □

From Proposition 5.2 one derives that a necessary condition for $G^1_\kappa \leq G^2_\kappa$ is that $G_1 \leq G_2$. Such a result proves extremely useful when generating Lorenz copulas that need to satisfy some specific conditions in terms of their zero sets and zero curves; more details in Section 5.3.

Many analytic properties of the copula $C(u,v)$ trace back to the non-negative finite-mean variable X, its Lorenz curve L and the associated mirrored Lorenz generator $\bar{L}$. In the following subsections these properties are collected per topic.

5.2.1. BOUNDS AND SINGULARITIES

The next proposition clarifies under which conditions a Lorenz copula replicates the Fréchet-Hoeffding lower bound.

Proposition 5.3. *Let $C(u,v)$ be a non-strict Archimedean copula, while $\bar{L}$ is its Lorenz generator obtained from the curve L of the non-negative finite-mean random variable X. The following statements are then equivalent:*

1. *$C(u,v)$ coincides with the Fréchet-Hoeffding lower bound $W(u,v) = \max(u+v-1,0)$;*
2. *X is characterized by perfect equality, i.e. $L(p) = L_{PE}(p)$;*
3. *$G = 0$.*

Proof. The goal is to show that $1 \Leftrightarrow 2$ and $2 \Leftrightarrow 3$.

Assume that $L(p) = L_{PE}(p)$. The mirrored Lorenz is $\bar{L}_{PE}(p) = 1-p$. By applying Definition 5.3, one gets $C(u,v) = u+v-1$ if $u+v > 1$, and $C(u,v) = 0$ otherwise. Hence $C(u,v) = W(u,v) = \max(u+v-1,0)$. The opposite implication is then straightforward.

To prove that 2 implies 3, it is sufficient to compute the Gini index in Equation (5.5) for $L(p) = p$, obtaining $G = 0$.

Finally $3 \Rightarrow 2$ holds since $L(p) \leq p$, therefore the only solution for the functional equation $\int_0^1 p - L(p)\,\mathrm{d}p = 0$ is $L(p) = p$, which concludes the proof. □

Regarding the Fréchet-Hoeffding upper bound, no Archimedean structure is able to replicate it exactly [5]. Therefore, an if-and-only-if characterization cannot be given either for a Lorenz copula. However, it is possible to show that, if a Lorenz curve converges point-wise to the perfect inequality case, and thus its Gini tends to 1, then the corresponding $C(u,v)$ will have the upper bound M as its limit.

Proposition 5.4. *Let $C(u,v)$ be the Lorenz copula generated by L and such that $C(u,v) \to M(u,v) = \min(u,v)$. Then $L \to L_{PI}$ (or $\bar{L} \to \bar{L}_{PI}$) and $G \to 1$.*

Proof. A necessary and sufficient condition for an Archimedean copula with generator ψ_θ parametrized by θ to attain the Fréchet-Hoeffding upper bound is given in [5], i.e.

$$\lim_{\theta \to \theta^*} \frac{\psi_\theta(p)}{\psi'_\theta(p)} = 0, \quad p \in (0,1). \tag{5.11}$$

In terms of Lorenz generators, for a Lorenz curve L_θ of parameter θ, such a condition clearly becomes

$$\lim_{\theta \to \theta^*} \frac{\bar{L}_\theta(p)}{\bar{L}'_\theta(p)} = 0, \qquad p \in (0,1). \tag{5.12}$$

The only Lorenz satisfying such a condition is the perfect inequality one, L_{PI}, the lower bound for every Lorenz curve, as per Proposition 5.1. In fact, all the other Lorenz curves are increasing and bounded, hence they cannot satisfy $\lim_{\theta \to \theta^*} \bar{L}'_\theta(p) = \infty$ for every $p \in (0,1)$, which is the only condition for Equation (5.12) to hold for a generic (non perfect inequality) Lorenz with $\lim_{\theta \to \theta^*} \bar{L}_\theta(p) \neq 0$, for $p \in (0,1)$.

This said, by Equation (5.5) the Gini index associated to the Lorenz generator will converge to 1, the value for perfect inequality. □

Following Propositions 5.3 and 5.4, a Lorenz copula may attain the Fréchet-Hoeffding bounds if the distribution of the underlying variable X can model both a completely even and a totally uneven distribution of wealth. As discussed in [4], this is not always the case, meaning that not all Lorenz copulas can reach the bounds W and M.

Another important feature of non-strict Archimedean copulas is the presence or the absence of a singular part. For a Lorenz copula this completely depends on the continuity of the quantile function of the underlying random variable X.

Proposition 5.5. *Let $C(u,v)$ be the Lorenz copula associated to the non-negative finite-mean random variable X. Then $C(u,v)$ exhibits a singular component if and only if X has a continuous density and a bounded support. Furthermore the singular component is placed on the zero curve.*

Proof. Since X has a continuous density, the Lorenz curve L is twice differentiable, and so is the generator $\bar{L}$. Hence, if present, any singular part must be located on the zero curve [5].

In order to quantify the mass of the singular part, one needs the Kendall distribution function of the Lorenz copula $C(u,v)$, which is

$$K(t) = t + \frac{\bar{L}(t)\mu}{F^{-1}(1-t)}, \qquad t \in [0,1]. \tag{5.13}$$

Thanks to Theorem 4.3.3 in [5], the C-measure of the zero curve is given by

$$K(0) = \frac{\bar{L}(0)\mu}{F^{-1}(1)} = \frac{\mu}{b}, \tag{5.14}$$

where $b \leq +\infty$ is the upper bound of the support of X and $\mu = E[X]$.

If X has no finite upper bound, clearly $b = +\infty$, and the $C-$measure of the zero curve is zero, thus concluding the proof. □

If the random variable X does not have a continuous density, Proposition 5.5 does not hold. However, one can easily verify that, by taking continuous non-differentiable Lorenz curves, as those arising from discrete random variables, it is possible to define Lorenz copulas with singular parts everywhere in the unit square, and not necessarily on the zero curve. The localization and the quantification of the $C-$measure for these singularities is quite simple, and it depends on the jumps in the quantile function of X. If the Lorenz copula exhibits a discontinuity in t, the $C-$measure corresponding to it is in fact $\mu\left(\frac{1}{F^{-1}(1-t^+)} - \frac{1}{F^{-1}(1-t^-)}\right)$.

5.2.2. DEPENDENCE AND INEQUALITY ORDERS

In the theory of copulas, a fundamental topic is the analysis of the dependence structure induced by a given copula function. For Lorenz copulas the dependence structure is directly connected to the underlying variable X and to the stochastic orders discussed in Subsection 5.1.3.

Proposition 5.6. *Let C_1 and C_2 be two Lorenz copulas associated to X_1 and X_2, non-negative random variables with finite means. One has that $C_1 \succ^{LTD} C_2$ if and only if $X_1 \succ^{*} X_2$.*

Proof. The proof is a straightforward application of Definition 5.2 and Theorem 5.2. In particular, note that the condition on the generators for the LTD order, $\frac{\mathrm{d}^2\psi_1(\psi_2^{[-1]}(p))}{\mathrm{d}^2 p} \geq 0$, can be rewritten in terms of the Lorenz generators as $\frac{\mathrm{d}^2 L_1(L_2^{[-1]}(p))}{\mathrm{d}^2 p} \geq 0$, which is exactly the condition for $X_1 \succ^* X_2$ according to Definition 5.2. □

If two Lorenz copulas are left tail ordered, then one of the associated random variables is more unequal than the other. In terms of Gini indices, one can easily verify that a necessary condition for $C_1 \succ^{\mathrm{LTD}} C_2$ is that $G_1 \geq G_2$.

Proposition 5.7. *Let C_1 and C_2 be two Lorenz copulas associated with the non-negative finite-mean random variables X_1 and X_2, with Gini indices G_1 and G_2. Then $C_1 \succ^{PK} C_2$ only if $G_1 \geq G_2$.*

Proof. By Theorem 5.2 a necessary and sufficient condition for $C_1 \succ^{\mathrm{PK}} C_2$ is $L_1(L_2^{[-1]}(u))$ being star-shaped.

Looking at derivatives, the star shape condition implies that $\frac{L_1(u)}{L_2(u)}$ is increasing. This is equivalent to $Y_{L_2} \succ^* Y_{L_1}$, where $Y_L \sim L$ (recall the last point in Proposition 5.1).

Therefore, remembering that the star order implies the Lorenz one [27], and setting $L^{(1)}(p) = \int_0^p L(t)\mathrm{d}t$, one gets that a necessary condition for the Kendall order is that

$$L_1^{(1)}(p) \geq L_2^{(1)}(p), \quad \forall p \in [0,1]. \tag{5.15}$$

Equation (5.15) can be rewritten in terms of the original quantile functions, i.e.

$$\int_0^p \int_0^u F_1^{-1}(t)\mathrm{d}t\mathrm{d}s \geq \int_0^p \int_0^s F_2^{-1}(t)\mathrm{d}t\mathrm{d}s, \qquad \forall p \in [0,1]. \tag{5.16}$$

This condition represents an ordering on non-negative random variables called inverse stochastic dominance of third degree (ISD(3)) between X_1 and X_2, and it can be shown [31] that the condition $G_1 \geq G_2$ is a necessary one for ISD(3) to hold. □

Here below a graphical summary of the relations among the orders cited in this section is provided.

$$\begin{array}{ccccccccc} \mathrm{LTD} & & \Rightarrow & & \mathrm{PK} & \Rightarrow & \mathrm{PQD} & \Rightarrow & \tau \\ \Updownarrow & & & & \Downarrow & & & & \\ * & \Rightarrow & \mathrm{L} & \Rightarrow & \mathrm{ISD(3)} & \Rightarrow & \mathrm{G} & & \end{array}$$

As stated, the LTD and * orders imply each other. LTD then implies PK and PQD, and from PQD one can derive that if $C_1(u,v) \succ^{PQD} C_2(u,v)$, then $\tau_1 > \tau_2$. Similarly, * implies L, which implies ISD(3). From ISD(3) one can finally obtain an order on the Gini indices.

5.2.3. UPPER TAIL DEPENDENCE

An interesting aspect of Lorenz copulas is the possibility of having different types of tail dependence. Once again, the underlying variable X plays a major role.

In [32], the upper tail behavior of Archimedean copulas is classified into three regimes, for which a characterization is offered in terms of generators:

1. Tail independence, if and only if
$$\psi'(1) < 0. \tag{5.17}$$
2. Asymptotic (upper) tail dependence, if and only if $\psi'(1) = 0$ and $\lim_{t\to 0} -\frac{t\psi'(1-t)}{\psi(1-t)} = 1$.
3. (Upper) Tail dependence, if and only if $\psi'(1) = 0$ and $\lim_{t\to 0} -\frac{t\psi'(1-t)}{\psi(1-t)} > 1$.

For Lorenz copulas, a necessary and sufficient condition for tail independence is that the support of the underlying non-negative finite-mean variable X has a lower bound strictly larger than zero. In fact, it is sufficient to observe that Equation (5.17) becomes

$$\bar{L}'(1) = -\frac{F^{-1}(0)}{\mu}, \tag{5.18}$$

where F^{-1} is the quantile function of X and $\mu = E[X] > 0$. Looking at the right-hand side of Equation (5.18) the condition on the lower bound is then evident.

5

According to the conditions in [32], asymptotic upper tail dependence requires $\bar{L}'(1) = 0$, therefore the lower bound of X needs to be 0: it cannot be negative, since X is non-negative, and it cannot be larger than 0, or it would fall in the case of tail independence.

In order to decide whether the upper tail dependence is asymptotic or not, one needs to study the behavior of $\lim_{p\to 0} -\frac{p\bar{L}'(1-p)}{\bar{L}(1-p)}$. In terms of X and of its cumulative distribution function F, if one sets $y = F^{-1}(x)$ and applies L'Hôpital's rule twice, such a limit becomes

$$\lim_{y\to 0} \frac{F''(y)F(y)}{(F'(y))^2} \leq 1. \tag{5.19}$$

If the limit above is strictly smaller than 1, one has tail dependence, while the asymptotic case appears when the limit is exactly 1.

Theorem 5.3. *A Lorenz copula shows asymptotic tail dependence when* $\lim_{y\to 0} \frac{F''(y)F(y)}{(F'(y))^2} = 1$, *that is if and only if the underlying X shows a lognormal-like left tail in a neighbourhood of zero.*

Proof. To prove the if part, it is sufficient to show that the lognormal distribution attains the equality in (5.19).

Let $\Phi(\cdot)$ be the distribution function of a standard normal, and $\phi(\cdot)$ the density. For a lognormal random variable Equation (5.19) becomes

$$\lim_{y\to 0} \frac{\Phi(\log(y))\left(\frac{1}{y^2}(\phi'(\log(y)) - \phi(\log(y)\right)}{\left(\frac{1}{y^2}\phi(\log(y)\right)^2} = \lim_{z\to -\infty} \frac{\Phi(z)(-1-z)}{\phi(z)}, \tag{5.20}$$

with $z = \log(y)$ and recalling that $\phi'(z) = -z\phi(z)$ [33].

Using L'Hôpital's rule twice, Equation (5.20) evolves into

$$\lim_{z\to -\infty} \frac{\phi'(z)(-1-z) - 2\phi(z)}{\phi''(z)} = \lim_{z\to -\infty} \frac{z^2+z-2}{z^2-1} = 1, \tag{5.21}$$

since $\phi''(z) = (z^2-1)\phi(z)$.

To prove the only if part, assume that there exists another Lorenz curve $\widetilde{L}(x)$, which is not the one of a lognormal random variable, and such that

$$\lim_{p\to 0} \frac{p\widetilde{L}'(p)}{\widetilde{L}(p)} = 1. \tag{5.22}$$

If such a Lorenz curve exists then it must be true that

$$\lim_{p\to 0} \frac{\frac{\widetilde{L}'(p)}{\widetilde{L}(p)}}{\frac{L'(p)}{L(p)}} = 1, \tag{5.23}$$

where $L(p)$ is the Lorenz curve of a lognormally distributed random variable with parameter $\sigma > 0$, i.e. $L(p) = \Phi(\Phi^{-1}(p) - \sigma)$.

In a neighbourhood of zero then the following differential equation is to hold:

$$\log \widetilde{L}(p)' = \frac{L'(p)}{L(p)}, \tag{5.24}$$

where $L'(p) = e^{-\sigma\left(\frac{1}{2}\sigma - \Phi^{-1}(p)\right)}$.

Solving Equation (5.24), one finds $\widetilde{L}(p) \propto \Phi(\Phi^{-1}(p) - \sigma)$, which is the lognormal Lorenz curve up to a constant. Therefore by the uniqueness of the solution of a differential equation (up to a shift), one can conclude that, to attain equality in Equation (5.19), a lognormal behavior in the vicinity of zero is needed. □

The presence or the absence of lognormality as well as the validity of the condition (5.19) may be cumbersome to check for a generic distribution function F with density f. Corollary 5.1 provides an easy-to-check sufficient condition for the presence of tail dependence in Lorenz copulas.

Corollary 5.1. *Let $f(x)$ be the density of a non-negative finite-mean variable X with lower bound 0, included or excluded. If $\lim_{x\to 0} f(x) > 0$ then the Lorenz copula generated by X exhibits upper tail dependence.*

Proof. It is easy to verify that if $\lim_{x\to 0} f(x) = c > 0$, then $\lim_{x\to 0} F''(x) = 0$. By substituting these values into Equation (5.19) one gets $\frac{0}{c^2} = 0 < 1$. Therefore the inequality in Equation (5.19) is strict and the presence of tail dependence is always verified. □

5.3. EXAMPLES OF LORENZ COPULAS

The present section is devoted to the illustration of some Lorenz copulas. Playing with Lorenz generators, it is not only possible to recover very well-known models, but–more interestingly–one can obtain brand new non-strict Archimedean copulas, with useful tail properties.

5.3.1. THE LOGNORMAL LORENZ COPULA

The Lognormal Lorenz copula is obtained by assuming X to be lognormally distributed, so that the generator is

$$\bar{L}_{LN}(p) = \Phi\left(\Phi^{-1}(1-p) - \sigma\right), \qquad p \in [0,1], \tag{5.25}$$

where Φ is the distribution function of a standard normal distribution, Φ^{-1} the corresponding quantile function, and $\sigma > 0$ is a scale parameter (equal to the standard deviation of $\log(X)$). Figure 5.1 shows examples of the lognormal generator for different values of σ.

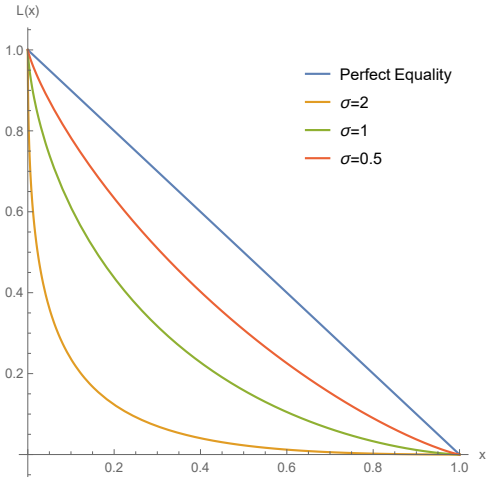


Figure 5.1: Examples of lognormal generators with different σ parameters.

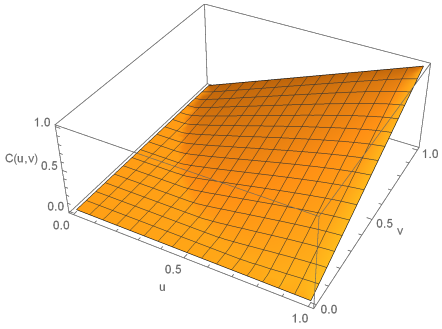
Notice that the quantity $\Phi(\Phi^{-1}(y)-\sigma)$, with $\sigma > 0$, is a well-known distortion function in actuarial mathematics, usually called Wang transform, and it has powerful applications in the fields of asset pricing, risk theory and utility theory [20, 34]. Furthermore, in terms of non-Newtonian calculus [11], it represents the pseudo-difference of a variable y and a constant σ: just notice that, given the continuity of Φ^{-1}, one has $\sigma = \Phi^{-1}(\Phi(\sigma))$, so that $\Phi(\Phi^{-1}(y) - \sigma) = \Phi\left(\Phi^{-1}(y) - \Phi^{-1}(\Phi(\sigma))\right)$ [12, 35].

Using Equation (5.6), the functional form of the lognormal Lorenz copula is:

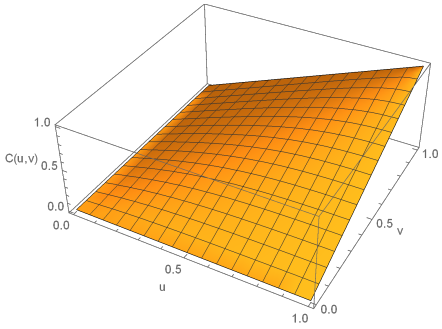
$$C_{LN}(u,v) = \max\left(1 - \Phi(\Phi^{-1}(\Phi(\Phi^{-1}(1-u) - \sigma) + \Phi(\Phi^{-1}(1-v) - \sigma)) + \sigma), 0\right). \tag{5.26}$$

Figure 5.2 shows the surface of $C_{LN}(u,v)$ for $\sigma = 0.5$ and $\sigma = 2$.

Since the lognormal variable X has an unbounded support (its right-end point is $x_F = +\infty$), Proposition 5.5 guarantees that the lognormal Lorenz copula has no singular part. Moreover, Theorem 5.3 tells us that such a copula is characterized by asymptotic tail dependence, whose strength grows with σ. In Figure 5.3 two simulations are given, and in both of them it is possible to notice the expected tail behavior: observations in the top right corner are more dependent.

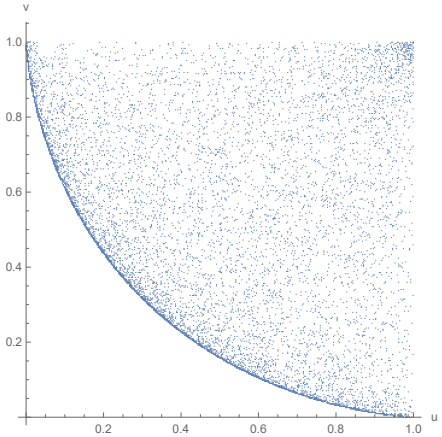


(a) $\sigma = 0.5$

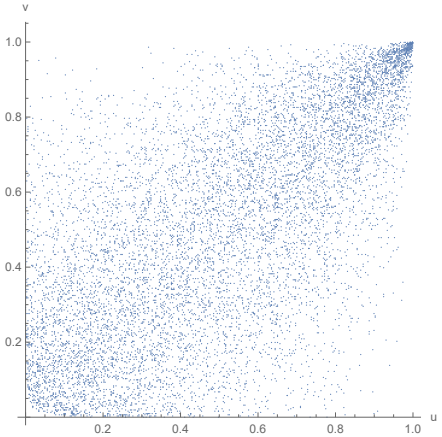


(b) $\sigma = 2$

Figure 5.2: Surfaces of a lognormal Lorenz copula for different values of σ.



(a) $\sigma = 0.5$



(b) $\sigma = 2$

Figure 5.3: Two simulations from a lognormal Lorenz copula for different σ.

The Kendall distribution function associated to $C_{LN}(u,v)$ is given by:

$$K_{LN}(t) = t + \Phi(\Phi^{-1}(1-t) - \sigma)e^{\frac{\sigma^2}{2} - \Phi^{-1}(1-t)}, \quad t \in (0,\infty), \tag{5.27}$$

which is trivially obtained by noting that the quantile function of a lognormal distribution rescaled by the mean is given by $e^{-(\frac{\sigma^2}{2} - \Phi^{-1}(1-t))}$.

From Equation (5.27), one can then obtain the value of the Kendall's τ, but this is only possible numerically, the analytical derivation being unfeasible. In Figure 5.4 the relation between τ and the parameter σ is presented. It is clear that monotonic dependence grows with σ. This is somehow expected by looking back at Figure 5.3, where not only tail dependence gets stronger as σ becomes larger, but the size of the zero set decreases. In the limit, for $\sigma \to \infty$, the lognormal Lorenz copula tends to the Fréchet-Hoeffding upper bound M (and to W for $\sigma \to 0$).

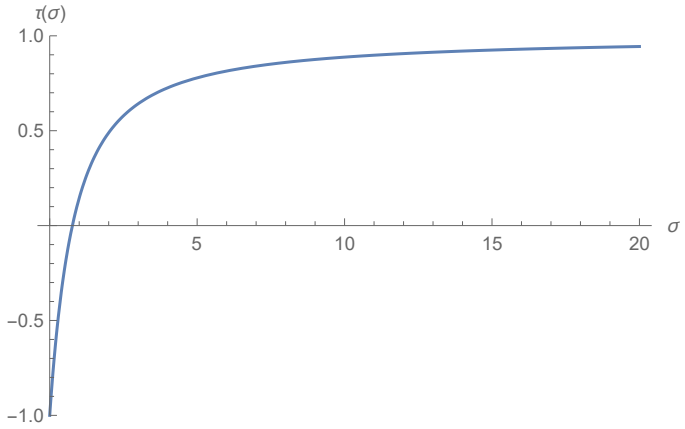


Figure 5.4: Lognormal Lorenz Copula Kendall's τ_{LN} as a function of σ.

Finally, some considerations in terms of stochastic orders. It is known that the lognormal distribution is star ordered in σ [4]. Namely, if $\sigma_1 > \sigma_2$ then $X_1 \succ^* X_2$, with $X_i \sim LN(\mu, \sigma_i)$, $i = 1, 2$. Thanks to Proposition 5.6 this means that the lognormal Lorenz copula is LTD ordered, and this also implies the positive quadrant dependence and the Kendall order.

5.3.2. THE SHIFTED EXPONENTIAL LORENZ COPULA

The shifted exponential Lorenz copula is obtained via the generator

$$\bar{L}_{SE}(p) = (1-p) + 2gp\log(p), \tag{5.28}$$

with $g \in (0, \frac{1}{2}]$. When $g = \frac{1}{2}$ the mirrored Lorenz curve in Equation (5.28) corresponds to that of a standard exponential random variable $X \sim \text{Exp}(\lambda)$. Notice that, for all exponentials, the (mirrored) Lorenz curve does not depend on λ, i.e. all exponentials share the same Lorenz curve, as observed in [26]. For $g < \frac{1}{2}$ the random variable X is shifted away from zero by a factor equal to $(1-2g)\lambda$. Figure 5.5 shows some examples of the generator $\bar{L}_{SE}(p)$ for different values of g.

The shifted exponential Lorenz copula obtained from Equation (5.28) is

$$C_{SE}(u,v) = \max\left(\exp\left(\frac{1}{2g} + W_{-1}\left(\frac{2g(u\log(u) + v\log(v)) - (u+v) + 1}{2g\exp\left(\frac{1}{2g}\right)}\right)\right), 0\right), \tag{5.29}$$

where W_{-1} is the lower-branch of the Lambert W function [36]. Two examples of $C_{SE}(u,v)$, for $g = 0.5$ and $g = 0.2$, are given in Figure 5.5. In the Appendix, the details of the derivation of Equation (5.29) are presented.

The following list contains some remarkable facts related to $C_{SE}(u,v)$:

1. The shifted exponential Lorenz copula has no singular part. This is a consequence of the unbounded support of the exponential distribution.

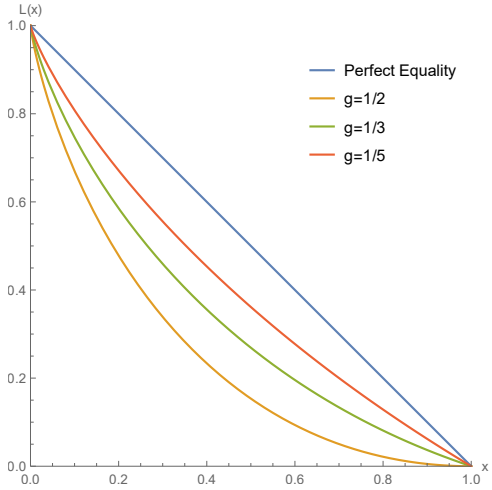


Figure 5.5: Examples of shifted exponential generators for different values of g.

2. The copula exhibits tail dependence only when $g = \frac{1}{2}$, i.e. when the support of the underlying random variable starts in 0. For all the other values of g, $C_{SE}(u,v)$ is tail independent. Figure 5.8 shows two simulations from $C_{SE}(u,v)$, with $g = 0.5$ and $g = 0.2$. As expected, the former case shows tail dependence, while the latter manifests tail independence.

3. The shifted Exponential random variables are star ordered with g [37]. Therefore, by Proposition 5.6, shifted exponential Lorenz copulas are ordered according to the LTD dependence order.

4. Since the Lorenz curve of perfect inequality is never attained by a shifted exponential random variable [4], Proposition 5.4 suggests that $C_{SE}(u,v)$ is always bounded away from the Fréchet-Hoeffding upper bound M.

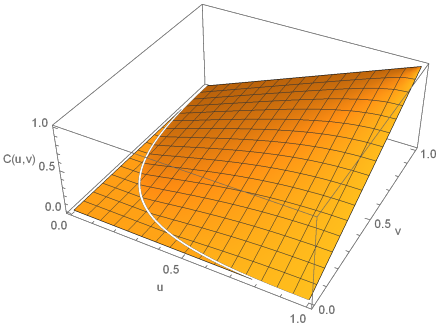
5. The Kendall's τ_{SE} of the shifted exponential Lorenz copula cannot be written in closed form, but only in terms of Gamma and Exponential Integral functions [36]. However, it can be easily evaluated numerically. Figure 5.7 shows its behaviour as a function of the parameter g. Interestingly, its range of variation is [-1,0.227], in line with the previous point.

5.3.3. THE PARETO LORENZ COPULA

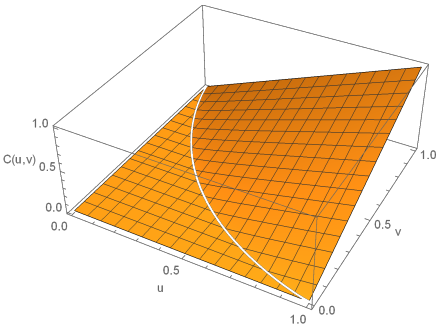
The Pareto Lorenz copula emerges when X is Pareto distributed with shape/tail parameter α and scale $x_m > 0$, with mirrored Lorenz curve equal to

$$\bar{L}_P(p) = 1 - p^{1-\frac{1}{\alpha}}, \qquad \alpha > 1. \tag{5.30}$$

For a Pareto random variable, $\alpha > 1$ is required in order to guarantee that $E[X] < \infty$, so that the Lorenz curve is defined. In Figure 5.9 some examples of L_P for varying α are given.



(a) $g = \frac{1}{2}$



(b) $g = \frac{1}{5}$

Figure 5.6: Surfaces of a shifted exponential Lorenz copula for different values of g.

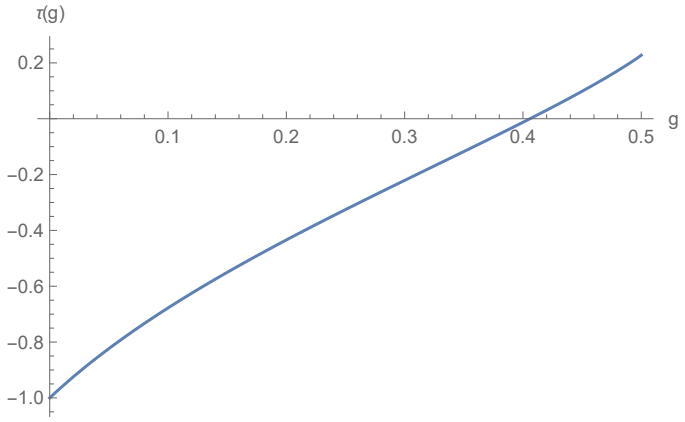


Figure 5.7: Shifted exponential Lorenz Copula Kendall's τ_{SE} as a function of g.

The Pareto Lorenz copula is

$$C_P(u,v) = \max\left((u^{1-\frac{1}{\alpha}} + v^{1-\frac{1}{\alpha}} - 1)^{\frac{1}{1-\frac{1}{\alpha}}}, 0\right), \tag{5.31}$$

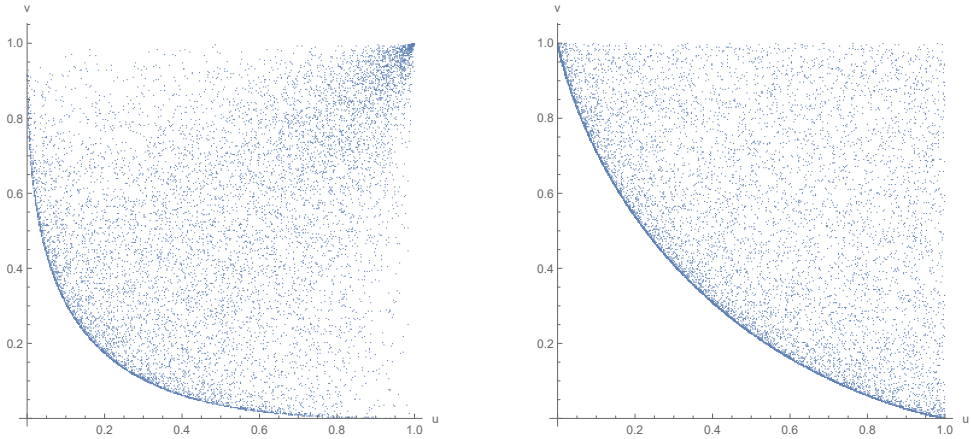
again with $\alpha > 1$.

It is worth noticing that, by setting $\theta = \frac{1}{\alpha} - 1$, the Pareto Lorenz copula coincides with the non-strict Clayton family, obtained for $\theta \in (-1, 0)$ [5].

Since the support of a Pareto random variable starts in $x_m > 0$, the Pareto Lorenz copula is upper tail independent for every choice of α. Moreover, since $X \sim Pareto(\alpha, x_m)$ is unbounded from above, $C_P(u,v)$ has no singular component.

As shown in the Appendix, the Pareto Lorenz copula is LTD ordered, as expected being a subset of the Clayton family. Recall that LTD then implies PK and PQD.

Finally, it is interesting to look at the role of the tail parameter α in the Kendall's τ of



(a) $g = 0.5$ (b) $g = 0.2$

Figure 5.8: Two simulations from a shifted exponential Lorenz copula for different g.

$C_P(u,v)$. One can easily verify that

$$\tau_P = \frac{1-\alpha}{1+\alpha} < 0, \qquad \alpha > 1. \tag{5.32}$$

Equation (5.32) can be re-written in terms of the Paretian Gini index $G_P = \frac{1}{2\alpha-1}$ [4], getting

$$\tau_P = \frac{G_P - 1}{3G_P + 1}. \tag{5.33}$$

Equation (5.33) shows that τ_P is an increasing function of $G_P \in [0,1]$, moving from -1 towards 0. This implies that the intensity of the association between two random variables coupled with a Pareto Lorenz copula decreases–in absolute value–with an increase in the inequality (in socio-economic terms) of the underlying Pareto random variable X.

Besides the Paretian case, it is worth stressing that, in general, the connection between τ and G is always rather interesting in Lorenz copulas.

5.3.4. THE UNIFORM LORENZ COPULA

The Uniform Lorenz Copula represents another interesting case. The underlying non-negative finite-mean variable X is taken to be uniformly distributed on $[a,b]$, with $0 \leq a < b < \infty$. One has

$$\bar{L}_U(p) = \frac{2a(1-p) + (b-a)(1-p)^2}{a+b}. \tag{5.34}$$

When $a = 0$ and $b = 1$, i.e. $X \sim U[0,1]$, Equation (5.34) simplifies to $\bar{L}_U(p) = (1-p)^2$. As usual, Figure 5.10 presents some examples of uniform Lorenz generators.

The uniform Lorenz copula is

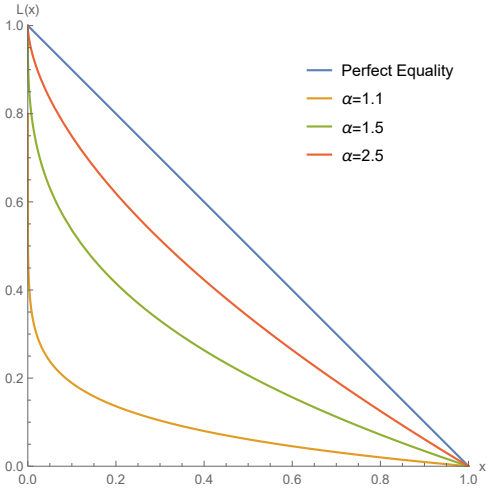


Figure 5.9: Examples of Pareto generators for different values of α.

$$C_U(u,v) = \max\left(1 - \frac{a - \sqrt{b^2(2+(-2+u)u+(-2+v)v)+2ab(u-u^2+v-v^2)+a^2(-1+u^2+v^2)}}{(a-b)}, 0\right),$$

and two examples of surfaces are given in Figure 5.11.

As far as the properties of $C_U(u,v)$ are concerned, one can observe the following:

1. The uniform Lorenz copula always possesses a singular part. From Equation (5.14), the $C-$measure is $\frac{a+b}{2b} > \frac{1}{2}$.
2. $C_U(u,v)$ exhibits tail dependence for $a = 0$, and tail independence for all $a > 0$.
3. The uniform family is star ordered with respect to a and b, therefore one can conclude that uniform Lorenz copulas are ordered according to the LTD (PK and PQD) dependence order.
4. The Gini index of the uniform family is $G_U = \frac{b-a}{3(a+b)}$, which can never be equal to one. Therefore, by Proposition 5.4, the uniform Lorenz copula will never attain its upper Fréchet-Hoeffding bound M.
5. Using Equation (5.3) it is possible to obtain a closed form formula for the Kendall's τ associated to the uniform Lorenz copula.

$$\tau_U = \frac{2a(a-b-a\log(ab))}{(a-b)^2}.$$

Observe that, if $a = 0$, one has $\tau_U = 0$ for every choice of b. But for $a = 0$ the uniform Lorenz copula necessarily exhibits tail dependence, hence this situation

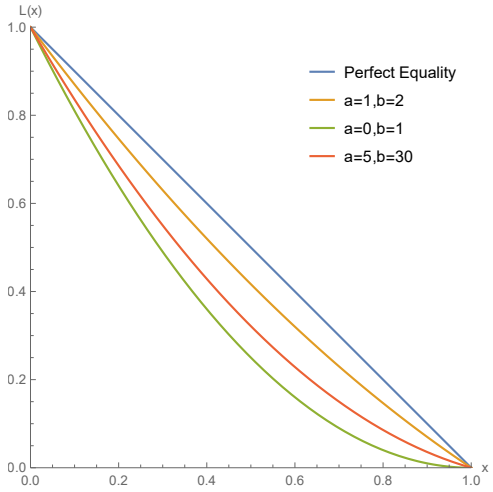
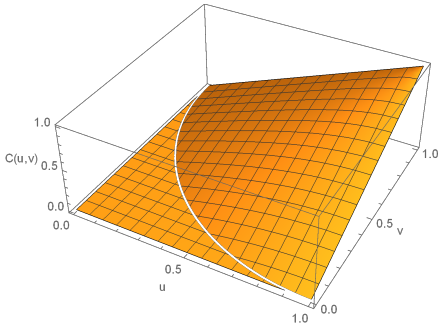
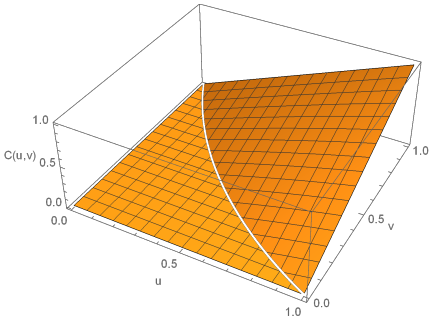


Figure 5.10: Examples of uniform Lorenz generators for some combinations of a and b.



(a) $a = 0, b = 1$

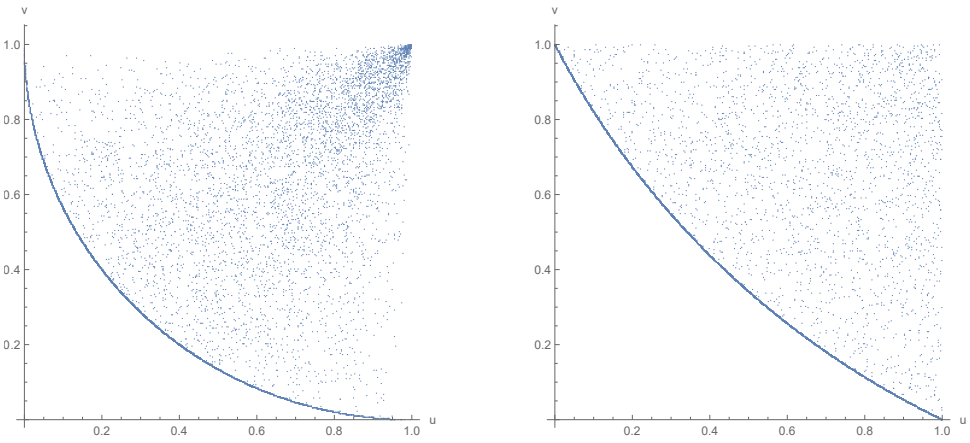


(b) $a = 1, b = 2$

Figure 5.11: Copula surface for Uniform Lorenz copula.

represents another pathological example of how measures of association should not be fully and acritically trusted when dealing with copulas [5, 9].

Figure 5.12 shows two samples drawn from two different uniform Lorenz copulas. The presence of tail dependence/independence given the values of a is evident.



(a) $a = 0, b = 1$ (b) $a = 1, b = 2$

Figure 5.12: Two simulations from a uniform Lorenz copula with $X \sim U[0,1]$ and $X \sim U[1,2]$.

5.4. ALCHEMIES AND MULTIPARAMETRIC EXTENSIONS

The class of Lorenz copulas is extremely rich and flexible. In the previous section a few examples were considered, starting from some well-known size distributions, but they only represent a small set of all the possible copulas one can actually generate. Just think about all the Lorenz curves currently available in the literature [4, 24, 27, 38, 39].

Besides flexibility, an appealing characteristics of the Lorenz approach to copulas is the possibility of importing into the Archimedean family many results developed in the studies of inequality. A particularly interesting example is represented by the so-called "alchemy of Lorenz curves" discussed in [27, 39]. The evocative term alchemy is used by Sarabia and Arnold to indicate a set of techniques for generating new Lorenz curves starting from given ones. Some relevant cases are presented in the following Proposition, for the proof of which see [27].

Proposition 5.8. *Let $L_1(p)$ and $L_2(p)$ be two Lorenz curves. Then, a new Lorenz curve $L(p)$ can be obtained, for instance, via*

1. *Exponentiation: $L(p) = L_1(p)^\alpha$, with $\alpha > 0$;*
2. *Composition: $L(p) = L_1(L_2(p))$;*
3. *(Generalized) Multiplication: $L(p) = L_1(p)^\alpha L_2(p)^\beta$, with $\alpha, \beta > 0$.*
4. *Convex combination: $L(p) = wL_1(p) + (1-w)L_2(p)$, with $w \in [0,1]$.*

5. *Maximization:* $L(p) = \max(L_1(p), L_2(p))$.

It is straightforward to verify that the statements in Proposition 5.8 are easily extended to any finite collection of Lorenz curves, so that, for example, $L(x) = \prod_{i=1}^{n} L_i(p)^{\alpha_i}$ and $\max(L_1(p), ..., L_n(p))$ are proper Lorenz curves.

Some of the transformations in Proposition 5.8 are already known for Archimedean generators. For instance, Nelsen's α and β families [5]–when restricted to non-strict generators–can be obtained by combining Statements 1. and 2. of Proposition 5.8. The same holds for most of the transformations presented in Proposition 1 of [40].

Furthermore, the alchemy of Lorenz curves can be used to create new multiparametric Lorenz copulas, providing a possible solution to the growing demand for more flexible copulas [5]. Consider, for instance, the Paretian generator in Equation (5.30). By applying exponentiation and multiplication one can easily obtain the following family of three-parameter Lorenz generators

$$\tilde{L}_{3P}(p) = (1-p)^{\eta}(1-p^{\theta})^{\gamma}, \tag{5.35}$$

with $\eta \geq 0$, $\theta \in (0,1)$ and $\gamma \geq 1$. The related Lorenz curves have been studied extensively in [41].

From Equation (5.35) it is possible to obtain a three-parameter Lorenz copula, whose properties can be studied using the results of Section 5.2. For example, one can quickly find out that the copula obtained from L_{3P} is almost never tail independent. This comes from the fact that $\tilde{L}'_{3P}(1) = 0$ for every choice of the parameters except for $\eta = 0$ and $\gamma = 1$.

By taking $\eta = 0$ in Equation (5.35), one obtains the mirrored Lorenz curve of a Sigh-Maddala random variable [42], which represents a pseudo-translation of a standard Pareto [4, 35], so that the new variable has its lower bound shifted to zero. Because of this new lower bound the original Paretian tail independence is lost.

By setting $\gamma = \frac{1}{\theta}$, Equation (5.35) becomes the famous Genest and Ghoudi's generator [43] behind copula 4.2.15 in [5]. Thanks to the Lorenz approach, it is immediate to study the properties of the associated copula. First notice that the new generator corresponds to the mirrored Lorenz curve of a Lomax random variable [4]. The Lomax distribution has a lower bound at zero and no upper bound, hence the associated copula is absolutely continuous and never tail independent. Moreover, looking at the behavior of the density in zero and using Corollary 5.1, one can conclude that the copula will be tail dependent for every choice of the parameter (the Lomax is indeed known as a lognormal-like distribution [4]). Finally, by noting that the Lomax family is star ordered with respect to the parameter θ, the associated family of copulas is stochastically ordered, as noted by Genest and Ghoudi [43]. In particular, Proposition 5.6 guarantees that the family is LTD (PK and PQD) ordered.

Table 5.1 summarizes and extends the results presented so far, listing some Lorenz generators and the properties of the related Lorenz copulas.

5.5. CONCLUSIONS

An alternative approach to the generation of non-strict bivariate Archimedean copulas has been proposed using the Lorenz curve, a powerful tool in the study of socio-economic inequality[3, 4, 22, 24] and risk management [26, 44–46]. The main advan-

tages of the Lorenz-generation of copulas are threefold. First, the great number of Lorenz curves available in the literature allows for the generation of a large amount of copulas, which include existing cases, but also brand new ones. Second, every Lorenz copula can be easily characterized looking at some basic features of the non-negative finite-mean variable X underlying the Lorenz generator. In particular, it has been shown that quantities like the Kendall's τ and properties like upper tail dependence and stochastic dominance can be inferred from X. Third, the possibility of importing into the world of copulas many of the results developed in the studies of inequality allows for many interesting considerations: from a novel perspective on the Gini index, as a measure of the distance of a Lorenz copula from its Fréchet-Hoeffding bounds, to the possibility of generating multiparametric copulas using some useful compositions rules for Lorenz curves.

Regarding the last point, it is interesting to notice that the opposite direction works as well. It is in fact possible to borrow tools from the theory of copulas and to apply them to the study of socio-economic inequality. Consider for example the Kendall's τ.

In terms of Lorenz curve, one has

$$\tau = 1 + 4 \int_0^1 \frac{L(p)\mu}{F^{-1}(p)} \mathrm{d}x. \tag{5.36}$$

Now, observe that $U(p) = \int_0^p \frac{\mu}{F^{-1}(t)} \mathrm{d}t$ is an increasing function. Therefore one can rewrite (5.36) as

$$\tau = 1 + 4 \int_0^1 L(p) \mathrm{d}U(p) = 4 \int_0^1 \frac{pF^{-1}(p) - L(p)\mu}{F^{-1}(p)} \mathrm{d}p - 1. \tag{5.37}$$

Following [47], the quantity τ as per Equation (5.37) is a valid inequality index with weighting function U. In particular, τ measures the distance between the Lorenz curve $L(p)$ and the line of perfect equality $L_{PE}(p) = p$, and in this it is similar to the Gini index. However, differently from the standard Gini, τ weights both $L(p)$ and $L_{PE}(p)$ for the actual value of wealth, as represented by the quantile function $F^{-1}(p)$. As a consequence, τ could be used for direct comparisons among countries, something not immediately possible using the Gini index, given its scale free nature [24].

To conclude, as far as future work is concerned, it would be interesting to investigate the possibility of extending the Lorenz approach to d-dimensional copulas, with $d \geq 3$. Viable solutions could be the exploitation of nested constructions [5], or the use of multivariate Lorenz curves, as for example the Lorenz zonoid of [48]. For this second direction, however, one needs to remember that there exist more definitions of multivariate Lorenz curves [4, 27], and there is no guarantee that they may all work for the purpose.

Generator	Underlying X	Also known as	Tail Dependence	Singular	Kendall's τ Range	Order
1. $1-x$	Perfect equality	M	TI	No	1	–
2. $\Phi(\Phi^{-1}(1-x)-\sigma)$	Lognormal	–	AD	No	$[-1,1]$	$*$
3. $(1-x)+2gx\log(x)$	Shifted Exponential	–	$\begin{cases} TD & \text{if } g=1/2 \\ TI & \text{if } g<1/2 \end{cases}$	No	$[-1,0.227]$	$*$
4. $1-x^{1-\frac{1}{a}}$	Pareto	Non-strict Clayton	TI	No	$[-1,0]$	$*$
5. $\frac{2a(1-x)+(b-a)(1-x)^2}{a+b}$	$U[a,b]$	–	$\begin{cases} TD & \text{if } a=0, b>0 \\ TI & \text{if } a>0, b>0 \end{cases}$	$\frac{a+b}{2b}$	$\begin{cases} 0 & \text{if } a=0, b>0 \\ [-1,0) & \text{if } 0<a<b<\infty \end{cases}$	$*$
6. $(1-x)^{\eta}(1-x^{\theta})^{\gamma}$	$\begin{cases} \text{Pareto} & \text{if } \eta=0, \gamma=1 \\ \text{Singh-Maddala} & \text{if } \eta=0 \\ \text{See [41]} & \text{otherwise} \end{cases}$	$\begin{cases} 4.2.15 & \text{if } \eta=0, \gamma=\frac{1}{\theta} \\ – & \text{otherwise} \end{cases}$	$\begin{cases} \text{TI} & \text{if } \eta=0, \gamma=1 \\ \text{TD} & \text{otherwise} \end{cases}$	No	$[-1,1]$	See [41]

Table 5.1: Some examples of Lorenz copulas with their properties. Legend: TI = tail independence, AD = asymptotic tail dependence, TD = tail dependence, $*$ = star order (which then implies the other orders as discussed in Section 5.1.3), – not available.

APPENDIX

The appendix collects some results removed from the main narrative of the chapter for the sake of space and readability.

SUBSECTION 1.3- STAR ORDER EQUIVALENCE

When dealing with the star order, one can show the equivalence between the standard definition in terms of quantile functions [27, 29] and Definition 5.2 via the Lorenz curve.

Recall that two non-negative random variables X_1 and X_2 are star ordered, i.e. $X_1 \succ^* X_2$, when the ratio of their quantile functions

$$\frac{F_2^{-1}(p)}{F_1^{-1}(p)}, \quad p \in [0,1], \tag{5.38}$$

is non-increasing in p. Or, equivalently when the ratio $\frac{F_1^{-1}(p)}{F_2^{-1}(p)}$ is non-decreasing.

By Definition 5.2, $X_1 \succ^* X_2$ when $L_1(L_2^{-1}(x))$ is convex.

For a generic Lorenz curve $L(p)$, associated to a non-negative random variable X with distribution function F and mean $\mu < \infty$, one has $L'(p) = \frac{F^{-1}(p)}{\mu}$. Now, assume that L_1 and L_2 are both twice differentiable. Then

$$L_2'(L_1^{-1}(p)) = \frac{F_2^{-1}(L_1^{-1}(p))}{\mu_2} \frac{\mu_1}{F_1^{-1}(L_1^{-1}(p))} \quad p \in [0,1]. \tag{5.39}$$

By the convexity of $L_2(L_1^{-1}(x))$, the previous equation is equivalent to $\frac{F_1^{-1}(x)}{F_2^{-1}(x)}$ being non-decreasing, since μ_1 and μ_2 are always positive, and $L_i^{-1}(p)$, $i = 1,2$, is a map from $[0,1]$ to itself.

SUBSECTION 3.2 - SHIFTED EXPONENTIAL LORENZ COPULA

Consider the Lorenz curve of the shifted exponential distribution, i.e.

$$L(p) = p + 2g(1-p)\log(1-p). \tag{5.40}$$

The mirrored Lorenz is easily obtained as $\bar{L}(p) = L(1-p)$, while its inverse is $\bar{L}^{-1}(y) = \left(1 - L^{-1}(y)\right) 1_{y \in [0,1]}$.

In order to get $L^{-1}(y)$ some manipulations are needed, starting from $L(p) = y$. In particular,

$$\begin{aligned}
-e^{\log(1-p)} + 2g e^{\log(1-p)} \log(1-p) &= y - 1 \\
e^{\log(1-p)} \left(\log(1-p) - \frac{1}{2g} \right) &= \frac{y-1}{2g} \\
\frac{e^{\frac{1}{2g}}}{e^{\frac{1}{2g}}} e^{\log(1-p)} \left(\log(1-p) - \frac{1}{2g} \right) &= \frac{y-1}{2g} \\
e^{\log(1-p) - \frac{1}{2g}} \left(\log(1-p) - \frac{1}{2g} \right) &= \frac{y-1}{2g e^{\frac{1}{2g}}}
\end{aligned}$$

By setting $z = \log(1-p) - \frac{1}{2g}$, and noting that $z < -1$ for every choice of $p \in [0,1]$ and $g \in [0, \frac{1}{2}]$, one gets

$$e^z z = \frac{y-1}{2g e^{\frac{1}{2g}}}. \tag{5.41}$$

Using the Lambert function W, defined as $f^{-1}(xe^x) = W(xe^x)$, Equation (5.41) becomes

$$z = W_{-1}\left(\frac{y-1}{2g e^{\frac{1}{2g}}}\right), \tag{5.42}$$

where the lower branch W_{-1} is chosen being $z < -1$ [36]. Recalling that $z = \log(1-p) - \frac{1}{2g}$, the inverse of the shifted exponential Lorenz is

$$L^{-1}(y) = \exp\left(W_{-1}\left(\frac{y-1}{2g e^{\frac{1}{2g}}}\right) + \frac{1}{2g}\right) - 1. \tag{5.43}$$

The rest of the derivation is straightforward: one needs to combine Equations (5.40) and (5.43) according to Definition 5.3.

Notice that the maximum appearing in Equation (5.29) takes care of the fact that $\bar{L}(u) + \bar{L}(v)$ may be larger than 1.

SUBSECTION 3.3 - STAR ORDER AND PARETO RANDOM VARIABLES

Assume that $X_1 \sim \text{Par}(x_m, \alpha_1)$ and $X_2 \sim \text{Par}(x_m, \alpha_2)$, with $1 < \alpha_1 < \alpha_2$. One can then show that $X_1 \succ^* X_2$.

Consider the quantile function of a Pareto random variable, $F^{-1}(p) = x_m(1-p)^{-\frac{1}{\alpha}}$. Equation (5.38) becomes

$$\frac{F_2^{-1}(p)}{F_1^{-1}(p)} = \frac{x_m(1-p)^{-\frac{1}{\alpha_2}}}{x_m(1-p)^{-\frac{1}{\alpha_1}}} = (1-p)^{-\frac{1}{\alpha_2}+\frac{1}{\alpha_1}}, \tag{5.44}$$

which is decreasing for every $\alpha_1 < \alpha_2$. Hence we can conclude that the shape parameter α orders Pareto random variables in the star sense.

Proposition 5.6 then guarantees that the Pareto Lorenz copulas C_1 and C_2 associated to X_1 and X_2 are such that $C_1 \succ^{\text{LTD}} C_2$.

REFERENCES

[1] A. Fontanari, P. Cirillo, and C. W. Oosterlee, *Lorenz-generated bivariate archimedean copulas,* Available at SSRN 3391514 (2019).

[2] A. B. Atkinson, *On the measurement of inequality,* Journal of economic theory **2**, 244 (1970).

[3] I. Eliazar, *A tour of inequality,* Annals of Physics **389**, 306 (2018).

[4] C. Kleiber and S. Kotz, *Statistical size distributions in economics and actuarial sciences*, Vol. 470 (John Wiley & Sons, 2003).

[5] R. B. Nelsen, *An introduction to copulas* (Springer, 2007).

[6] M. O. Lorenz, *Methods of measuring the concentration of wealth,* Publications of the American statistical association **9**, 209 (1905).

[7] A. Sklar, *Fonctions de repartition an dimensions et leurs marges,* Publications de l'Institut Statistique de l'Université de Paris **8**, 229 (1959).

[8] M. Kendall, *Multivariate analysis* (Charles Griffin & Company Ltd, 1975).

[9] T. Mikosch, *Copulas: Tales and facts,* Extremes **9**, 3 (2006).

[10] C. Alsina, B. Schweizer, and M. J. Frank, *Associative functions: triangular norms and copulas* (World Scientific, 2006).

[11] M. Grossman and R. Katz, *Non-Newtonian Calculus* (Lee Press, 1972).

[12] J. Meginniss, *Non-newtonian calculus applied to probability, utility and bayesian analysis,* Proceedings of the American Statistical Association: Business and Economic Statistics Section **1**, 405 (1980).

[13] A. J. McNeil and J. Nešlehová, *Multivariate archimedean copulas, d-monotone functions and 1-norm symmetric distributions,* The Annals of Statistics **37**, 3059 (2009).

[14] M. Flores, E. de Amo Artero, F. Durante, and J. Sánchez, *Copulas and Dependence Models with Applications: Contributions in Honor of Roger B. Nelsen* (Springer, 2017).

[15] J. Avérous and J.-L. Dortet-Bernadet, *Dependence for archimedean copulas and aging properties of their generating functions,* Sankhyā: The Indian Journal of Statistics , 607 (2004).

[16] E. Di Bernardino and D. Rullière, *A note on upper-patched generators for archimedean copulas,* ESAIM: Probability and Statistics **21**, 183 (2017).

[17] S. Koenig, H. Kazianka, J. Pilz, and J. Temme, *Estimation of nonstrict archimedean copulas and its application to quantum networks,* Applied Stochastic Models in Business and Industry **31**, 464 (2015).

[18] C. Genest and L.-P. Rivest, *Statistical inference procedures for bivariate archimedean copulas,* Journal of the American statistical Association **88**, 1034 (1993).

[19] P. Capéraa, A.-L. Fougeres, and C. Genest, *A stochastic ordering based on a decomposition of kendall's tau,* in *Distributions with given marginals and moment problems* (Springer, 1997) pp. 81–86.

[20] A. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools: Concepts, Techniques, and Tools* (Princeton University Press, 2015).

[21] J. L. Gastwirth, *A general definition of the lorenz curve,* Econometrica: Journal of the Econometric Society , 1037 (1971).

[22] I. Eliazar and M. H. Cohen, *Hierarchical socioeconomic fractality: The rich, the poor, and the middle-class,* Physica A: Statistical Mechanics and its Applications **402**, 30 (2014).

[23] C. Gini, *Measurement of inequality of incomes,* The Economic Journal **31**, 124 (1921).

[24] S. Yitzhaki and E. Schechtman, *The Gini methodology: A primer on a statistical methodology* (Springer, 2012).

[25] B. C. Arnold, *The lorenz curve: evergreen after 100 years,* in *Advances on income inequality and concentration measures* (Routledge, 2001) pp. 34–46.

[26] A. Fontanari, P. Cirillo, and C. W. Oosterlee, *From concentration profiles to concentration maps. new tools for the study of loss distributions,* Insurance: Mathematics and Economics **78**, 13 (2018).

[27] B. C. Arnold and J. M. Sarabia, *Majorization and the Lorenz order with applications in applied mathematics and economics* (Springer, 2018).

[28] A. Balbás, J. Garrido, and S. Mayoral, *Properties of distortion risk measures,* Methodology and Computing in Applied Probability **11**, 385 (2008).

[29] M. Shaked and J. G. Shanthikumar, *Stochastic orders* (Springer, 2007).

[30] J. M. Sarabia, *A general definition of the leimkuhler curve,* Journal of Informetrics **2**, 156 (2008).

[31] P. Muliere and M. Scarsini, *A note on stochastic dominance and inequality measures,* Journal of Economic Theory **49**, 314 (1989).

[32] A. Charpentier and J. Segers, *Tails of multivariate archimedean copulas,* Journal of Multivariate Analysis **100**, 1521 (2009).

[33] J. K. Patel and C. B. Read, *Handbook of the normal distribution*, Vol. 150 (CRC Press, 1996).

[34] S. S. Wang, *A class of distortion operators for pricing financial and insurance risks,* Journal of Risk and Insurance **67**, 15 (2000).

[35] E. Pap, *Pseudo-analysis,* IFAC Proceedings Volumes **34**, 743 (2001), 1st IFAC Symposium on System Structure and Control 2001, Prague, Czechoslovakia, 27-31 August 2001.

[36] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Vol. 55 (Courier Corporation, 1965).

[37] R. D. Gupta and D. Kundu, *Theory & methods: Generalized exponential distributions,* Australian & New Zealand Journal of Statistics **41**, 173 (1999).

[38] D. Chotikapanich, *Modeling income distributions and Lorenz curves*, Vol. 5 (Springer Science & Business Media, 2008).

[39] J. M. Sarabia, *Parametric lorenz curves: Models and applications,* in *Modeling income distributions and Lorenz curves* (Springer, 2008) pp. 167–190.

[40] C. Genest, K. Ghoudi, and L.-P. Rivest, *Understanding relationships using copulas, by edward frees and emiliano valdez, january 1998,* North American Actuarial Journal **2**, 143 (1998).

[41] J.-M. Sarabia, E. Castillo, and D. J. Slottje, *An ordered family of lorenz curves,* Journal of Econometrics **91**, 43 (1999).

[42] S. K. Singh and G. S. Maddala, *A function for size distribution of incomes,* in *Modeling income distributions and Lorenz curves* (Springer, 2008) pp. 27–35.

[43] C. Genest and K. Ghoudi, *Une famille de lois bidimensionnelles insolite,* Comptes Rendus - Academie des Sciences Paris Série 1 **318**, 351 (1994).

[44] A. Fontanari, I. Eliazar, P. Cirillo, and C. Oosterlee, *Portfolio risk and quantum majorization of correlation matrices,* SSRN Doi: 10.2139/ssrn.3309585 (2019), http://dx.doi.org/10.2139/ssrn.3309585.

[45] H. Shalit and S. Yitzhaki, *Mean-gini, portfolio theory, and the pricing of risky assets,* Journal of Finance **39**, 1449 (1984).

[46] H. Shalit and S. Yitzhaki, *The mean-gini efficient portfolio frontier,* Journal of Financial Research **28**, 59 (2005).

[47] F. Mehran, *Linear measures of income inequality,* Econometrica: Journal of the Econometric Society , 805 (1976).

[48] G. Koshevoy and K. Mosler, *The lorenz zonoid of a multivariate distribution,* Journal of the American Statistical Association **91**, 873 (1996).

6

CONCLUSION AND FUTURE WORK

6.1. SUMMARY OF THE THESIS

In this thesis, we applied the theory behind the mathematical object called Lorenz curve, to solve a variety of problems related to quantitative risk managements. Specifically, credit risk management, in Chapters 2 and 3, portfolio risk and dependence in Chapters 4 and 5.

In particular, the leitmotif throughout the thesis has been the understanding of the Lorenz curve as a transformation of positive random variables highlighting their variability properties through the convexification of the quantile function: the more curvature in the Lorenz curve, the higher the dispersion of the underlying random variable.

We now summarize the conclusions drawn in each of the previous chapters.

6.1.1. GINI BASED RISK MEASURES: CONCENTRATION PROFILE

In Chapter 2 we tackled the problem of measuring the tail variability of the loss distribution and the reliability of the Expected Shortfall. We proposed to study the Lorenz curve associated to the truncated losses random variable, where the truncation level is given by the Value-at-Risk. In order to summarize the *continuum* of information conveyed by such a Lorenz curve, we used the Gini index of the Lorenz curve at each truncation level.

We obtained a new tool, called Concentration Profile, which maps each Value-at-Risk for a given probability level α into the Gini index of the corresponding VaR_α-truncated losses random variable. In such a way, we built a measure that, as the Lorenz curve, characterizes the loss random variable but is easier to interpret in terms of tail variability. From then built a graphical tool, called Concentration map, which can be used to compare portfolio losses once a specific risk preference has been inputted. With the Concentration Adjusted Expected Shortfall we were then able to summarize the information of both the Expected Shortfall and the Concentration Profile into just one measure. Finally, with the Concentration Profile being a measure of tail behaviour, we explored the possibility of using such a transformation in the context of Extreme Value Theory as a measure of tail threshold selection. In particular, our measure shows comparable results

with the most common non-parametric techniques for threshold selection such as the Hill plot and the Mean-excess plot.

6.1.2. GINI ESTIMATOR UNDER INFINITE VARIANCE

In this chapter, we addressed the problem of consistency of the non-parametric estimator for the Gini index when the underlying stochastic environment presents heavy tails, particularly when the finiteness of the second moment of the losses random variable is not guaranteed. Understanding the behaviour of the Gini index estimator is crucial also for the Concentration Profile introduced in Chapter 2.

The heavy tails setting poses important issues of convergence of the non-parametric estimator of the Gini index, since the usual Central Limit Theorem result with Gaussian limit cannot be applied.

By exploiting the Generalized Central Limit Theorem, we showed how the limiting distribution for the non-parametric estimator of the Gini index exists and it is an α-stable random variable totally skewed to the right. This observation is quite important in practice, since the loss of symmetry of the limiting distribution translates to a less robust estimator and to a bias in the pre-asymptotic behaviour of the Gini non-parametric estimator.

Finally, we built a naive bias-correction term that can be added to the non-parametric Gini estimator to improve its performances in finite samples.

6.1.3. QUANTUM MAJORIZATION FOR FINANCIAL CORRELATION MATRICES

An ordering on the space of market correlation matrices, called *quantum majorization*, has been defined exploiting the concept of majorization for symmetric positive semidefinite matrices. Recalling results from quantum mechanics, we built the financial correlation risk analogue of the Von Neumann entropy order for density matrices.

Additionally, we showed how most of the measures commonly used in finance to synthesize the risk embedded in a correlation matrix are isotonic with respect to quantum majorization providing an additional argument to use such order in finance.

We also argue that the presence of quantum majorization itself should be checked when dealing with market correlations, since its presence is a strong signal of risk in the market.

Finally, a version of the Lorenz curve based on a unitary transformation of the correlation matrix has been developed and used to test the presence of the quantum majorization order between the spectrum of historical correlation matrices for the Industrial Dow Jones index.

6.1.4. LORENZ-GENERATED ARCHIMEDEAN COPULAS

The Lorenz copula has been proposed as class of non-strict, bivariate Archimedean copulas. In particular, leveraging on the convex distortion interpretation of the Lorenz curve, we used the decreasing convex dual of the Lorenz curve, the mirrored Lorenz curve, as generator for Archimedean copulas.

Using this construction we were able to study most of the fundamental properties of such generated copulas simply by looking at features of the random variable associated to the Lorenz curve. For example, the tail behaviour of the Lorenz curve was linked to

the behaviour of the lower bound of the underlying random variable and multivariate stochastic ordering properties of the copula to the univariate variability orders such as the star order and the Lorenz order.

6.2. FUTURE DIRECTIONS

As we have tried to sketch this thesis, the Lorenz curve is an extremely powerful object, with several interpretations, from socio-economic variability measure to convex distortion of the identity map.

Here below, we collect some ideas for future works.

6.2.1. LORENZ CURVE AND PICKANDS

In multivariate Extreme Value Theory, in particular in the bivariate case, an important role is played by the so-called extreme value copulas [1].

Extreme value copulas arise as the possible limit of copulas of component-wise maxima of i.i.d. (or strongly-mixing stationary) random sequences [2]. These copulas prove to be very useful in risk management, to model joint extremes, or when dealing with data characterized by positive and possibly asymmetric dependence.

In the bivariate case, a copula C is an extreme value copula if and only if it can be represented as

$$C\left(y^{1-u}, y^{u}\right) = y^{P(u)}, \qquad y, u \in [0,1], \tag{6.1}$$

where $P : [0,1] \to [1/2, 1]$ is a convex function, called the Pickands dependence function, satisfying

$$\max(u, 1-u) \leq P(u) \leq 1.$$

When $P(u) = 1$ for every $u \in [0,1]$, the random variables Y and Z are completely tail independent. Conversely, when $P(u) = u$ or $P(u) = 1-u$, we are in the case of perfect tail dependence. For other values of $P(u)$, different degrees of tail dependence are reached [2, 3].

The Pickands dependence function is therefore a convex distortion of the case of independence and the degree of dependence induced by the copula in Equation (6.1) increases as the Pickands dependence function turns towards its lower bound.

In view of the interpretation of the Lorenz curve as a convex distortion, it comes natural to apply the Lorenz curve to generate new parametric models for Pickand dependence functions and thus for extreme value copulas.

A possibility could be to apply a 45-degrees rotation to the Lorenz curve and a scaling by a factor $\sqrt{2}$ in order to match the space of the Lorenz curves with the space of the Pickands dependence functions. Such operation would lead to the following expression for the Pickands dependence function, in terms of Lorenz curve $L(p)$:

$$P(u) = 1 + u - K^{-1}(u), \tag{6.2}$$

with $K^{-1}(u)$ the inverse of $K(p) = (L(p) + p)/2$.

When dealing with the Pickands function for bivariate dependence, a common quantity of interest is the *Coefficient of upper tail dependence* λ_U [3], defined as

$$\lambda_U = 2\left(1 - P\left(\frac{1}{2}\right)\right). \tag{6.3}$$

λ_U takes values in $[0,1]$ and it relates to the vertical distance between the independence case, defined as $P(u) = 1$ for all $u \in [0,1]$, and the middle point of the Pickands function $P\left(\frac{1}{2}\right)$. It reaches the value $\lambda_U = 0$ if the underlying bivariate random vector is tail independent, and $\lambda_U = 1$ when the underlying bivariate random vector is fully tail dependent.

The coefficient λ_U is therefore a simple way to check the degree of dependence of the random vector characterized by a given Pickands function. However, λ_U has different pitfalls, one of the biggest being that it is just a central measure: it gives the degree of dependence by evaluating the Pickands function just in its central point.

It is not difficult to see that, when the strongest dependence between Y and Z is in a direction different from the antidiagonal, thus resulting in an asymmetric Pickands function, the index λ_U may lead to wrong conclusions on the degree of dependence.

To overcome this type of problems, a more suitable statistics could be the entire area between the independence case and the Pickands function, suitably rescaled in order to stay between 0 and 1. Given the relation between the Pickands function $P(u)$ and the Lorenz curve $L(p)$, this would correspond to using the Gini index of the latter.

The advantage of using the Gini index is that its is scale and rotation invariant, therefore, once the Lorenz curve associated to a given bivariate vector is obtained, the associated Gini index will have exactly the same value of the one computed directly on the Pickands function.

Finally other indices build for the analysis of the Lorenz curve such as the vertical and horizontal maximal distances [4] could be used to study the asymmetry of the Pickands dependence function and its associated impact on tail dependence.

6.2.2. LORENZ CURVE AND COPULAS

In Chapter 5 we built Archimedean copulas by the use of the mirrored Lorenz curve as the generator function. However, it turns out that this is not the only way to generate copulas from Lorenz curves.

Here, we propose two other possible generative schemes: one based on the Kendall's distribution function, $K(v) := P(C(U,V) \leq v)$, and another based on the copula's diagonal, $\delta(t) := C(t,t)$.

It is known since the works of Genest [5, 6] that in the case of the Archimedean copula the Kendall distribution function uniquely recovers the generator of the copula $\phi(x)$ through the following formula:

$$\phi(x) = \exp \int_k^x \frac{1}{t - K(t)} \mathrm{d}t, \tag{6.4}$$

with $k \in (0,1)$ an arbitrary chosen constant. By definition, $K(t)$ is an increasing function with $K(1) = 1$, in particular, if the copula is absolutely continuous then $K(0) = 0$. Additionally, by the Fréchet bounds, $t \leq K(t) \leq 1$, $\forall t \in [0,1]$, Genest et al [5] also showed how, if the copula is Archimedean, the $K(t)$ is concave.

Increasing concave functions with the above properties are known in the context of wealth inequality studies as Leimkuhler curves $M(t)$ [7], which are related to the Lorenz curve through the following expression:

$$M(t) = 1 - L(1-t). \tag{6.5}$$

Given the above results, combining Equations (6.4) and (6.5) given a Lorenz curve, the generator of an absolutely continuous Archimedean copula can be recovered:

$$\phi(x) = \exp \int_k^x \frac{1}{t-1-L(1-t)} \mathrm{d}t \tag{6.6}$$

For example, given the Lorenz curve of the exponential distribution, see Section 2.8 for its construction, the generator of the independence copula $\phi(x) = -\log(x)$ is obtained.

Alternatively to the above method it is possible to associate a Kendall distribution function to a Lorenz curve by noting that $K(t) = L^{-1}(x)$, where $L^{-1}(x)$ is the continuous inverse of the Lorenz curve.

Finally, by a similar geometric argument it is possible to generate diagonal functions $\delta(t)$ using Lorenz curves. Also in this case, the Fréchet bounds provide the proper upper and lower bounds to frame $\delta(t)$ is terms of Lorenz curve and allow to interpret concentration of wealth bounds in terms of dependence bounds and vice-versa.

Recognizing the *lorenzian* structure of these copula features it is not only important because it unlocks the possibility of generate new models, but also because it allows to use the concentration indices in the copula context.

For example it is straightforward to show that in the case of the Kendall distribution function the Kendall's τ, a common measure of association, is proportional to the Gini index of the associated Lorenz curve.

The Gini index, however, is only one of the main indices that can be derived from the Lorenz curve. It would then be interesting to study what feature of the dependence are captured by different concentration indices and their respective probabilistic interpretation in terms of multivariate probabilities.

6.2.3. LORENZ CURVE AND DISTORTION RISK MEASURES

In risk management, *distortion measures* play an important role [8].

Distortion risk measures are a type of risk measures based on the Choquet integral. They originated from the works of Wang on fair insurance pricing [9–11]. A distortion risk measure for the loss random variable $X \sim F(x)$ is defined as:

$$\rho(X) = \int_\chi x \mathrm{d}g(F(x)) \tag{6.7}$$

where $g : [0,1] \to [0,1]$ is called the distortion measure.

From Equation (6.7) it is possible to observe that the role of the distortion measure g is to re-assign the probability weights on the outcome space χ of X. Note that if g is the identity map then no distortion occurs and Equation (6.7) reduces to the expectation of X, while if $g(x) = 0$, $\forall x \in [0,1)$, and $g(1) = 1$, $\rho(X) = \sup X$, which is also known as the *worst case risk measure* [12].

In [8], it has been proven that a distortion risk measure is coherent if and only if the distortion function g is increasing and convex and additionally any coherent risk measure can be written as (6.7) for some increasing convex function $g : [0,1] \to [0,1]$. From this result, it should be clear that the distortion functions that span coherent risk measures must be Lorenz curves of some underlying random variable and, additionally,

that the more spread out the random variable the more weight is assigned to tail events making the risk measure more conservative.

For example, consider the so-called Wang transform: [10]:

$$g(x) = \Phi(\Phi^{-1}(x) - a), \tag{6.8}$$

where Φ is the c.d.f of a standard normal distribution and $a \in \mathbb{R}$ is a parameter controlling the shape of g. If $a \geq 0$ the Wang transform is convex and thus coherent, while if $a \leq 0$ the Wang transform is concave. It can be shown that Equation (6.8) for $a \geq 0$ coincides with the Lorenz curve of a log-normal distribution with volatility a [13].

Recognizing the Lorenz structure of distortion functions can be useful when building such measures since the behaviour of the distortion is related to the c.d.f. of common distributions [14]. However, belonging to a different domain, the Lorenz curves and the concentration indices are usually not yet considered in the risk management literature.

6.2.4. OPTIMAL TRANSPORT AND THE LIFT ZONOID

In optimal transport theory, the aim is to allocate, or transport, resources from one point to another in the most efficient way, in other words, by minimizing some given cost function [15].

In a more general setting this problem can be framed in terms of *moving* mass, distributed according some distribution function $f(x)$ from a subset of $\mathbb{R}^d$, into another, possibly different subset of $\mathbb{R}^d$ with distribution function $g(x)$. Note that since the distribution functions integrate to one no mass is lost in the process.

It can be shown that for $d = 1$ under the L^2 distance the optimal transport plan to move $f(x)$ into $g(x)$ is given by the gradient of some convex increasing function [16]. Clearly, the Lorenz curve satisfies these conditions and so it can be understood as the building block for optimal transport plans when L^2 distance is used. However, when moving to higher dimensions, $d \geq 2$, finding optimal conditions for transport plans is not trivial and there is not, so far, a uniquely determined way of defining the transport plan itself [17].

Difficulties of operating in a higher dimensional setting have been encountered in the Lorenz related literature as well, because of the non-uniqueness of the notion of quantile. The problem seems to have been definitively solved only recently with the works of Mosler of the so-called Lift Zonoid [18].

Understanding the analogy between the optimal transport plan and Lorenz curve in the one-dimensional case leads to the following set of questions to be answered. Is the multidimensional Lorenz curve, also knows as the Lift Zonoid, an optimal transport map as well, and if so under which conditions can we study the properties of such transport plan by looking at the multivariate vector generating the multivariate Lorenz curve?

We believe that from this connection both fields of optimal transport and multivariate Lorenz could benefit generating even more material to work with.

REFERENCES

[1] G. Gudendorf and J. Segers, *Extreme-value copulas,* in *Copula theory and its applications* (Springer, 2010) pp. 127–145.

[2] J. A. Tawn, *Bivariate extreme value theory: models and estimation,* Biometrika **75**, 397 (1988).

[3] M. Falk, J. Hüsler, and R.-D. Reiss, *Laws of small numbers: extremes and rare events* (Springer Science & Business Media, 2010).

[4] I. Eliazar, *A tour of inequality,* Annals of Physics **389**, 306 (2018).

[5] C. Genest and L.-P. Rivest, *Statistical inference procedures for bivariate archimedean copulas,* Journal of the American statistical Association **88**, 1034 (1993).

[6] P. Capéraa, A.-L. Fougeres, and C. Genest, *A stochastic ordering based on a decomposition of kendall's tau,* in *Distributions with given marginals and moment problems* (Springer, 1997) pp. 81–86.

[7] B. C. Arnold and J. M. Sarabia, *Majorization and the Lorenz order with applications in applied mathematics and economics* (Springer, 2018).

[8] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition* (Princeton university press, 2015).

[9] S. S. Wang, V. R. Young, and H. H. Panjer, *Axiomatic characterization of insurance prices,* Insurance: Mathematics and economics **21**, 173 (1997).

[10] S. S. Wang, *A class of distortion operators for pricing financial and insurance risks,* Journal of risk and insurance **67**, 15 (2000).

[11] S. Wang, *Premium calculation by transforming the layer premium density,* ASTIN Bulletin: The Journal of the IAA **26**, 71 (1996).

[12] H. Föllmer and A. Schied, *Stochastic finance: an introduction in discrete time* (Walter de Gruyter, 2011).

[13] C. Kleiber and S. Kotz, *Statistical size distributions in economics and actuarial sciences.*, Vol. 470 (John Wiley & Sons, 2003).

[14] D. Chotikapanich, *Modeling income distributions and Lorenz curves*, Vol. 5 (Springer Science & Business Media, 2008).

[15] C. Villani, *Optimal transport: old and new*, Vol. 338 (Springer Science & Business Media, 2008).

[16] R. J. McCann *et al.*, *Existence and uniqueness of monotone measure-preserving maps,* Duke Mathematical Journal **80**, 309 (1995).

[17] C. de Valk and J. Segers, *Stability and tail limits of transport-based quantile contours,* arXiv preprint arXiv:1811.12061 (2018).

[18] G. Koshevoy and K. Mosler, *The lorenz zonoid of a multivariate distribution,* Journal of the American Statistical Association **91**, 873 (1996).

ACKNOWLEDGEMENTS

Writing this dissertation has been one of the greatest challenges I have dealt with so far in my career. I couldn't have done it without the help of numerous people I have met during this journey.

First of all I would like to thank my supervisors Prof. Dr. Ir. Kees Oosterlee and Dr. Pasquale Cirillo. Thank you for giving me the chance of undertaking this path in the first place, supporting me in every step and guiding me in my professional and personal growth.

Secondly, I would like to thank all the committee members: Prof. dr. F. H. J. Redig, Prof. dr. A. Pascucci, Prof. dr. M. Bonetti, Prof. dr. P. J. C. Spreij and Prof. dr. ir. M. H. Vellekoop for the useful comments, feedback and the time dedicated to read my final work. I would also like to give special thanks to Prof. Iddo Eliazar in Tel Aviv and Prof. Nassim Taleb in New York. I have really enjoyed our collaborations and I am looking forward to undertake new challenges together. Also special thanks to Prof. Sonia Petrone, Prof. Pietro Muliere and Prof. Simone Padoan from Bocconi University for supervising my Master thesis and introducing me for the first time to the world of probability and statistics.

From Bocconi University I would also like to mention Paolo, Isadora, Giulia and Stefano. Thank you for welcoming me in the big *Bocconi's PhD family* and guiding me in my first steps as PhD student.

I am grateful to the European Union, in particular to the Marie Sklodowska-Curie grant and the Horizon 2020 program for actually sponsoring my PhD program.

I also would like to express my gratitude to Dr. Diederik Fokkema for supervising me during my time at EY, introducing me to the *real world* challenges and to have always shown genuine interest in my work.

I would like to thank Ki Wai for the great support both scientific and psychological. I will always remember our *Math Sundays* in Amsterdam, your critical approach to my sloppy mathematical statements and our *Dim Sum*. You helped me a lot to become a better mathematician.

On the same line, special thanks also to Eni, Federico, Bruno, Mario, Francesca and Kai Lun, you really helped a lot through my entire PhD, I am extremely grateful to all of you.

Thanks to Anastasia for being an example of dedication, grit and effort.

A special thanks to Martina for being a trustworthy friend and bearing all my complaints and offering me so many after-lunch coffees!

Of course I cannot forget about Daniele Imperiale, you are a great friend, thank you for inspiring me everyday. I look up to you: *stay convex*!

A stand alone acknowledgement is needed also for Dr. Alessandra Cipriani. Thanks for all the Coffecompany weekends, the work on graphs, the mathematical discussions

and the life choices counseling. Thank you for taking time to helping me, teaching me and making me more self-confident of my math skills.

Thanks also to the PhD students and stuff members, past and present, of Applied Probability and Statistics in Delft and the CWI group for all the early lunches, the coffees, the discussions or just for being friendly smiley faces on rainy days.

During my PhD I had the chance to travel a lot, visit new places, meeting new people. Among them I would like to recall David from Tel Aviv, from the day we meet at Shneor Caffe you have always been on my side. Thank you for the numerous discussion about math and politics. And Sharon for all those nights outside Port Said. I am looking forward to visit you soon again.

Also, I would like to acknowledge the staff of the Coffecompany Delft, for providing me with the *usual* medium cappuccino and granola yogurt for the past 4 years and my *greek family* in Ios island of the Safran bar whose provided me with the best office view one can possibly imagine.

Living in the Netherlands has not always been easy, I would like to thank those people that ease my staying in this country starting with Matteo Bonica who followed me all the way from Trento with whom I shared an uncountable number of adventures from Delft, to Ios and hopefully in Melburne soon. Thanks also to Andrea (your Netflix account saved me so many times), Alfredo and Fabio (great times in Amsterdam), Marieke (for your directness), Mila (for helping me fighting the *Delftpression*) and Roos (for making me daydream warm places during cold winters and the legal advises). Unfortunately life sometimes takes different routes, I wish you all the luck and thank you again for spending some of your time with me.

Finally, I would have never been able to make it without the help of Nada, Evelyn, Dorothee, Stefanie and Carl that guided me in all the logistics of this PhD journey. Also, I have to acknowledge the help of Barbara, and the entire Bergonzoni Family for offering friendship and medical consulting when needed and Bella for helping with the dutch translation of this thesis' summary.

Thank you also to Daniele and Silvia Tomasi. You are like brother and sister to me.

Some final honorable mentions are needed for Chiara and Carlo de Battaglia, Matteo Martin, Giulo Rangoni, Stefano Waller, Francesco Marchetti and Tristano Slopm, thanks for making me feel home very time I come back to Trento and/or Milano. Finally thank you to my kick-boxing friends a big *Ooss* to you, and to the cats, swans (white or black) and stars (conditioning on dutch weather) that I spotted during these years always leaving me with a smile.

I am sure I have forgotten someone. I apologize, I will make it up to you with a beer next time I see you.

This thesis is dedicated to my parents, my mum Chiara and my father Maurizio, thank you.

CURRICULUM VITÆ

Andrea FONTANARI

14-10-1990 Born in Trento, Italy.

EDUCATION

2015–2019 PhD in Applied Mathematics
Delft University of Technology, Delft, The Netherlands
Thesis: Lorenz-based quantitative risk management
Promotors: Prof. dr. ir. C. W. Oosterlee, Dr. P. Cirillo

2013–2015 Master of Science in Economics
Bocconi University, Milan, Italy

2009–2013 Bachelor of Science in Economics
Bocconi University, Milan, Italy