



HAL
open science

The Gini coefficient and negative income: A probabilistic approach

Karim Kilani

► **To cite this version:**

Karim Kilani. The Gini coefficient and negative income: A probabilistic approach. 2022. hal-03911502

HAL Id: hal-03911502

<https://hal.science/hal-03911502>

Preprint submitted on 22 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Gini coefficient and negative income: A probabilistic approach

Karim Kilani

December 2022

Abstract

This paper discusses the Gini index when a mass of individuals in the population sharing the wealth suffers a loss. This question has been addressed in the statistical framework while here it is treated from a probabilistic point of view, which allows us to be as exhaustive as possible. The assumptions are the most general ones on probability distributions. We determine a condition on the Lorenz function that generates a Gini index that exceeds unity. We construct an adjusted Gini index, i.e. one that remains within limits between zero and one, according to a method advocated in the statistical framework. After having recovered the recommended index in the statistical case, we apply the method to continuous distributions, namely the uniform distribution and the generalized Pareto distribution, thus showing the operability of the method.

Keywords. Adjusted Gini index; Lorenz function; Negative income; Quantile function

JEL classification. C43; D63

1 Introduction

The treatment of inequality in the case of wealth distributions with a strictly negative component is rather sparsely addressed in the literature on this research question. However, the issue is quite relevant from an economic point of view, for example in the framework of a cost-benefit analysis aiming at evaluating policies where some individuals would suffer significant losses while others would make profits.

If we look at this problem by analyzing the usual Lorenz curve, the latter immediately decreases, and dramatically so if the negative component of the wealth distribution is significant, before rising again and ending up in the usual area where we expect it to be, just beneath the diagonal of perfect equality. As a result, the Gini index, which measures the area between that diagonal and the Lorenz curve, can exceed unity, which is quite detrimental to the applicability of this index. This happens when the integral of the negative part of the Lorenz function exceeds that corresponding to the positive part.

In this paper, we follow the proposal of Chen et al. [1982], who published a rather intuitive and simple way to adjust the Gini index in order to keep it within the proper limits, i.e. between zero and one. We generalize their approach, which was limited to the case of statistical data, to the probabilistic framework including any distribution with finite and strictly positive mean. Among other things, we propose a new interpretation of their adjusted index in terms of a ratio involving mathematical expectations of random variables of wealth rather than integrals of portions of the Lorenz function. We will also examine the case of some usual distributions in order to show the operational character of this approach, beyond the reservations that could be formulated about this adjusted index.

The rest of the paper is organized as follows: after having specified in section 2 the

framework and the very general assumptions on the wealth distributions we study, we recall in section 3 the definition of the Lorenz function as an integral of the quantile function, as well as the essential properties of the latter two functions that will be useful for the sequel. In the section 4, we remind the definition of the Gini index and we show that for distributions which have a fairly large negative part, this index can exceed unity, as has already been pointed out since the papers of Schutz [1951] or Chen et al. [1982]. We then explain, following the method advocated by Chen et al. [1982], how to adjust the Gini index by adding twice the area of the negative part of the Lorenz curve to the denominator of the Gini index. In section 5, we treat the case of discrete distributions and show that we do indeed recover the adjusted index of the statistical case as accurately derived by Berrebi and Silber [1985] in a very short note correcting the paper of Chen et al. [1982]. Finally, in the section 6, we treat two cases of continuous distributions: the uniform distribution where the values of the Gini index and the adjusted index are explicit and a generalization of the Pareto distribution where the Gini index is explicit while the adjusted index requires some basic numerical computations.

2 Assumptions

We consider a wealth distributed among the individuals of a population. It can be a discrete distribution, a continuous one, or even any mixture of these two types of distributions. The so-called distribution is modeled by means of a real random variable W defined on a probability space where P denotes its probability function.

In this paper, we allow that a part of the individuals in the population receives a nega-

tive income. However, we make the assumption that the mathematical expectation of W , the wealth, is finite and strictly positive $0 < \mu \equiv E(W) < \infty$. This means that we analyze the inequalities generated by the distribution of wealth in a population where gains more than compensate for losses.

3 The Lorenz Function

Let us recall the definition of the quantile function, widely used in the literature on inequalities, which is the generalized inverse of the cumulative distribution function $F(x) \equiv P(W \leq x)$. More formally, it is defined by [Parzen, 1979]

$$w(p) \equiv \sup \{x : F(x) \leq p\}, 0 \leq p < 1.$$

This is the maximum wealth held by $100 \times p\%$ of the least wealthy individuals. We recall that the quantile function is non-decreasing and continuous to the left [Resnick, 2008].

The Lorenz function for its part is defined as follows [Gastwirth, 1972, Arnold, 2015]

$$L(p) = \frac{1}{\mu} \int_0^p w(z) dz, 0 \leq p \leq 1. \quad (1)$$

This is the share of cumulative wealth held by $100 \times p\%$ of the least wealthy individuals.

It is clear that $L(0) = 0$ and that, since the integral of the quantile function is the mathematical expectation [see, e.g. Parzen, 2004], we have $L(1) = (1/\mu) \int_0^1 w(p) dp = 1$. Thanks to the fundamental theorem of analysis, L is a derivable function whose derivative is $L'(p) = w(p)/\mu, 0 \leq p \leq 1$. It is thus a convex function, decreasing in the interval

$[0, F(0)]$ and thus non-positive. On the other hand, it will be strictly increasing in the interval $(F(0), 1]$. There exists thus a unique proportion p^* in the last mentioned interval such that

$$L(p) \leq 0 \iff 0 \leq p \leq p^*. \quad (2)$$

By continuity, we therefore have

$$L(p^*) = 0. \quad (3)$$

The share of cumulative wealth held by individuals with a level of wealth below $w(p^*)$ is zero. Moreover, if the variable is almost surely non-negative, we will have $p^* = F(0)$. Otherwise, if the wealth distribution exhibits a non-zero mass negative part, we will have $p^* > F(0)$, with $L(p) < 0$ if $0 < p < p^*$ and $L(p) > 0$ if $p > p^*$. We can deduce that the positive and negative parts of the function of L which respectively verify

$$L^+(p) = \max(L(p), 0); L^-(p) = -\min(L(p), 0). \quad (4)$$

In the rest of the paper, we will compare the current distribution whose Lorenz function is L to two extreme distributions of wealth. The first is the one where wealth is distributed in a perfectly equitable way, i.e. all individuals receive the same wealth, and whose Lorenz function is given by

$$\bar{L}(p) \equiv p, 0 \leq p \leq 1. \quad (5)$$

Recall that the Lorenz function satisfies the following inequality: $L \leq \bar{L}$, formally proved in [Thompson, 1976, Theorem 1] which deals with the general case we are considering here, i.e. a possibly negative part for the wealth distribution but with a strictly pos-

itive mean. In general, this inequality is strict, except when the distribution concentrates all its mass in a single value, in which case there is equality over the whole probability interval.

The second limiting Lorenz function is the one where the entire wealth is held by an infinitesimal fraction of the individuals of the population, which is given by

$$\underline{L}(p) \equiv \begin{cases} 0, & 0 \leq p < 1; \\ 1, & 0 \leq p < 1. \end{cases} \quad (6)$$

4 The Gini index and its adjustment

One way of thinking of the Gini index, which is far from being unique, is the following: We seek to measure the distance that separates the Lorenz function L from the two extreme functions \bar{L} and \underline{L} given respectively by (5) and (6), and representing respectively a completely egalitarian and a totally unequal situation .

More formally, recall the definition of the \mathcal{L}^1 -norm which states

$$\|f\| \equiv \int_0^1 f(p)dp, \quad (7)$$

where f is any non-negative integrable function. The Gini index can be expressed with respect to this norm in the following way:

$$G \equiv \frac{\|\bar{L} - L\|}{\|\bar{L} - \underline{L}\|}. \quad (8)$$

It is thus the ratio between the distance which separates L from \bar{L} and the distance separat-

ing \underline{L} and \bar{L} . Note that $G \geq 0$, with equality only in the case of an almost-surely constant wealth distribution.

Using the expressions for the extreme Lorenz functions seen in the above section, we can rewrite the Gini index in a more usual form

$$G = 2 \int_0^1 [p - L(p)] dp = \frac{2}{\mu} \int_0^1 w(p) p dp - 1. \quad (9)$$

The second form is obtained through an integration by parts which will prove useful in the rest of the paper.

Another way to write the Gini index is to make use of the positive and negative parts of the Lorenz function. Using also the fact that $\|\bar{L} - \underline{L}\| = \|\bar{L}\| - \|L^+\| + \|L^-\|$, we deduce the following alternative writing

$$G \equiv \frac{\|\bar{L}\| - \|L^+\| + \|L^-\|}{\|\bar{L}\|}.$$

We conclude that a necessary and sufficient condition that positions the value of G with respect to unity is given by:

$$G \lesseqgtr 1 \iff \|L^-\| \lesseqgtr \|L^+\|. \quad (10)$$

Thus, it may happen that the wealth distribution admits a negative part that has excess mass, generating a very high level of inequality, so that the Gini index even goes beyond unity.

A reasonable idea would be to modify the denominator of the Gini index given in its

form (8) to obtain an adjusted index that remains within the usual range, i.e. between 0 and 1. The revised index could be designed as follows

$$G^* \equiv \frac{\|\bar{L} - L\|}{\|\bar{L}\| + \|L^-\|}. \quad (11)$$

Since the difference between the denominator and numerator of the expression on the right-hand side of the above equation is $\|L^+\|$, it is clear that we do have $G^* \leq 1$.

Equivalent and arguably more operational ways of expressing this adjusted Gini index are as follows

$$G^* = \frac{G}{1 + 2\|L^-\|} = 1 - \frac{2\|L^+\|}{1 + 2\|L^-\|}. \quad (12)$$

This adjustment is in every respect equivalent to the one suggested by Schutz [1951] and formalized, but only in the statistical case, by Chen et al. [1982], which we shall verify in the next section.

In the same vein as the expression for the Gini index obtained in (9), both these alternative expressions for the integral for the additional area used in the index adjustment will be useful in the following

$$\|L^-\| = - \int_0^{p^*} L(p) dp = \frac{1}{\mu} \int_0^{p^*} w(p) p dp. \quad (13)$$

Let us consider again the expression (9) which gives the Gini index, in the second integral form. By performing a change of variable, we obtain

$$G = \frac{1}{\mu} \int_0^1 w(\sqrt{p}) dp - 1.$$

Since $w(\sqrt{p})$ is nothing else than the quantile function of the maximum of two variables W_1, W_2 , independent and of the same law as W ,¹ so that

$$G = \frac{E(\max(W_1, W_2)) - \mu}{\mu}$$

By noticing that $|W_1 - W_2|/2 = \max(W_1, W_2) - (W_1 + W_2)/2$, we deduce this other expression of the Gini index

$$G = \frac{E(|W_1 - W_2|/2)}{\mu}. \quad (14)$$

In this form, the index is interpreted as the average of the mean absolute half-difference between the wealth of two individuals taken at random from the population, divided by the mean wealth. If we now consider the expression for the adjusted index given by (12), it can also be written, by combining it with the latest equation highlighted, as

$$G^* = \frac{E(|W_1 - W_2|/2)}{\mu + 2 \int_0^{p^*} w(p) p dp}.$$

By applying a change of variable to the integral that appears in the denominator of the above equation, we can write this expression in the following way

$$G^* = \frac{E(|W_1 - W_2|/2)}{\mu + 2(p^*)^2 \int_0^1 w(p^* \times p) p dp}.$$

We notice that $w(p^* \times p)$ corresponds to the quantile of the truncation of the variable W to the right of $w(p^*)$, i.e. the variable where we retain only the realizations which satisfy the

¹Since the cumulative distribution function of $\max(W_1, W_2)$ is given by $F^2(x)$, the inequality $F^2(x) \leq p$ becomes equivalent to the following: $F(x) \leq \sqrt{p}$.

condition $W \leq w(p^*)$. By denoting W^* this truncated variable, we end up with

$$G^* = \frac{E(|W_1 - W_2|/2)}{\mu + (p^*)^2 E(\max(W_1^*, W_2^*))}, \quad (15)$$

where W_1^*, W_2^* , are two independent random variables having the same distribution as W^* . Thus, the adjusted index adds to the mathematical expectation of wealth in the denominator of the Gini index the weighted mathematical expectation of the maximum wealth of two individuals taken at random independently from the subpopulation of less wealthy individuals whose gains and losses just about even out.

In Expression (15), the weighting of the mathematical expectation by the square of p^* may seem surprising at first sight. It is due to the fact that we draw two truncated random variables which must be less than $w(p^*)$. The probability of obtaining them corresponds precisely to this weighting.

5 The case of discrete distributions

Let us consider an ordered list of $n \geq 2$ wealth data $w_1 \leq \dots \leq w_n$ and their associated (nonzero) probabilities p_1, \dots, p_n , this list could be finite or infinite (in which case $n = \infty$).

Throughout this section, we will assume that the mean wealth is strictly positive $\mu = \sum_1^n w_i p_i > 0$, in accordance with our assumptions in the previous sections.

If W is the random variable that takes the values w_1, \dots, w_n , with associated probabilities p_1, \dots, p_n , the Lorenz function corresponding to that discrete variable is

$$L(p) = \sum_1^i y_j + \left(\frac{p - \pi_i}{p_i} \right) y_i, \quad \pi_{i-1} \leq p < \pi_i, \quad i = 1, \dots, n, \quad (16)$$

where $y_i \equiv p_i w_i / \sum_1^n p_j w_j$ is the share of total wealth held by individuals with a wealth value of w_i while $\pi_i \equiv \sum_1^i p_j$ are cumulative probabilities. It goes without saying that we have $\sum_i^n y_i = 1$. In order for our notations to be completely defined, we assume that a sum whose lower bound is greater than its upper bound is zero.

To obtain the Gini index, one may use the expression of the index G in its second integral form which appears in (9), which gives us

$$G = 2 \sum_1^n \int_{\pi_{i-1}}^{\pi_i} (y_i / p_i) p dp = \sum_1^n \gamma_i - 1, \quad (17)$$

where

$$\gamma_i \equiv (2\pi_i - p_i)y_i, \quad i = 1, \dots, n.$$

Let us then define the unique index $k \geq 0$ that verifies the following condition²

$$\left(\sum_1^k y_i \leq 0 \right) \wedge \left(\sum_1^{k+1} y_i > 0 \right). \quad (18)$$

Necessarily, p^* lies within the interval $[\pi_k, \pi_{k+1})$. Solving (3) with L given by (16), we find an explicit value for this critical probability given by

$$p^* = \pi_k - \frac{\sum_1^k y_i}{y_{k+1}} p_{k+1}. \quad (19)$$

We can now calculate an expression for the term that adjusts the Gini index using (13)

²In the case where the distribution is strictly positive, i.e. when $w_1 > 0$, we will arbitrarily set, preserving the hypothesis that a sum where the upper and lower bounds are reversed, $k = 0$. In this case, the value of p^* defined in (19) will be zero.

with the value of p^* calculated just above, to obtain

$$2\|L^-\| = \sum_1^k \gamma_i + \left(\sum_1^k y_i \right) \left(\frac{\sum_1^k y_i}{y_{k+1}} p_{k+1} - 2 \sum_1^k p_i \right). \quad (20)$$

We conclude that the adjusted Gini index given by (12) for discrete distributions is, by combining the expressions (17) and (20), equal to the following ratio

$$G^* = \frac{\sum_1^n \gamma_i - 1}{1 + \sum_1^k \gamma_i + (\sum_1^k y_i) \left[(\sum_1^k y_i / y_{k+1}) p_{k+1} - 2 \sum_1^k p_i \right]}. \quad (21)$$

In the statistical case treated by Chen et al. [1982] and which corresponds to the particular case where $p_i = 1/n$ and $\gamma_i = (2i - 1)y_i/n$, $i = 1, \dots, n$, with n finite, we obtain, by replacing these values in (21), the expression below

$$G^* = \frac{(2/n) \sum_1^n iy_i - 1/n - 1}{1 + (2/n) \sum_1^k iy_i + (1/n) (\sum_1^k y_i) \left[\sum_1^k y_i / y_{k+1} - (2k + 1) \right]}. \quad (22)$$

This expression corresponds precisely to the one found by Berrebi and Silber [1985], who corrects an error that crept into the formula of a footnote of the paper of Chen et al. [1982].

As a very simple illustration, we use the data provided in Schutz [1951], taken up by Chen et al. [1982], on the wealth of ten families listed as $(-500, -300, -300, -100, 200, 300, 300, 400, 500, 500)$, with an average value of 100.

The Gini index, by applying formula (17), is thus: $G = 2.94 - 1 = 1.94$ and exceeds unity in this case. It is clear, from the sixth row of the table where $\sum_1^6 y_j = 0$, that we have, according to (18), $k = 6$; by applying (19), it follows: $p^* = \pi_6 = 0.8$. Therefore, $1 + 2\|L^-\| = 1 + \sum_1^6 \gamma_i = 1 + 1.14 = 2.14$, so that $G^* = 1.94/2.14 = 0.9065$.

Table 1: Computation of standard and adjusted Gini indices for Schutz data

i	w_i	p_i	π_i	y_i	$\sum_1^i y_j$	γ_i	$\sum_1^i \gamma_j$
1	-500	0.1	0.1	-0.5	-0.5	-0.05	-0.05
2	-300	0.2	0.3	-0.6	-1.1	-0.24	-0.29
3	-100	0.1	0.4	-0.1	-1.2	-0.07	-0.36
4	200	0.1	0.5	0.2	-1.0	0.18	-0.18
5	300	0.2	0.7	0.6	-0.4	0.72	0.54
6	400	0.1	0.8	0.4	0.0	0.60	1.14
7	500	0.2	1.0	1.0	1.0	1.80	2.94

Figure 1: The Lorenz curve corresponding to the Schutz data

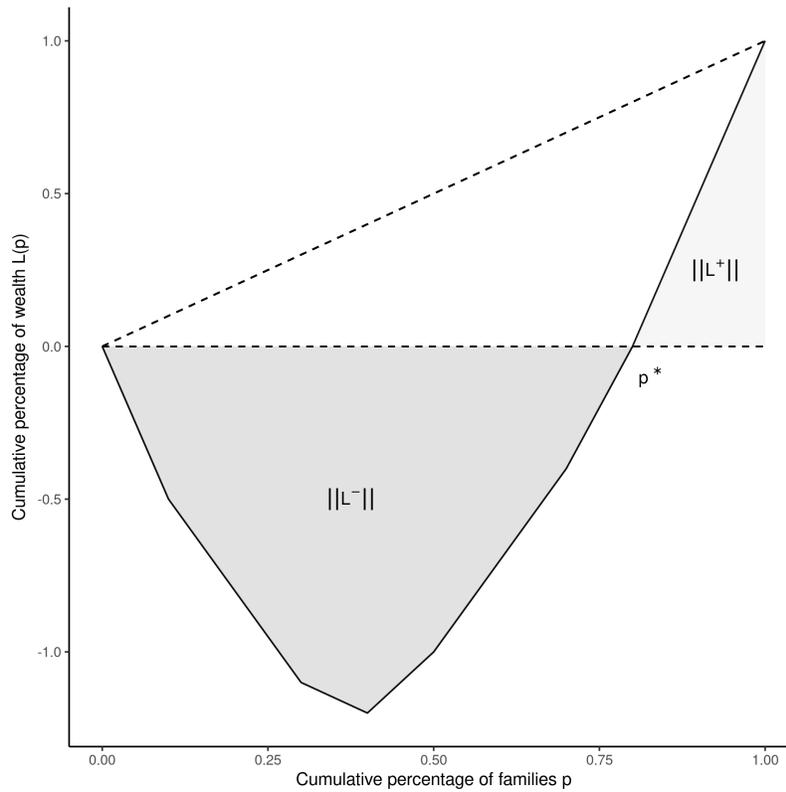


Figure 1 shows the Lorenz curve for the Schutz data. The area of the negative part,

$\|L^-\|$, which is shaded in dark gray, is much larger than the one corresponding to the positive part, $\|L^+\|$, shaded in light gray. The curve crosses the x -axis at $p^* = 0.8$.

6 Some continuous distributions

We will now provide some illustrations for continuous distributions with negative support.

Uniform distribution

We analyze the uniform law on the interval $[-a, 2-a]$, where $0 \leq a < 1$, whose Lorenz function is

$$L(p) = \frac{p(p-a)}{1-a}, \quad 0 \leq p \leq 1 \quad (23)$$

It has a negative part on the interval $(0, a)$, and we have $p^* = a$. We obtain the following Gini index

$$G = \frac{1}{3(1-a)}. \quad (24)$$

It decreases with a , from infinity near -1 to $1/3$ in the neighborhood of $a = 0$, passing by 1 when $a = 1/3$. Noticing that

$$\|L^-\| = \frac{1}{1-a} \int_0^a (2p-a)pdp = \frac{a^3}{6(1-a)},$$

the adjustment of the Gini index is as follows

$$G^* = \frac{1}{3(1-a) + a^3}. \quad (25)$$

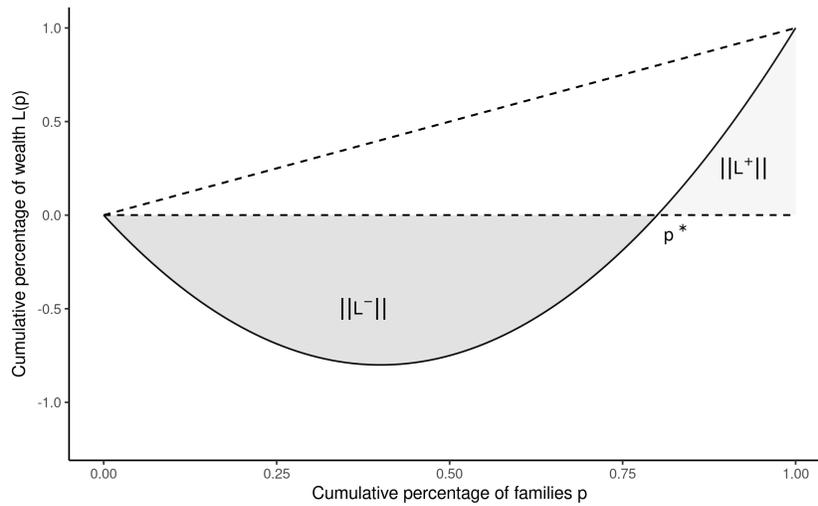
The following table provides values for the Gini index and the adjusted index.

Table 2: Standard and adjusted Gini index for the uniform distribution on $[-a, 2 - a]$

a	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
G	0.33	0.41	0.50	0.62	0.78	1.00	1.33	1.89	3.00	6.33
G^*	0.33	0.37	0.42	0.47	0.54	0.62	0.71	0.80	0.90	0.97

Figure 2 shows the Lorenz curve of a uniform distribution of wealth taken in the interval $(-0.8, 1)$ and allocated, this time, among a continuum of families. The curve is this time a piece of parabola with a rather pronounced negative part, as for the Schutz data, and becomes positive from $p^* = 0.8$.

Figure 2: Lorenz curve of the uniform distribution on $(-0.8, 1)$



Generalized Pareto distribution

We now consider a generalization of the Pareto distribution whose Lorenz function is of the following form:

$$L(p) = p - \frac{1-p}{1-a} \left[\frac{(1-p)^{-\xi} - 1}{\xi} \right], \quad 0 \leq p \leq 1, \quad (26)$$

where $0 \leq a < 1$ is a scaling parameter whose interpretation is provided below, and where $\xi < 1$ as a shape parameter.

We assume that the mean of the underlying wealth distribution is $1 - a$, without loss of generality, to ensure a simple understanding of the parameters of the distribution. Indeed, since $L'(0) = -a/(1-a) < 0$, the value $-a$ represents the lower bound on the support of the distribution. The distribution contains a positive and a negative part. However, we have to consider three different patterns depending on the scale parameter of the distribution:

1. $0 \leq \xi < 1$. The support is $[-a, \infty)$;
2. If $\xi = 0$, by applying l'Hospital's rule, it is an exponential distribution with as Lorenz function: $L(p) = p + (1-p) \ln(1-p)/(1-a)$, $0 < p < 1$;
3. $\alpha < 0$. The support is in this case bounded and given by: $[-a, 1 - a - 1/\xi]$.

For all these cases, the Gini index amounts to

$$G = \frac{1}{(1-a)(2-\xi)}. \quad (27)$$

The value of p^* as a nonzero solution of $L(p^*) = 0$ cannot be obtained explicitly, as in the case of the uniform distribution discussed above. It can only be evaluated numerically.

The Lorenz function given by (26) satisfies the following differential equation

$$L(p) = p - \frac{(1-p)L'(p)}{1-\xi} - \frac{a(1-p)}{(1-a)(1-\xi)}, 0 \leq p \leq 1.$$

Recalling that the adjustment area can be seen in two alternative forms in (13), we obtain the following equation

$$-||L^-|| = \frac{(p^*)^2}{2} + \frac{||L^-||}{1-\xi} - \frac{a}{(1-a)(1-\xi)} \left[p^* - \frac{(p^*)^2}{2} \right].$$

Solving the latter equation, we obtain

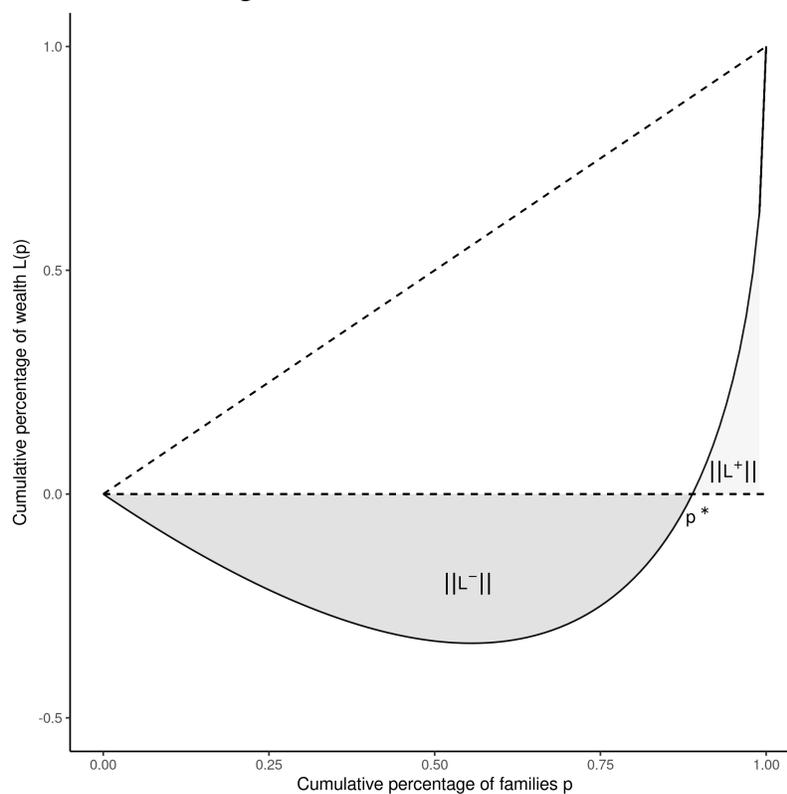
$$2||L^-|| = \frac{p^*[2a - (1-\xi + a\xi)p^*]}{(1-a)(2-\xi)},$$

Thus, summarizing the previous results, we find the following adjusted Gini index

$$G^* = \frac{1}{(1-a)(2-\xi) + p^*[2a - (1-\xi + a\xi)p^*]}. \quad (28)$$

Figure 3 corresponds to the Lorenz curve of a generalized Pareto distribution with $a = 0.5$ and $\xi = 0.5$, with a Gini index equal to $G = 4/3$. The value where the curve crosses the x-axis, computed numerically, is $p^* = 0.8889$. This allows us to find the adjusted Gini index of $G^* = 0.9557$.

Figure 3: Lorenz curve of the generalized Pareto distribution with $a = 0.5$ and $\xi = 0.5$



References

B. C. Arnold. *Pareto distributions*. Chapman and Hall/CRC, 2015.

Z. M. Berrebi and J. Silber. The Gini coefficient and negative income: A comment. *Oxford Economic Papers*, 37(3):525–526, 1985.

C.N. Chen, T.W. Tsaur, and T.S. Rhai. The gini coefficient and negative income. *Oxford Economic Papers*, 34(3):473–478, 1982.

- J. L. Gastwirth. The estimation of the lorenz curve and gini index. *The review of economics and statistics*, 54(3):306–316, 1972.
- E. Parzen. Nonparametric statistical data modeling. *Journal of the American statistical association*, 74(365):105–121, 1979.
- E. Parzen. Quantile probability and statistical data modeling. *Statistical Science*, 19(4): 652–662, 2004.
- S. I. Resnick. *Extreme values, regular variation and point processes*. Springer, 2008.
- R. R. Schutz. On the measurement of income inequality. *The American Economic Review*, 41(1):107–122, 1951.
- W. A. Thompson. Fisherman’s luck. *Biometrics*, 32(2):265–271, 1976.