Statistica & Applicazioni - Vol. XIV, n. 2, 2016, pp. 123-144

A COMPARISON OF INCOME DISTRIBUTIONS MODELS THROUGH INEQUALITY CURVES

Alberto Arcagni*
Francesco Porro*

SUMMARY

In this paper the features of some distribution models used in the literature to depict income distributions are analysed. The analysis is based on the inequality curves generated by such models. In particular, the role of the parameters related to the inequality curves is investigated, also by considering the influence of their variations from a pointwise perspective.

Keywords: *Income Distribution, Inequality, Partial Orders, Inequality Curves.*

1. INTRODUCTION

The analysis of the inequality plays an important role in the domain of socio-economics sciences. Originally, it has been introduced for the study of income distributions, but later on, it has been widely used in several research areas (see for example Jacobson, Milman, Kammen, 2005; Naeem, 2009; Bhattacharya, Ghosh, Mande, 2015). In the literature many tools for evaluating the inequality has been proposed. The investigation of their features is still interesting for applications, but also from the methodological point of view.

Statistical models can be used for the approximation of empirical distributions, in order to obtain information from the data by the parameter estimates and their interpretation. Therefore the analysis of how the distribution parameters affect the inequality measures and curves is useful to achieve a deeper comprehension of the social phenomena and economical conditions. This knowledge may be essential in supporting the decisions of policy makers.

It is not possible to compile an exhaustive list of papers in the literature, dealing with the inequality analysis, among others: Gini (1914), Dagum (1977), Dagum (1980), Dancelli (1989), Zenga (2007), McDonald and Ransom (2008), Garcia and Alaiz (2011), and Arcagni and Porro (2014).

The main purpose of the present paper is an investigation about how the parameters affect the inequality in the most used income distribution models, with particular attention on which areas of such distributions are more sensitive to parameter

* Dipartimento di Statistica e Metodi Quantitativi - Università degli Studi di Milano-Bicocca - Via Bicocca degli Arcimboldi, 8 - 20126 - MILANO (e-mail: alberto.arcagni@unimib.it; ✉: francesco.porro1@unimib.it).

variations. A particular attention will be paid to the tails of the distribution, since the left one represents the low incomes, and the right one the high incomes. For this reason a pointwise analysis of inequality allows the identification of the population areas more involved by any economic policy.

The present paper is organized as follows: in Section 2 the considered distribution models are presented; Section 3 is devoted to the inequality tools: inequality curves, related indices and partial orders. Section 4 presents the definition and the features of the Zenga-84 inequality curve, while Section 5 provides for each distribution model, several plots of the two Zenga inequality curves with different values of the parameters, and the description of the role played by them. In Section 6 some final remarks are discussed.

2. THE DISTRIBUTION MODELS

In order to perform the aforementioned analysis, some distribution models will be considered. The models used in the present investigation are:

- the Pareto Type II distribution depending on the three parameters $\mu > 0, \sigma > 0$, and $\alpha > 1$ (in order to have non-negative support and finite expectation), with probability density function (see for further details Arnold, 2008):

$$f(x;\alpha,\sigma,\mu) = \begin{cases} \dfrac{\alpha}{\sigma}\left[1+\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-(\alpha+1)} & x > \mu \\ 0 & \text{otherwise;} \end{cases} \tag{1}$$

- the generalized lognormal distribution depending on the three parameters $\theta_1 \in R, \theta_2 > 0, \alpha \geq 0$, with probability density function (see for details Brunazzo and Pollastri, 1986; Vianelli, 1982):

$$f(x;\alpha,\theta_1,\theta_2) = \begin{cases} \frac{(\alpha+1)^{\frac{\alpha}{\alpha+1}}}{2\Gamma\left(\frac{1}{\alpha+1}\right)\theta_2 x}\exp\left(-\frac{1}{\alpha+1}\left|\frac{\log x-\theta_1}{\theta_2}\right|^{\alpha+1}\right) & x > 0 \\ 0 & \text{otherwise;} \end{cases} \tag{2}$$

- the Singh-Maddala distribution, also known in the literature as Burr distribution, depending on the three parameters $a > 0, b > 0$ and $q > 1/a$ (in order to have finite expectation), with probability density function (refer to McDonald and Ransom, 2008; Singh and Maddala, 1976):

$$f(x;b,a,q) = \begin{cases} aqx^{a-1}b^{-a}[1+(x/b)^a]^{-(q+1)} & x > 0 \\ 0 & \text{otherwise;} \end{cases}$$

- the Dagum distribution (proposed in Dagum, 1977), depending on the three parameters $\theta > 0, \lambda > 0; \delta > 1$ (in order to have finite expectation), with probability density function:

$$f(x;\theta,\lambda,\delta) = \begin{cases} \theta\lambda\delta x^{-\delta-1}(1+\lambda x^{-\delta})^{-(\theta+1)} & x > 0 \\ 0 & \text{otherwise;} \end{cases}$$

- the quite recent Zenga distribution (proposed in Zenga, 2010, and investigated in Zenga *et al.*, 2010; Zenga *et al.*, 2011), depending on the three parameters $\mu > 0, \alpha > 0$, and $\theta > 0$, with probability density function:

$$f(x;\mu,\alpha,\theta) = \begin{cases} \dfrac{1}{2\mu\beta(\alpha,\theta)}\left(\dfrac{x}{\mu}\right)^{-3/2}\displaystyle\int_0^{\frac{x}{\mu}} k^{\alpha-1/2}(1-k)^{\theta-2}\,dk & 0 < x < \mu \\ \dfrac{1}{2\mu\beta(\alpha,\theta)}\left(\dfrac{x}{\mu}\right)^{-3/2}\displaystyle\int_0^{\frac{\mu}{x}} k^{\alpha-1/2}(1-k)^{\theta-2}\,dk & x > \mu, \end{cases}$$

where $\beta(\alpha,\theta)$ denotes the Beta function, defined by:

$$\beta(\alpha,\theta) = \int_0^1 t^{\alpha-1}(1-t)^{\theta-1}, \qquad \alpha > 0, \theta > 0.$$

It is well-known that the distribution parameters can be classified by their roles. From the inequality point of view, an important class of parameters is represented by the scale parameters. The following definition characterizes such class.

DEFINITION 1 (Scale parameter)
Let $\{F_\alpha,\ \alpha > 0\}$ be a family of distribution functions. Then α is a scale parameter of such family if

$$F_\alpha(x) = F_1\left(\frac{x}{\alpha}\right).$$

Since the inequality does not depend on the scale parameters, it is useful to identify them. For this reason in the following, the Pareto Type II distribution will be considered, using the alternative parametrization given by the replacement of μ/σ with θ. The probability density function (1) becomes then:

$$f(x;\alpha,\sigma,\theta) = \begin{cases} \dfrac{\alpha}{\sigma}\left[1+\left(\dfrac{x}{\sigma}-\theta\right)\right]^{-(\alpha+1)} & x > \theta\sigma \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0, \sigma > 0$, and $\alpha > 1$. Using this new parametrization, σ is a scale parameter, therefore it does not affect the inequality.

3. THE INEQUALITY CURVES AND THE RELATED INDICES

In the following, the most used inequality curves are considered. It is worth noting that they can be defined through the mean of the whole population and the means of specific groups.

Given a population and a random variable X evaluated on it, for each $p \in (0,1)$, the population can be split into two non overlapping groups: the first one, called *lower group* that consists of the proportion p of people with the lowest values of X, and the second one called *upper group* composed by all the others. Once the population is split into the lower and the upper group, the means of X in these two groups can be computed, obtaining the *lower* and the *upper mean*. The two following definitions delineate these two means.

DEFINITION 2 (Lower mean)
Let X be a continuous random variable, with distribution function F, and support $[a, b]$, where $0 \leq a < b \leq +\infty$. For any $p \in [0, 1]$, the lower mean $\bar{M}_{(p)}$ is defined as

$$\bar{M}_{(p)} = \begin{cases} \frac{1}{p}\int_0^p F^{-1}(t)dt & p \in (0, 1] \\ a & p = 0. \end{cases}.$$

DEFINITION 3 (Upper mean)
Let X be a continuous random variable, with distribution function F, and support $[a, b]$, where $0 \leq a < b \leq +\infty$. For any $p \in [0, 1]$, the upper mean $\overset{+}{M}_{(p)}$ is defined as

$$\overset{+}{M}_{(p)} = \begin{cases} \frac{1}{1-p}\int_p^1 F^{-1}(t)dt & p \in [0, 1) \\ b & p = 1. \end{cases}$$

In the Definitions 2 and 3 the lower mean and the upper mean have been extended by continuity in $p = 0$ and in $p = 1$, respectively. It is easy to verify that for a random variable X with expected value μ, the following formula holds true:

$$\mu = p\bar{M}_{(p)} + (1-p)\overset{+}{M}_{(p)}, \qquad \forall p \in [0, 1],$$

with the convention that whether the support of X is not finite:

$$(1-p)\overset{+}{M}_{(p)} = 0 \qquad \text{if } p = 1.$$

Now the definitions of some inequality curves can be presented.

DEFINITION 4
Let X be a non-negative continuous random variable, with positive and finite expected value μ, and distribution function F. Then:

- *the Lorenz curve of X is (see Lorenz, 1905; Gastwirth, 1971):*

$$L_X(p) = \frac{1}{\mu}\int_0^p F^{-1}(t)dt = \frac{p\bar{M}_{(p)}}{\mu}, \qquad p \in [0, 1];$$

- *the $\delta(p)$ curve of X, proposed by Gini, is (see Zenga, 1984, 1987; Basulto and Busto Guerrero, 2010):*

$$\delta_X(p) = \frac{\log(1-p)}{\log\left(1 - \frac{1}{\mu}\int_0^p F^{-1}(t)dt\right)} = \frac{\log(1-p)}{\log\left(1 - \frac{p\bar{M}_{(p)}}{\mu}\right)}, \qquad p \in (0, 1);$$

- *the $\lambda(p)$ curve of X, proposed by Zenga, is (see Zenga, 1984):*

$$\lambda_X(p) = 1 - \frac{\log\left(1 - \frac{1}{\mu}\int_0^p F^{-1}(t)dt\right)}{\log(1-p)} = 1 - \frac{\log\left(1 - \frac{p\bar{M}_{(p)}}{\mu}\right)}{\log(1-p)}, \qquad p \in (0,1);$$

- *the Gini curve of X is (see Zenga, 1984):*

$$G_X(p) = 2\left[p - \frac{1}{\mu}\int_0^p F^{-1}(t)dt\right] = 2p\left[1 - \frac{\bar{M}_{(p)}}{\mu}\right], \qquad p \in [0,1];$$

- *the Bonferroni curve of X is (see Bonferroni, 1930):*

$$B_X(p) = \frac{1}{p\mu}\int_0^p F^{-1}(t)dt = \frac{\bar{M}_{(p)}}{\mu}, \qquad p \in (0,1];$$

- *the reversed Bonferroni curve of X is:*

$$V_X(p) = 1 - \frac{1}{p\mu}\int_0^p F^{-1}(t)dt = 1 - \frac{\bar{M}_{(p)}}{\mu}, \qquad p \in (0,1];$$

- *the Zenga I(p) curve of X is (see Zenga, 2007):*

$$I_X(p) = 1 - \frac{(1-p)\int_0^p F^{-1}(t)dt}{p\int_p^1 F^{-1}(t)dt} = 1 - \frac{\bar{M}_{(p)}}{\overset{+}{M}_{(p)}}, \qquad p \in (0,1).$$

Remark 1
It is worth noting that from the definitions it follows that $\forall p \in (0,1)$:
- $\delta_X(p) = \frac{\log(1-p)}{\log(1-L_X(p))}$
- $\lambda_X(p) = 1 - \frac{1}{\delta_X(p)}$
- $G_X(p) = 2[p - L_X(p)]$;
- $B_X(p) = \frac{L_X(p)}{p}$;
- $V_X(p) = 1 - B_X(p)$;
- $I_X(p) = \frac{p - L_X(p)}{p[1-L_X(p)]}$.

An important feature that some inequality curves share is about uniformity: some curves among the aforementioned ones are constant for some kinds of the distribution model. More in details, in Zenga (1984) it is shown that both the $\delta(p)$ and the $\lambda(p)$ curves of X are constant if X is a Pareto Type II with $\mu = \sigma$ or alternatively $\theta = 1$ (i.e. the classical Pareto Type I), and in Polisicchio (2008) it is proved that the Zenga $I(p)$ curve is constant for a Pareto Type I with a particular restriction on the domain, and parameter $\alpha = 0.5$.

Starting from the aforementioned inequality curves, the inequality degree can be summarized in a single value, by using a synthetic measures. The definitions of some well-known indices are stated in the following.

DEFINITION 5

Let X be a non-negative continuous random variable, with positive and finite expected value μ, and let $L_X(p)$, $\delta_X(p)$, $\lambda_X(p)$, $G_X(p)$,$B_X(p)$, $V_X(p)$, and $I_X(p)$ be the corresponding Lorenz, $\delta(p)$, $\lambda(p)$, Gini, Bonferroni, reversed Bonferroni, and Zenga curves, respectively. Then:

- *the Gini index of X is (see Gini, 1914):*

$$R_X = 1 - 2\int_0^1 L_X(p)dp = \int_0^1 G_X(p)dp;$$

- *the δ index of X is (see Zenga, 1987):*

$$\delta_X = \int_0^1 \delta_X(p)dp;$$

- *the λ index of X is (see Zenga, 1984):*

$$\lambda_X = \int_0^1 \lambda_X(p)dp;$$

- *the Bonferroni index of X is (see Bonferroni, 1930):*

$$B_X = 1 - \int_0^1 B_X(p)dp = \int_0^1 V_X(p)dp;$$

- *the Zenga index of X is (see Zenga, 2007):*

$$I_X = \int_0^1 I_X(p)dp.$$

Figure 1 shows the behaviour of the inequality curves in the square $[0,1]^2$. The graph of the $\delta_X(p)$ curve lies out of such square, therefore its modification, the $\lambda_X(p)$ curve, is preferable. For this reason, in the remaining of the paper, the $\delta(p)$ curve is not analized. It is important to note that the four inequality indices R_X, δ_X, B_X and I_X correspond to the area beneath $G_X(p)$, $\delta_X(p)$, $V_X(p)$, and $I_X(P)$, respectively. Hence such four curves are more intuitive, since they satisfy the ‘‘increasing ranking’’ as described in Arcagni and Porro (2014).

Figures 1 also shows that some of the aforementioned inequality curves have some special points, which always belong to their graph, whatever the underlying distribution model is. The Lorenz curve is always equal to 0 if $p = 0$ and equal to 1 if $p = 1$. The Gini curve takes on value zero at $p = 0$ and $p = 1$. If $p = 1$, the Bonferroni curve always takes value 1, the reversed Bonferroni value 0, and

$$\lim_{p\to 0^+} B_X(p) = \frac{a}{E(X)} \qquad \text{and} \qquad \lim_{p\to 0^+} V_X(p) = 1 - \frac{a}{E(X)}.$$

The $\lambda(p)$ curve satisfies

$$\lim_{p\to 0^+} \lambda_X(p) = 1 - \frac{a}{E(X)},$$

where a is the lower bound of the support of the random variable X. Also the Zenga $I(p)$ curve has not such special points, since the values assumed for both the extreme values of p depend on the support of the distribution model. As proved in Polisicchio (2008), such dependence is given by:

$$\lim_{p\to 0^+} I_X(p) = 1 - \frac{a}{E(X)} \qquad \text{and} \qquad \lim_{p\to 1^-} I_X(p) = 1 - \frac{E(X)}{b}, \tag{3}$$

where a and b denote the lower and the upper bound of the support of the random variable X, respectively. If the upper bound b of the support of the random variable X (with positive and finite expectation) is not finite, the second limit in (3) becomes:

$$\lim_{p\to 1^-} I_X(p) = \lim_{b\to +\infty} 1 - \frac{E(X)}{b} = 1.$$

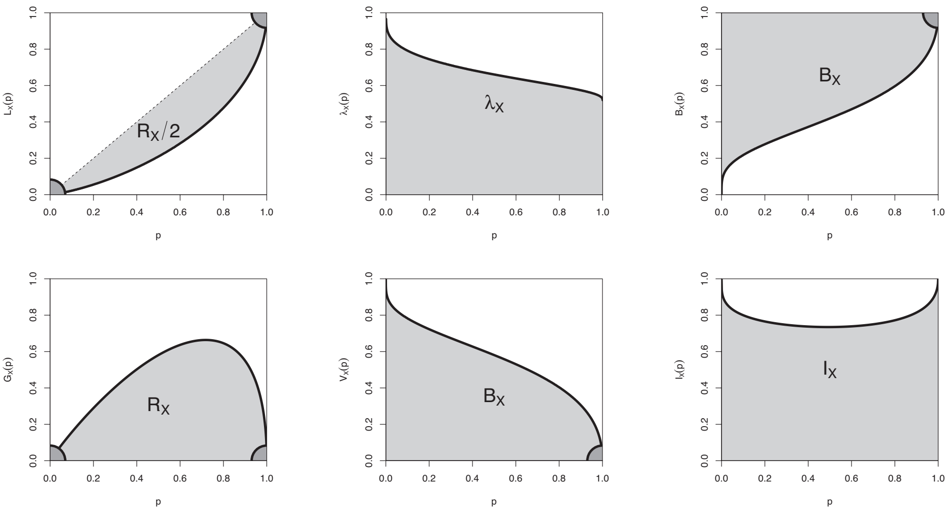

FIGURE 1. - *Inequality curves and related synthetic indices*

REMARK 2

At this point a very brief historical review may be usefull. Gini (1914) proposed to associate at each $p \in (0, 1)$ the relative concentration measure of a given variable X

$$\rho_X(p) = \frac{p - L_X(p)}{p},$$

that is the relative difference of the Lorenz curve $L_X(p)$ with respect to p. As synthetic inequality measure he suggested the weighted mean of the $\rho_X(p)$ with weights p, that is:

$$R_X = \frac{\int_0^1 \rho_X(p) \cdot p\, dp}{\int_0^1 p\, dp} = \frac{1}{1/2} \int_0^1 \frac{p - L_X(p)}{p} \cdot p\, dp = 2 \int_0^1 \Big(p - L_X(p) \Big)\, dp,$$

which corresponds to the formula of the Gini index, previously provided. Some years later Bonferroni (1930) proposed a concentration diagram based on the comparison between the lower mean $\bar{M}_{(p)}$ and the total mean μ of the variable X. For Bonferroni the inequality raises as the relative difference $V_X(p) = \frac{\mu - \bar{M}_{(p)}}{\mu}$ increases. As synthetic measure, he introduced the Bonferroni index

$$B_X = 1 - \int_0^1 B_X(p)\, dp = \int_0^1 V_X(p)\, dp = \int_0^1 \frac{\mu - \bar{M}_{(p)}}{\mu}\, dp,$$

by averaging the quantities $V(p)$. A decade after, De Vergottini (1940) showed the equivalence $V_X(p) = \rho_X(p)$, i.e.:

$$\frac{\mu - \bar{M}_{(p)}}{\mu} = \frac{p - L_X(p)}{p}.$$

As consequence, he showed that

$$R_X = \frac{1}{1/2} \cdot \int_0^1 \frac{\mu - \bar{M}_{(p)}}{\mu} \cdot p\, dp.$$

Following the same approach of De Vergottini, many other synthetic inequality indices proposed in the literature can be seen as weighted mean of the quantities $V_X(p)$, with different weights (for further details, see for example Amato, 1987; Zenga, 1990).

The inequality indices provide a total order in the set of the non-negative random variables, since, given two non-negative random variables, it is always possible to compute the values of one index and to compare such values.

The comparison of the inequality curves does not provide a total order, but only a partial one: the inequality curves of two random variables can intersect. It is not difficult to observe that the partial order based on one inequality curve ($L(p), G(p), \lambda(p), B(p), V(p)$ or $I(p)$) is coherent with the total order based on the related synthetic inequality index (R, λ, B, or I). Further details about these partial orderings are described in the next subsection.

An important result that originates from the inequality curves regards the partial orders induced by them. Here the definitions.

DEFINITION 6
Let X and Y be two non-negative random variables, both with finite and positive expectation. X is said to be larger (or more unequal) than Y in the

- *partial order based on the Lorenz curve, ($X \geq_L Y$), if*
$$L_X(p) \leq L_Y(p), \text{ or equivalently, } G_X(p) \geq G_Y(p), \qquad \forall p \in (0,1)$$
- *partial order based on the $\lambda(p)$ curve, ($X \geq_\lambda Y$), if*
$$\lambda_X(p) \geq \lambda_Y(p), \qquad \forall p \in (0,1)$$
- *partial order based on the Bonferroni curve, ($X \geq_B Y$), if*
$$B_X(p) \leq B_Y(p), \text{ or equivalently, } V_X(p) \geq V_Y(p), \qquad \forall p \in (0,1)$$
- *partial order based on the Zenga $I(p)$ curve, ($X \geq_I Y$), if*
$$I_X(p) \geq I_Y(p), \qquad \forall p \in (0,1).$$

The partial orders based on inequality curves are related also with other kinds of orderings. The next one is the definition of a well-known ordering, particularly useful in the domain of risk assessment.

DEFINITION 7 (Stochastic dominance of the second order)
Let X and Y be two non-negative random variables, with distribution functions F_X and F_Y, respectively. X is said to be second-order stochastically dominant over Y ($X \geq_2 Y$), if
$$\omega_X(z) = \int_{-\infty}^{z} F_X(t)dt \leq \int_{-\infty}^{z} F_Y(t)dt = \omega_Y(z), \qquad \forall z \in \mathbb{R}.$$

All the aforementioned partial orders are related, since the following Lemma holds (for further details see Shaked and Shanthikumar, 2007; Polisicchio and Porro, 2009).

LEMMA 1
Let X and Y be two non-negative random variables, both with finite and positive expectation. Then:
$$X \geq_L Y \Leftrightarrow X \geq_\lambda Y \Leftrightarrow X \geq_B Y \Leftrightarrow X \geq_I Y.$$
Moreover if $E(X) = E(Y)$, then it holds that:
$$X \geq_L Y \Leftrightarrow X \geq_\lambda Y \Leftrightarrow X \geq_B Y \Leftrightarrow X \geq_I Y \Leftrightarrow X \leq_2 Y.$$

The Figure 2 shows some examples of the relationships involved in the Lemma 1. Three random variables X, Y and W with expected value equal to 1 are represented through their probability density functions (f_X, f_Y, and f_W), their distribution functions (F_X, F_Y, and F_W), and functions ω as introduced in Definition 7. After that, their inequaliy curves are drawn.

Using the aforementioned partial orders, it can be observed that the random variable W is less unequal than (and stochastically dominant over) X and Y, while the random variables X and Y cannot be compared, since their inequality curves intersect. Nevertheless, it is worth to observe that $I_X(p) < I_Y(p)$ for values of p close to 1. Taking into account that the behaviour of the Zenga $I(p)$ curve in a left neighbourhood of 1 is ruled by the second limit in (3), it follows that the right tail of the probability density function of X is ''less heavy'' than the one of Y. Vice versa, for low values of p it can be observed that $I_X(p) > I_Y(p)$, meaning that the left tail of the density function of X is ''heavier'' than the one of Y.

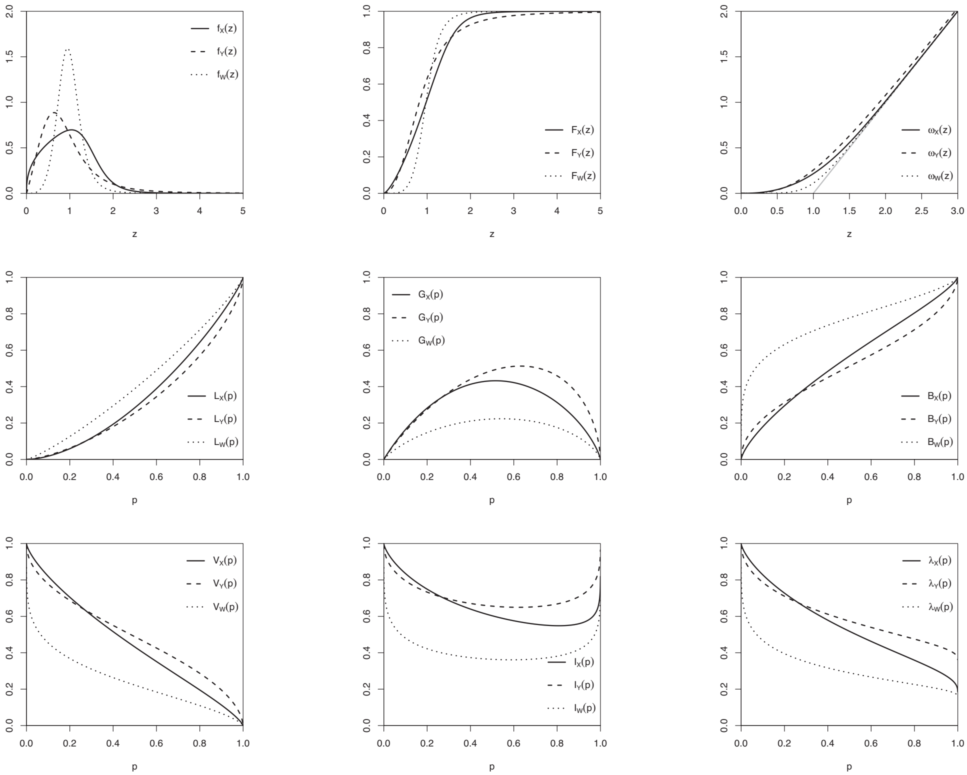

FIGURE 2. - *Graphical representation of Lemma 1*

All these observations are confirmed by an analysis of the plot of the probability density functions, where it is possible to realize that the probability density function of W takes on higher values around its mode, and it is less spread out than the probability density function of Y. On the contrary, the inequality curves of the random variables X and Y cross one each other, therefore these two variables cannot be compared using the partial orders.

These relationships between the probability density functions are well-founded only if the two random variables have the same expectation. Such final requirement can be always easily satisfied, whenever the distribution model has a scale parameter.

5. A "DIFFERENT" INEQUALITY CURVE: THE ZENGA-84 CURVE

Beyond the aforementioned curves, in the literature another curve has been proposed for measuring the inequality. It was initially introduced in Zenga (1984), hence in the following it will be labelled as Zenga-84 inequality curve $Z(p)$. In the literature many papers deal with this curve with methodological studies (see Aly and Hervas, 1999; Jedrzejczak, 2013; Dancelli, 1989, 1990; Kleiber and Kotz, 2003; Pollastri, 1987; Arnold, 2015a; Arcagni, 2017) and applications (see Salvaterra, 1990; Zenga, 1991; Jedrzejczak and Kubacki, 2013), just to mention some of them. The definition of such curve is the following one.

DEFINITION 8
Let X be a non-negative continuous random variable with probability density function f, strictly increasing distribution function F, and positive finite expectation μ. Let $Q(x)$ be the first incomplete moment of X, defined by:

$$Q(x) = \frac{1}{\mu}\int_0^x tf(t)dt.$$

If $F^{-1}(p)$ and $Q^{-1}(p)$ denote the inverse functions of F and Q, respectively, then the Zenga-84 inequality curve $Z_X(p)$ of X is defined as the relative difference of $F^{-1}(p)$ with respect to $Q^{-1}(p)$, that is:

$$Z_X(p) = \frac{Q^{-1}(p) - F^{-1}(p)}{Q^{-1}(p)} = 1 - \frac{F^{-1}(p)}{Q^{-1}(p)}, \qquad p \in [0, 1]. \tag{4}$$

As the other inequality curves, also this one originates a synthetic index, by the following definition.

DEFINITION 9
Let X be a non-negative continuous random variable, with positive and finite expected value μ, and let $Z_X(p)$ be the corresponding Zenga-84 inequality curve. Then the inequality index ζ_X of the random variable X is defined as

$$\zeta_X = \int_0^1 Z_X(p)dp = 1 - \int_0^1 \frac{F^{-1}(p)}{Q^{-1}(p)}dp.$$

As shown in Figure 3, it is interesting to note that the $Z(p)$ curve always assumes values in the interval $[0, 1]$, and it satisfies the ‘‘increasing ranking’’ as described in Arcagni and Porro (2014). Such feature can be emphasized, by considering the partial orderings. As for the other curves, a partial order based on $Z(p)$ can be defined in the following way.

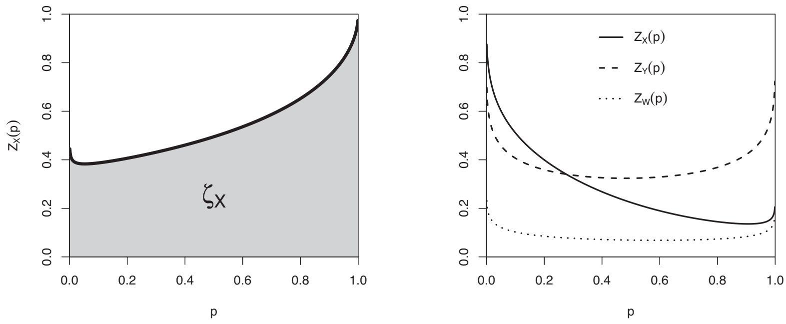

FIGURE 3. - *The Zenga-84 inequality curve $Z(p)$ and the inequality index ζ (left panel). The Zenga-84 curve for the same variables X, Y, W used in Figure 2 (right panel)*

DEFINITION 10
Let X and Y be two non-negative random variables, both with finite and positive expectation. X is said to be larger (or more unequal) than Y in the partial order based on the Zenga-84 $Z(p)$ curve, $(X \geq_Z Y)$, if

$$Z_X(p) \geq Z_Y(p), \qquad \forall p \in (0, 1)$$

This curve is ‘‘different’’ from the others, since there are no results stating that the partial order based on this curve is equivalent to the partial order based on one of the other inequality curves considered in the previous sections. This means that if one random variable is more inequal than another one in the ordering based on the Lorenz curve (or based on one of the other curves), it is not (almost yet) proved that the same relationship holds for the order based on the $Z(p)$ curve, and vice versa.

An important characteristic of the Zenga-84 curve is that it is constant for the classical lognormal distribution: if the random variable X has a distribution given by formula (2), with parameter $\alpha = 1$, then:

$$Z_X(p) = 1 - e^{-\frac{\theta^2}{2}}, \qquad \forall p \in [0, 1]. \tag{5}$$

6. DETAILS OF THE ANALYSIS

As mentioned in the previous sections, the aim of this paper is a detailed analysis regarding the role of the distribution parameters from the inequality point of view. To do this, one (or even more) tool for evaluating the inequality has to be chosen. In the previous sections several inequality curves have been defined.

The distribution models introduced in Section 2 have been selected for this analysis basically for two main reasons. The first one is that they are the most used models for representing the income in the literature; the second one is related to the number of their parameters. All these models have just three parameters.

For all the models, one parameter (or a function of it) plays the role of scale parameter, therefore it does not affect the inequality at all, meaning that all the inequality indicators do not depend on it. The remaining two parameters are instead related to inequality: in each model, if one is fixed, the other one is a direct or inverse inequality indicator. This means that a variation of the value of such parameter modifies the inequality level.

Beyond the aforementioned parameter classification, the present analysis will try to identify the parameters that mainly affect only a part of the distribution, for instance the lowest or the biggest values. Many papers in the literature about inequality and income distributions deal with this kind of analysis (see for example Dagum, 1977, 1980; Dancelli, 1989; Garcia Perez and Prieto Alaiz, 2011). The analysis will be illustrated by using graphs, and supported by analytical considerations. The graphs will show the inequality curves of the models using parameters ranges containing the most usual values observed in the applications from the real data. Extreme values are also considered for providing a more complete set of considerations.

In order to restrict the number of figures, the present paper describes the results by showing only two inequality curves. Among all the curves described in Section 3, the Zenga $I(p)$ curve is considered but, for the aforementioned relationships about partial orders, the same results hold also for the other curves of the section. The choice of the $I(p)$ curve is justified by the fact that the graphical analysis is more immediate since the differences among the curves are easier to be remarked. That happens also at both the extremes of the interval $[0, 1]$: as previously remarked, in those values of p, the value of the $I(p)$ curve depends on the distribution, therefore also the behaviour of such curve in their neighbourhood captures some features of the models, that are not caught by all the other inequality curves. The second drawn curve is the Zenga-84 one, for its characteristics already detailed in the previous section, which differ from the other curves.

The plots in this Section have been realized by using the Software R (R Core Team, 2015). Since a closed-form expression of the considered inequality curves does not exist for all the examined distribution models, in some cases a numerical approximation has been used.

6.1 *The Pareto Type II model*

The formula of the $I(p)$ curve for the Pareto Type II model is:

$$I(p) = \frac{\alpha\left[(1-p)^{-1/\alpha} - 1\right]}{p\left[(\alpha-1)(\theta-1) + \alpha(1-p)^{-1/\alpha}\right]}, \qquad p \in (0, 1).$$

Figure 4 shows the $I(p)$ and the $Z(p)$ curves for such distribution with different combinations of parameters. As expected it highlights that whether one parameter is fixed, the other one is an inverse inequality indicator (more details regarding the roles of the parameters are described in Arnold, 2008, 2015b). This characteristic can be observed in the rows and in the columns of the figure: if one parameter is fixed and the other one increases then the two inequality curves move to the bottom. More in detail, α modifies the shape of the inequality curves by affecting the inequality in the whole distribution; the parameter θ affects the left tail and the center of both the curves mainly for low values of α. Moreover, the $I(p)$ curve assumes different values in a right neighbourhood of 0, since the minimum value of the support of the variable depends on θ.

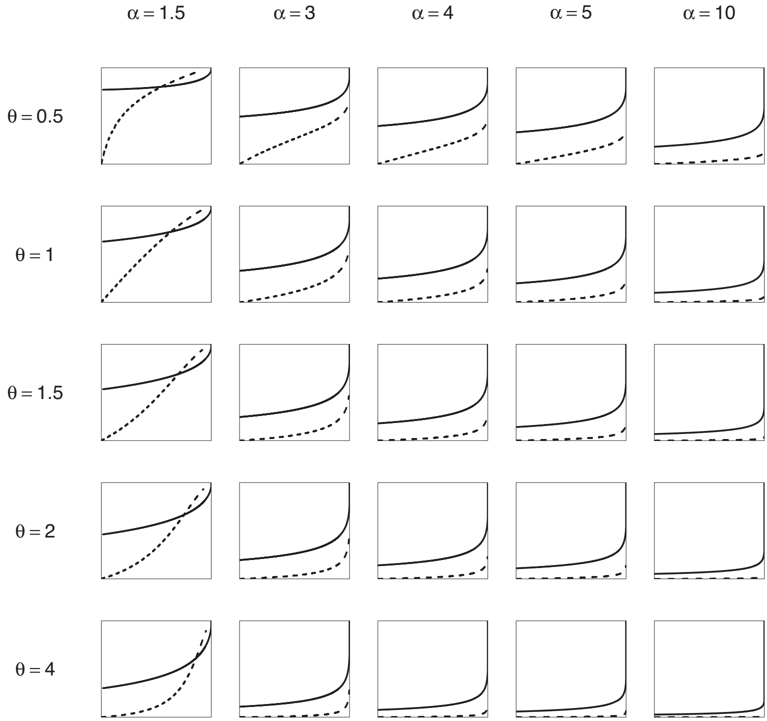

FIGURE 4. - *$I(p)$ (continue) and $Z(p)$ (dashed) curves of the Pareto Type II model*

6.2 *The generalized lognormal model*

In the literature, a closed-form expression of the $I(p)$ and $Z(p)$ curves for the generalized lognormal model has been found only for the special case where $\alpha = 1$, which corresponds to the usual lognormal distribution. For this reason, as previously argued, the graphs of both the curves (for $\alpha \neq 1$) have been obtained by numerical calculations. In Figure 5 some $I(p)$ and $Z(p)$ curves of the generalized lognormal model are shown. The parameter θ_1 is a transformation of the scale parameter, therefore it does not affect the inequality. When α is constant, θ_2 "lifts up" both the inequality curves along their entire length. On the contrary, α is an inverse inequality indicator and its most relevant influence is on the right tail. The support of the variable does not depend on any of the shape parameters, in fact the extreme values of the $I(p)$ curve are always equal to one, but the behaviour of such curve for low values of θ_2 suggests that in these cases the model has not heavy tails. For the generalized lognormal model with $\alpha = 1$ the $Z(p)$ curve is a straight line, with equation given by (5). It is interesting to observe that the Zenga-84 curve is convex for $\alpha < 1$ and concave for $\alpha > 1$.

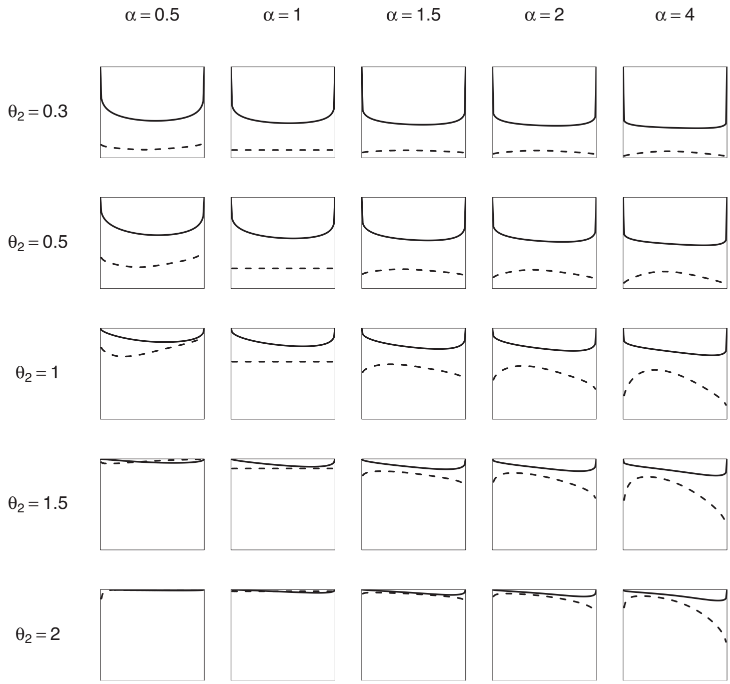

FIGURE 5. - *$I(p)$ (continue) and $Z(p)$ (dashed) curves of the generalized lognormal model*

6.3 *The Singh-Maddala model*

The Zenga inequality $I(p)$ curve for the Singh-Maddala model is given by:

$$I(p) = \frac{p - \beta\left(1 - (1-p)^{\frac{1}{q}}; 1 + \frac{1}{a}, q - \frac{1}{a}\right)}{p\left[1 - \beta\left(1 - (1-p)^{\frac{1}{q}}; 1 + \frac{1}{a}, q - \frac{1}{a}\right)\right]}, \qquad p \in (0,1),$$

where $\beta(x; \alpha, \theta)$ denotes the regularized incomplete Beta function, defined by:

$$\beta(x; \alpha, \theta) = \frac{1}{\beta(\alpha, \theta)} \int_0^x t^{\alpha-1}(1-t)^{\theta-1}, \qquad x \in [0,1], \alpha > 0, \theta > 0.$$

Figure 6 shows the behaviour of the two inequality curves for different values of the shape parameters of the Singh-Maddala model. The analysis by columns of the grid of plots shows that the parameter q is an inverse inequality indicator, and that it does not affect the behaviour of the $I(p)$ curve near the boundaries. The rows of the grid show that a is an inverse inequality indicator too, but it has more influence for extreme values of p. Moreover the remarkable asymmetric U-shape of the $I(p)$ curve reveals that the large values of the variable provide a lower contribution to the inequality level than the small ones. This feature can be observed more clearly by the decreasing trend of the $Z(p)$ curve. It is worth noting that for this curve, for low values of a, the parameter q affects also the right tail. More details regarding the role of the parameters from the inequality point of view can be found in Wilfling and Kramer (1993) and Kleiber (1999).

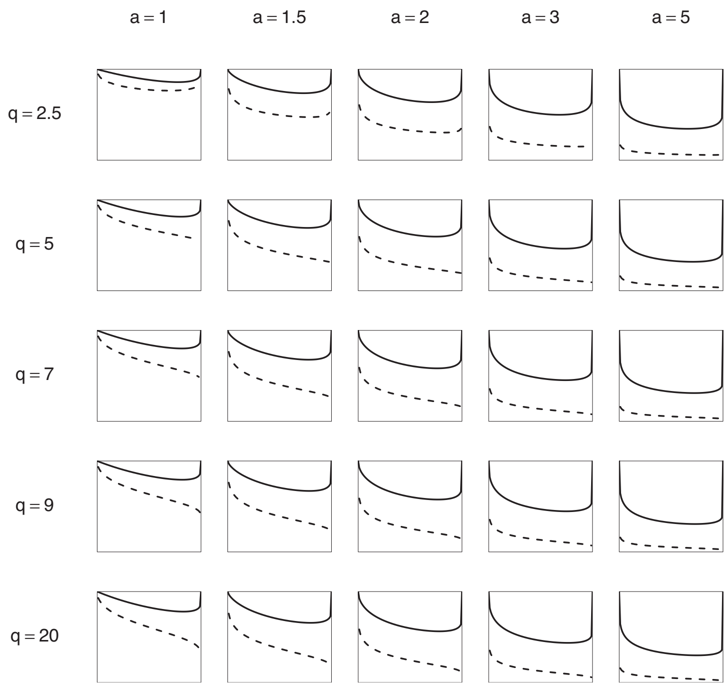

FIGURE 6. - *$I(p)$ (continue) and $Z(p)$ (dashed) curves of the Singh-Maddala model*

6.4 *The Dagum model*

The formula of the Zenga $I(p)$ curve for the Dagum model is:

$$I(p) = \frac{p - \beta\left(p^{\frac{1}{\theta}}; \theta + \frac{1}{\delta}; 1 - \frac{1}{\delta}\right)}{p\left[1 - \beta\left(p^{\frac{1}{\theta}}; \theta + \frac{1}{\delta}; 1 - \frac{1}{\delta}\right)\right]} \qquad p \in (0, 1).$$

A deep investigation about the roles of the parameters from the inequality point of view can be found in Dagum (1977) and Dancelli (1989).

The plots of the $I(p)$ and $Z(p)$ curves for Dagum model provided in Figure 7 confirm that both the shape parameters are inverse inequality indicators, but the influence of the parameter θ is more evident for low values of p, while the parameter δ has a more regular effect on the whole curves. These characteristics of the parameters are coherent with the ones obtained in Dancelli (1989).

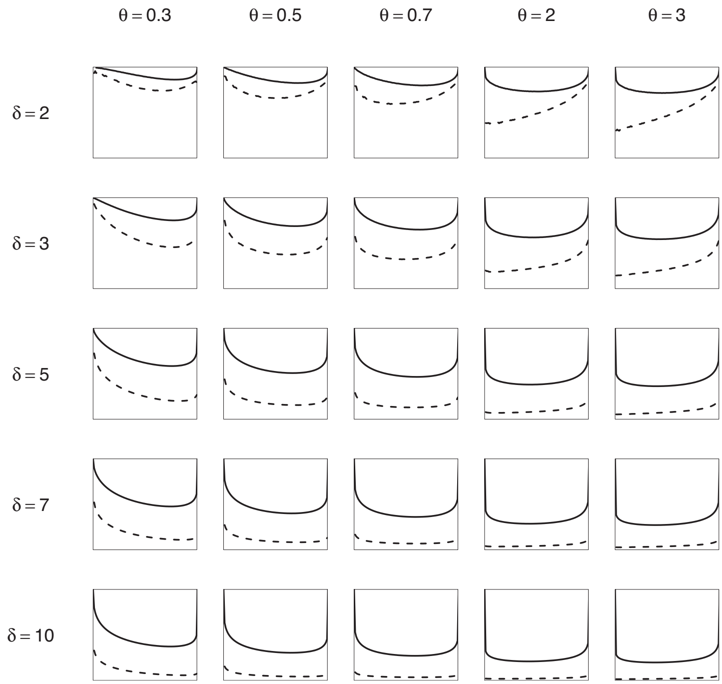

FIGURE 7. - *$I(p)$ (continue) and $Z(p)$ (dashed) curves of the Dagum model*

6.5 *The Zenga model*

As for the lognormal model, in the literature a closed-form expression for the $I(p)$ curve of the Zenga distribution is not available. For this reason, all the plots in this subsection have been obtained numerically.

The features of the parameters, from the inequality point of view, are investigated in Arcagni and Porro (2013) and Porro (2015). From Figure 8 it can be observed that α is an inverse inequality indicator and θ a direct one. Both the shape parameters influence the whole two inequality curves, even if, only for big values of the parameter θ, the effect of α on the $Z(p)$ curve seems to be more significant for values of p around 0.5. From the grid of the plots, it can be observed that the two curves seem to be quite symmetric, mainly as α increases.

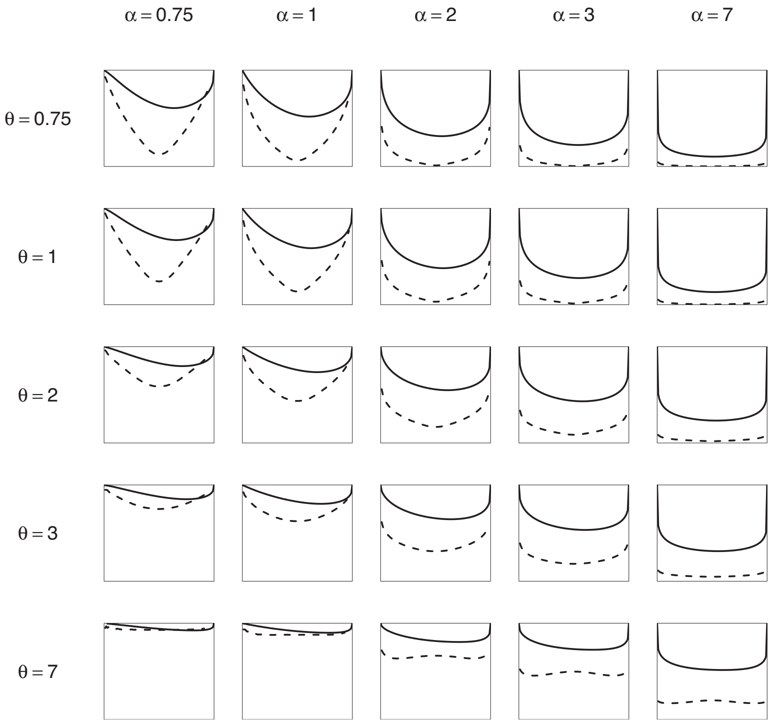

FIGURE 8. - $I(p)$ *(continue) and* $Z(p)$ *(dashed) curves of the Zenga model*

6.6 *Summary of the results*

Table 1 summarizes the results of the analysis. Each shape parameter of the considered distribution model has been classified as Direct or Inverse inequality indicator. Moreover, for the two considered inequality curves, the areas of the distribution most influenced by each parameter are highlighted, by using a ‘‘x’’ mark.

7. CONCLUSIONS

This paper focuses on the comparison of income distribution models through their inequality curves. Five of the most representative models have been selected because they can be considered suitable for such comparison. Such analysis is inspired by the one described in Dancelli (1989). The present analysis is performed by using two inequality curves: the $I(p)$ and the $Z(p)$ curves. The former has been selected for representing other inequality curves, which originate equivalent partial orderings; the latter because of its characteristics, quite different from the ones owned by all the other inequality curves. The distribution models are chosen with the same number of parameters, and they are parametrized in order to have two shape parameters

TABLE 1. - *Summary of the results*

Models and shape parameters	Inequality indicator	Most influence for $I(p)$			Most influence for $Z(p)$		
		Left tail	Center	Right tail	Left tail	Center	Right tail
Pareto Type II							
α	Inverse	x	x	x	x	x	x
θ	Inverse	x	x		x	x	
		(for small α)	(for small α)		(for small α)	(for small α)	
Gen. Lognorm.							
α	Inverse			x			x
θ_2	Direct	x	x	x	x	x	x
Singh-Maddala							
a	Inverse	x		x	x		x
q	Inverse		x			x	x
							(for small a)
Dagum							
θ	Inverse	x			x		
δ	Inverse	x	x	x	x	x	x
Zenga							
α	Inverse	x	x	x	x	x	x
						(for big θ)	
θ	Direct	x	x	x	x	x	x

and a scale parameter. This setting allows to perform comparisons and to provide some interesting conclusions regarding the shape of the probability density functions with the same expectation. Even if in some cases the two curves look different, it is also important to remark the considerable coherence of the results summarized in Table 1: the differences are very limited, and confirm the adequacy of both these tools to measure the inequality degree.

ACKNOWLEDGEMENTS

The authors would like to thank the two anonymous referees for their constructive comments and suggestions, which helped to improve the manuscript.

REFERENCES

Aly E.A., Hervas M.O. (1999). Nonparametric Inference for Zenga's Measure of Income Inequality. *Metron*, **57**, 69-84.

Amato V. (1987). Contributo della Scuola Statistica Italiana alla costruzione di classi di indici di concentrazione. In M. Zenga (Ed.), *La distribuzione personale del reddito: problemi di formazione, di ripartizione e di misurazione* (pp. 3-25). Vita e Pensiero, Milano.

Arnold B.C. (2008). Pareto and Generalized Pareto Distributions. In D. Chotikapanich (Ed.), *Modeling Income Distributions and Lorenz Curves*. Springer, New York City, NY.

Arnold B.C. (2015a). On Zenga and Bonferroni curves. *Metron*, **73**, 25-30.

Arnold B.C. (2015b). *Pareto Distributions - Second Edition*. CRC Press, Boca Raton.

Arcagni A., Porro F. (2013). On the parameters of Zenga distribution. *Statistical Methods & Applications*, **22**(3), 285-303.

Arcagni A., Porro F. (2014). The graphical representation of inequality. *Revista Colombiana de Estadistica*, **37**(2), 419-436.

Arcagni A. (2017). On the decomposition by sources of the Zenga 1984 point and synthetic inequality indexes. *Statistical Methods & Applications*, **26**, 113-133.

Basulto Santos J., Busto Guerrero J.J. (2010). Gini's concentration ratio (1908-1914). *Electronic Journal for History of Probability and Statistics*, **6**, 1-42.

Bhattacharya T., Ghosh T.S., Mande S.S, (2015). Global profiling of carbohydrate active enzymes in human gut microbiome, *PLOS ONE*, **10**(11). Available at: https://doi.org/10.1371/journal.pone.0142038

Bonferroni C.E. (1930). *Elementi di Statistica generale*. Sebeer, Firenze.

Brunazzo A., Pollastri A. (1986). Proposta di una nuova distribuzione: La lognormale generalizzata. In *Scritti in onore di Francesco Brambilla I* (pp. 57-83). Edizioni Bocconi Comunicazione, Milano.

Dagum C. (1977). A new model of personal income distribution: specification and estimation. *Economie Appliquée*, **30**(3), 413-437.

Dagum C. (1980). Inequality measures between income distributions with applications. *Econometrica*, **48**(7), 1791-1803.

Dancelli L. (1989). Confronti fra le curve di concentrazione $Z(p)$ e $L(p)$ nel modello di Dagum. *Statistica Applicata*, **1**(4), 399-415.

Dancelli L. (1990). On the behaviour of the $Z(p)$ concentration curve. In C. Dagum and M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty. Studies in contemporary Economics* (pp. 111-127). Springer Verlag, Berlin.

De Vergottini M. (1940). Sul significato di alcuni indici di concentrazione. *Giornale degli Economisti e Annali di Economia*, **2**(5/6), 317-347.

Gini C. (1914). *Sulla misura della concentrazione e della variabilita dei caratteri*. Premiate officine grafiche C. Ferrari, Venezia.

Garcia Perez C., Prieto Alaiz M. (2011). Using the Dagum model to explain changes in personal income distribution. *Applied Economics*, **43**, 4377-4386.

Gastwirth J.L. (1971). A general definition of the Lorenz curve. *Econometrica*, **39**, 1037-1039.

Jacobson A., Milman A.D., Kammen D. (2005). Letting the (energy) Gini out of the bottle: Lorenz curves of cumulative electricity consumption and Gini coefficients as metrics of energy distribution and equity. *Energy Policy*, **33**(14), 1825-1832.

Jedrzejczak A. (2013). Asymptotic properties of some estimators for income inequality measures - a simulation study. In *Proceedings 59th ISI World Statistics Congress* (pp. 5474-5479), CPS203.

Jedrzejczak A., Kubacki J. (2013). Estimation of income inequality and the poverty rate in Poland, by region and family type. *Statistics in Transition - New series*, **14**(3), 359-378.

Kleibert C. (1999). On the Lorenz order within parametric families of income distributions. *Sankhya*, **61**, 514-517.

Kleiber C., Kotz S. (2003). *Statistical size distributions in economics and actuarial sciences*. John Wiley & Sons, Hoboken, NJ.

Lorenz M.O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, **9**(70), 209-219.

McDonald J.B., Ransom M. (2008). *The generalized beta distribution as a model for the distribution of income: estimation of related measures of inequality*. Springer, New York City, NY.

Naeem S. (2009). Ecology: Gini in the bottle. *Nature*, **458**, 579-580.

Polisicchio M. (2008). The continuous random variable with uniform point inequality measure $I(p)$. *Statistica & Applicazioni*, **VI**(2), 137-151.

Polisicchio M., Porro F. (2009). A comparison between Lorenz $L(p)$ curve and Zenga $I(p)$ curve. *Statistica Applicata*, **21**(3-4), 289-302.

Pollastri A. (1987). Le curve di concentrazione $L(p)$ e $Z(p)$ nella distribuzione lognormale generalizzata. *Giornale degli Economisti e Annali di Economia*, **11-12**, 639-663.

Porro F. (2015). Zenga distribution and inequality ordering. *Communications in Statistics - Theory and Methods*, **44**(18), 3967-3977.

R Core Team (2015). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. Vienna, Austria. *http://www.R-project.org/

Salvaterra T. (1990). Comparison among concentration curves and indexes in some empirical distributions. In C. Dagum and M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty. Studies in contemporary Economics* (pp. 194-214). Springer Verlag, Berlin.

Shaked M., Shanthikumar J.G. (2007). *Stochastic Orders*. Springer, New York City, NY.

Singh S.K., Maddala G.S. (1976). A Function for size distribution of incomes. *Econometrica*, **44**, 963-970.

Vianelli S. (1982). Sulle curve lognormali di ordine *r* quali famiglie di distribuzioni di errori di proporzione. *Statistica*, **II**, 156-176.

Wilfling B., Kramer W. (1993). The Lorenz-ordering of Singh Maddala income distributions. *Economic Letters*, **43**, 53-57.

Zenga M. (1984). Proposta per un indice di concentrazione basato sui rapporti tra quantili di popolazione e quantili di reddito. *Giornale degli Economisti e Annali di Economia*, **43**(5-6), 301-326.

Zenga M. (1987). *Il contributo degli italiani allo studio della concentrazione - Prima parte: dal 1895 al 1915*. Vita e pensiero, Milano.

Zenga M. (1990). Concentration curves and concentration indexes derived from them. In C. Dagum and M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty, Studies in contemporary Economics* (pp. 94-110). Springer Verlag, Berlin.

Zenga M. (1991). Impiego delle tabelle di cograduazione per la determinazione dell'indice puntuale di concentrazione $Z(p)$. *Statistica Applicata*, **3**, 283-291.

Zenga M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica & Applicazioni*, **V**(1), 3-27.

Zenga M. (2010). Mixture of Polisicchio's truncated Pareto distributions with beta weights. *Statistica & Applicazioni*, **VIII**(1), 3-25.

Zenga M., Pasquazzi L., Zenga Ma. (2010). First applications of a new three parameter distribution for non-negative variables. *Statistica & Applicazioni*, **VIII**(2), 131-147.

Zenga M., Polisicchio M., Zenga Ma., Pasquazzi L. (2011). More on M.M. Zenga's new three-parameter distribution for non-negative variables. *Statistica & Applicazioni*, **IX**(1), 5-33.