Hacettepe Journal of Mathematics and Statistics
Volume 46 (6) (2017), 1159 – 1174

Some properties of the Canberra inequality index

Shahryar Mirzaei*, Gholam Reza Mohtashami Borzadaran†‡, Mohammad Amini§ and Hadi Jabbari ¶

Abstract

In this paper, we study some properties of the Canberra index which is based on Canberra distance function. Some features that required to an inequality index are investigated. Also we present explicit expression for the Canberra curve in some important inequality distributions. Further, we compare the Canberra curve with the traditional Lorenz curve. A simulation study based on fitted distribution to real income data is performed in order to investigate the asymptotic behavior of the proposed sampling estimator. Finally, the superiority of this index is illustrated by means of a real data set.

Keywords: Canberra index, Distance function, Income distribution, Lorenz curve.

2000 AMS Classification: 62P20, 91B82, 30F45.

Received : 26.11.2015 *Accepted* : 24.01.2016 *Doi* : 10.15672/HJMS.201611115602

1. Introduction

One of the important issues in income inequality indices is finding an appropriate mathematical metric to measure disparity in population. There have been considerable efforts in finding the appropriate measures among such a plethora of choices because it is of fundamental importance to have particular properties for detection inequality in the best way. It should be noted that the importance of finding suitable distance measures cannot be overemphasized yet. A number of distance measures to exploit new inequality indices have been proposed and extensively studied by Gini [6], Pietra [14], Bonferroni [3], Mehran [13], Kakwani [9], Chakravarty [5] and Zenga [18, 19].

There are metrics which have not received much attention even though they satisfy the usual properties required to an income inequality index. Such a metric is the Canberra

*Ferdowsi University of Mashhad, Mashhad, Iran, Email: sh.mirzaei@stu.um.ac.ir
†Ferdowsi University of Mashhad, Mashhad, Iran, Email: grmohtashami@um.ac.ir
‡Corresponding Author.
§Ferdowsi University of Mashhad, Mashhad, Iran, Email: m-amini@um.ac.ir
¶Ferdowsi University of Mashhad, Mashhad, Iran, Email: jabbarinh@um.ac.ir

distance. The Canberra metric, introduced by Lance and Williams [11] as a software metric, is a weighted version of the classic L_1 distance family (see Cha [4]) which naturally extends to a metric on symmetric groups. The role of the Canberra distance as a stability indicator for measuring income inequality was first described in Subramanian [17] as a measure of disarray between the Lorenz curve and the equality line of ranked non-negative incomes.

In this paper, we present some properties of the Canberra inequality index and related curve which is based on the Canberra distance function. In this work, the discussed measure is investigated for continuous models, but it can be also applied to discrete and empirical distributions.

The article is organized as follows. Section 2 contains some preliminaries and the basic tools which will be used in the next sections. In Section 3, we present a general definition of Canberra index which is based on the Canberra distance function. Some properties that are required to an inequality measure including determining the distribution, scale invariance and translation effect are investigated. An important application of the Canberra curve is that it can be used to define curve ordering. This application is discussed in Section 4. We present explicit expressions for the Canberra index and corresponding curve with graphical representation in some important inequality distributions in Section 5. In Section 6, we propose the Canberra sampling estimator. Section 7 performs a simulation study based on Singh Madalla, fitted distribution to income of the US family for 2013 in order to investigate the asymptotic normal behavior of the discussed estimator. Next, to show the sensivity of the Canberra curve for low income, a real data set is presented in Section 8. Finally, conclusions are given in the last section.

2. Basis tools

The Lorenz curve is a graphical representation of the cumulative income distribution. It was developed by Max O. Lorenz [12] for representing inequality in the wealth distributions. The Lorenz curve can usually be represented by a function $L(p)$, where p, the cumulative portion of the population, is represented on the horizontal axis, and $L(p)$, the cumulative portion of the total wealth or income, is represented on the vertical axis. Let $\mathcal{L}$ be the class of all non-negative continuous random variables with positive finite expectations ($\mu = E(X) > 0$). For a random variable X in $\mathcal{L}$ with cumulative distribution function (cdf) F, we define its inverse distribution function by

$$F^{-1}(t) = \inf\{x : F(x) \geq t\}, \qquad t \in [0, 1].$$

According to Zenga [19], the Lorenz function corresponding to X is defined by

$$L(p) = \frac{1}{E(X)} \int_0^p F^{-1}(t)dt, \qquad p \in [0, 1]. \tag{2.1}$$

The Lorenz function indicates the cumulative percentage of total income held by a cumulative proportion p of the population. To visualize proportions (2.1), like Figure 1, we plot the points $(p, L(p))$.

As the result of this, we obtain the curve L called the Lorenz curve. The curve L is well defined on the entire interval [0, 1], with values $L(0) = 0$ and $L(1) = 1$. It can be noted that the Lorenz curve is always below the diagonal

$$I(p) = p, \qquad p \in [0, 1].$$

The diagonal I, on the other hand, is also a Lorenz curve. Indeed, assuming that all the incomes are equal. Thus the interpretation of I (the straight line) represents perfect equality and any departure from this 45° line represents inequality.

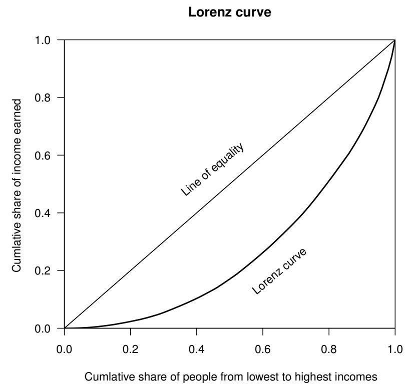

Figure 1. An example of Lorenz curve

Based on the discussion about Lorenz curve and line of equality, it now becomes natural to measure the economic inequality by using some distance $d(I, L)$ between the egalitarian Lorenz curve (I) and the actual one L, then we can consider $d(I, L)$ as a measure of economic inequality in the population. The main idea behind the construction of $d(\cdot, \cdot)$ is based on the fact that we are merely interested in measuring the distance between I and L. This implies that we are really interested only in the functional $\mathcal{D} = d(I, L)$ defined on the set of all Lorenz curves. It is natural to require the functional $\mathcal{D}$ be such that

(1) $\mathcal{D}(L) \geq 0$,
(2) $\mathcal{D}(L) = 0$ if $L = I$,
(3) $\mathcal{D}(L') \geq \mathcal{D}(L'')$ whenever ${L_n}' \leq {L_n}''$.

According to what has been discussed above, inequality measures have been constructed as distance functions. Some important indices in terms of distance measures have been shown in Table 1.

Table 1. Income inequality indices based on distance functions

Inequality index	$d(p, L(p))$	
Pietra	$\max_{0 \leq p \leq 1} (p - L(p))$	
Gini	$2 \int_0^1 (p - L(p))dp$	
E-Gini	$2[\int_0^1 (p - L(p))^{\alpha} dp]^{\frac{1}{\alpha}}$	$\alpha \geq 1$
S-Gini	$v(v-1) \int_0^1 (p - L(p))(1-p)^{v-2} dp$	$v \geq 1$
Bonferroni	$\int_0^1 (1 - \frac{L(p)}{p})dp$	
Zenga	$\int_0^1 \frac{p - L(p)}{p(1 - L(p))} dp$	

3. The Canberra index

In this section, we study the Canberra measure and related curve which are derived based on Canberra distance function. Also some basic properties of this measure have been investigated.

3.1. Definition of the Canberra index.

3.1. Definition. Let X be a random variable belonging to $\mathcal{L}$ class. The Canberra index (C) based on Canberra distance is defined as

$$C = d(p, L(p)) = \int_0^1 C(p)dp, \tag{3.1}$$

under the Canberra curve, defined by

$$C(p) = \frac{p - L(p)}{p + L(p)}, \qquad p \in (0, 1]. \tag{3.2}$$

Since the Lorenz function, in terms of expectation of X, can be obtained as

$$L(p) = \frac{pE(X|X \leq F^{-1}(p))}{E(X)}, \tag{3.3}$$

then the Canberra curve can be rewriten as

$$\begin{aligned} C(p) &= \frac{E(X) - E(X|X \leq F^{-1}(p))}{E(X) + E(X|X \leq F^{-1}(p))}, \\ &= \frac{\mu - \mu(p)}{\mu + \mu(p)}, \end{aligned}$$

where $\mu(p)$ is the partial mean of proportion p of the ordered income (poorest) population.

If all the recipients in the population have the same quantity, the Canberra curve coincides with the line of perfect equality that joins the coordinate points $(0, 1), (0, 0), (1, 0)$. Also, the line of maximum inequality joins the coordinate points $(0, 1), (1, 1), (1, 0)$. If the random variable X tends to the situation of minimal inequality, the $C(p)$ curve tends to the function

$$C_m(p) = 0, \qquad p \in (0, 1),$$

while, if the random variable X tends to the situation of maximal inequality, the $C(p)$ curve tends to 1, for all $p \in (0, 1)$, that is

$$C_M(p) = 1, \qquad p \in (0, 1).$$

It is easy to see that $\lim_{p \to 0^+} C(p)$ depends on the random variable originating the Canberra curve. This means that near the boundary of the left domain, the $C(p)$ curve is related to the distribution of X, and therefore it is more explanatory than the Lorenz curve. So, the $C(p)$ curve is more suitable for inequality analysis around lower values of X.

Similar to other income inequality indices, the Canberra index can be obtained from the mean value of the $C(p)$ curve but it represents the area below the $C(p)$ curve. In Figure 2 from a geometric point of view, the Canberra index is the area between the Canberra curve and the line of perfect equality. This index is 0 in maximum equality and 1 in perfect inequality. The new inequality index satisfies all the main properties required to a synthetic inequality measure (see Subrimanian [17]). It is symmetric in the sense of its invariance under any permutation of incomes. The symmetry property follows from the fact that we have defined C directly on an ordered distribution. It satisfies the Pigue-Dalton condition, a postulate which states that an income transfer from a richer

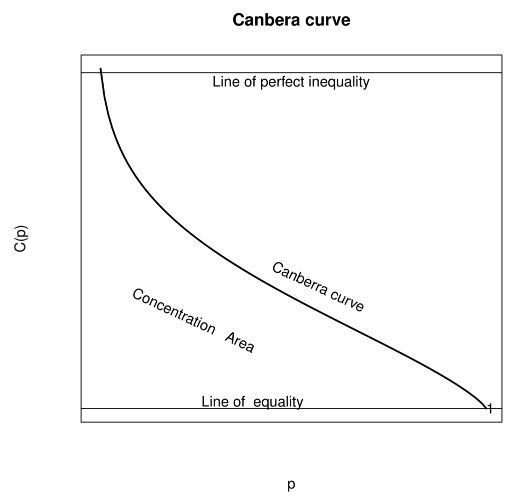

Figure 2. An example of Canberra curve

to a poorer individual, other things remaining the same including their relative rank in distribution, decreases the extent of income inequality.

3.2. Determination of the distribution. One important property of an inequality curve is determining the corresponding income distribution. The following proposition states this property for the Canberra inequality curve.

3.2. Proposition. *Let X be a random variable belonging to $\mathcal{L}$ class with $C(p)$ Canberra curve. The distribution of X is uniquely determined by $C(p)$.*

Proof. Based on (2.1) and (3.2), the relation between the $C(p)$ and $F^{-1}(t)$ can be written as

$$\int_0^p F^{-1}(t)dt = \mu \frac{p(1-C(p))}{C(p)-1}, \tag{3.4}$$

the result is achieved by differentiating (3.4) as

$$F^{-1}(p) = \mu \frac{d}{dp}[\frac{p(1-C(p))}{C(p)-1}],$$

thus F^{-1} will determine F.

□

3.3. Scale invariance. In this subsection, an argument can be advanced in favor of the requirement that inequality curves and measures should be scale invariant, i.e., that the inequality associated with a random variable X should be the same as that associated with the random variable $Y = aX$ for any $a > 0$.

3.3. Proposition. *The Canberra curve and corresponding index are scale invariant.*

Proof. Let $Y = aX$ for any $a > 0$, then

$$F_Y^{-1}(p) = aF_X^{-1}(p), \qquad p \in (0,1), \tag{3.5}$$

where F_X^{-1} and F_Y^{-1} are the inverse function of X, Y respectively. Also, we have

$$E(Y) = aE(X). \tag{3.6}$$

Since $L_X(p) = \frac{1}{E(X)} \int_0^p F_X^{-1}(t)dt$, is scale invariant, from (3.5) and (3), then the Canberra curve as a function of Lorenz curve, also is scale invariant. So, we have

$$C_Y(p) = \frac{p - L_Y(p)}{p + L_Y(p)} = \frac{p - L_X(p)}{p + L_X(p)} = C_X(p). \tag{3.7}$$

Consequently, the scale invariant of the Canberra index follows from (3.7). □

3.4. Translation effect. Another important property of an inequality measure is its translation effect. Here, we analyze how some translations influence the Canberra measure. In the following proposition, it will be shown that the Canberra index is consistent with translation.

3.4. Proposition. *The Canberra measure is consistent with translation.*

Proof. Let $Y = X + b$ for any $b > 0$, $C_X(p)$ and $C_Y(p)$ be point measures of the Canberra index of X and Y, respectively. So, we have

$$\begin{aligned} C_Y(p) &= \frac{\mu_Y - \mu_Y(p)}{\mu_Y + \mu_Y(p)}, \\ &= \frac{(\mu_X + b) - (\mu_X(p) + b)}{(\mu_X + b) + (\mu_X(p) + b)}, \\ &= C_X(p) \times \frac{\mu_X + \mu_X(p)}{\mu_X + \mu_X(p) + 2b}. \end{aligned}$$

For every fixed value $b > 0$, the ratio $\frac{\mu_X + \mu_X(p)}{\mu_X + \mu_X(p) + 2b}$ assumes values in $(0, 1)$. Hence

$$C_Y(p) < C_X(p), \qquad p \in (0, 1). \tag{3.8}$$

Consequently, the consistency of the Canberra index follows from (3.8). □

3.5. Remark. An analogous consideration holds for any $b < 0$, in this case $C_Y > C_X$.

4. Stochastic orders based on $C(p)$ curve

An important application of the inequality curves is that they can be used to define some orderings. Such orderings allow the comparison of distributions in terms of inequality. This kind of comparison within the same model allows to understand how the distribution parameters influence the inequality. In this section, a result about the partial order based on Lorenz curve and the partial order based on Canberra curve is presented. It will be shown that the two curves establish two equivalent partial orders. Next we include the definition of the well-known ordering based on the Lorenz curve.

4.1. Definition. Let X and Y be random variables belonging to $\mathcal{L}$ class. The Lorenz order $\leq_L$ on $\mathcal{L}$ is defined by,

$$X \leq_L Y \Leftrightarrow L_X(p) \geq L_Y(p), \qquad p \in [0, 1].$$

If $X \leq_L Y$, then X exhibits less inequality than Y in the Lorenz sense. From the graphical point of view, the random variable X is smaller than Y in this order, if its Lorenz curve lies above the Lorenz curve of Y for all $p \in (0, 1)$. Analogously, to the ordering based on the Lorenz curve, other orders have been applied in inequality analysis. The links among some different orders and their relationships with inequality have been deeply studied (see Sarbia [16], Kleiber and Kotz [10], Polisicchio and Porro [15] and Arnold [2]).

In analogy to Lorenz ordering, the ordering curve are defined for the Canberra inequality curve as follows.

4.2. Definition. Let X and Y be random variables belonging to $\mathcal{L}$ class. The Canberra order $\leq_C$ on $\mathcal{L}$ is defined by

$$X \leq_C Y \Leftrightarrow C_X(p) \leq C_Y(p), \qquad p \in (0,1].$$

If $X \leq_C Y$, then X exhibits less inequality than Y in the Canberra sense.

4.3. Lemma. *Let $X, Y \in \mathcal{L}$. Then*

$$X \leq_L Y \Leftrightarrow X \leq_C Y. \tag{4.1}$$

Proof. By definition, $X \leq_L Y$, means $L_X(p) \geq L_Y(p)$ for any fixed $p \in (0,1)$, also, since the Canberra curve can be written in the form

$$C(p) = 1 - \frac{2}{[\frac{p}{L(p)}] + 1},$$

and conversely

$$L(p) = \frac{p}{[\frac{2}{1-C(p)}] - 1},$$

it is evident that $X \leq_L Y \Leftrightarrow X \leq_C Y$. □

The stochastic order based on Bonferroni and Zenga curves appear to be essentially equivalent to the stochastic order based on Lorenz curve (Arcagni and Porro [1] and Arnold [2]). Since the Canberra and Lorenz order are equivalent in (4.1), hence, the relation based on curve orderings follows from Proposition 4.4.

4.4. Proposition. *Let $X, Y \in \mathcal{L}$. Then the following statements are equivalent:*

$$X \leq_C Y \Leftrightarrow X \leq_L Y \Leftrightarrow X \leq_B Y \Leftrightarrow X \leq_Z Y,$$

where $\leq_B$ and $\leq_Z$ denote the Bonferroni order and the Zenga order, respectively.

For the definitions and features of the Bonferroni order and the Zenga order stochastic dominance, we refer the reader to Arcagni and Porro [1] and Arnold [2].

5. The Behavior of the Canberra curve in some income models

The comparison between the Lorenz and the Canberra curves arise in the comparison of the curves for some income models. In this section, two curves are presented. The Lorenz curve is the oldest but also the most used nowadays despite its forced behavior. The Canberra curve is the most recent although related to the Lorenz curve and it hasn't a forced behavior. It can assume different shapes which allow to distinguish different situations in terms of inequality.

Several densities have been proposed in the literature to model the income distribution. Of course all these densities are defined for a positive support. The most simple distributions, and consequently the widely used ones are exponential, uniform and Pareto. These distribution models just examined have one or two parameters, in spite of that, their curves have different behaviors. In order to fit better tails, three-parameter distributions are proposed. We shall examine the Singh-Maddala distribution as a three-parameter distribution.

5.1. Exponential model. Let X be a random variable with exponential distribution $F_X(x) = 1 - e^{-\lambda x} I_{(x>0)}$, where $\lambda > 0$ is its parameter distribution, then the $L(p)$ curve is

$$L(p) = p + (1-p)\log(1-p), \qquad p \in [0,1],$$

and from that, the Canberra curve can be obtained as

$$C(p) = -\frac{(1-p)\log(1-p)}{2p + (1-p)\log(1-p)}, \qquad p \in (0,1].$$

It is important to note that the scale parameter λ is not an inequality indicator, in fact the Canberra curve, Lorenz curve and inequality measures derived from them don't depend on λ (see Figure 3).

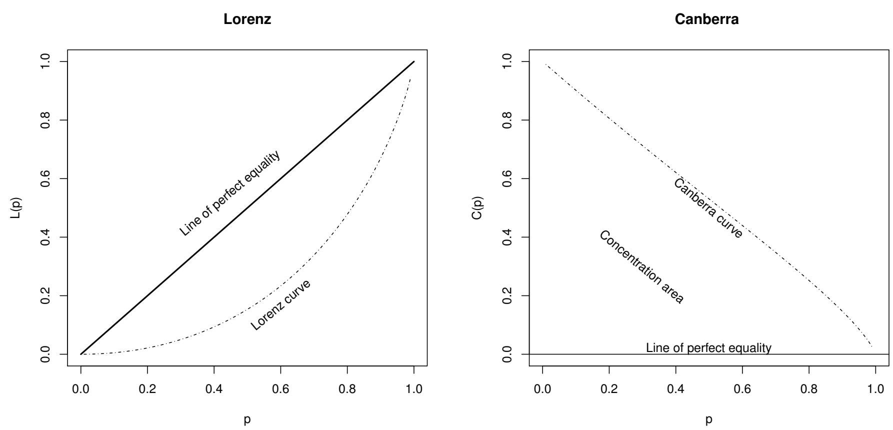

Figure 3. Lorenz and Canberra curves of exponential distribution

5.2. Pareto model. A random variable X follows a Pareto distribution if its distribution function is

$$F(x) = 1 - (\frac{x}{x_0})^{-\theta} I_{(x>x_0)},$$

where $x_0 > 0$ and $\theta > 1$. In this case,

$$L(p) = 1 - (1-p)^{(1-\frac{1}{\theta})}, \qquad p \in [0,1],$$

and from that, the Canberra curve can be obtained as

$$C(p) = \frac{p - [1-(1-p)^{(1-\frac{1}{\theta})}]}{p + [1-(1-p)^{(1-\frac{1}{\theta})}]}, \quad p \in (0,1]. \tag{5.1}$$

Both curves do not depend on scale parameter x_0. It can be seen that, for any fixed $p \in (0,1)$ as θ increases, the Lorenz curve increases too (whereas the inequality decreases). For the $C(p)$ curve, since for any fixed $p \in (0,1)$ the partial derivative of $C(p)$ with respect to θ is negative, it follows that if θ increases, then the $C(p)$ curve decreases, and so inequality does. Therefore, for both curves, the distribution parameter θ is an inverse inequality indicator. Suppose $\Theta = \{\theta \mid 1 < \theta < \infty\}$ be a parameter space of Pareto model. In this case, we have

$$\lim_{\theta \to 1} L(p,\theta) = 0 \qquad \text{and} \qquad \lim_{\theta \to \infty} L(p,\theta) = p.$$

Some $L(p)$ and $C(p)$ curves for the Pareto model are drawn in Figures 4. Each curve corresponds to a different choice of the distribution parameter θ. It is worth noting that

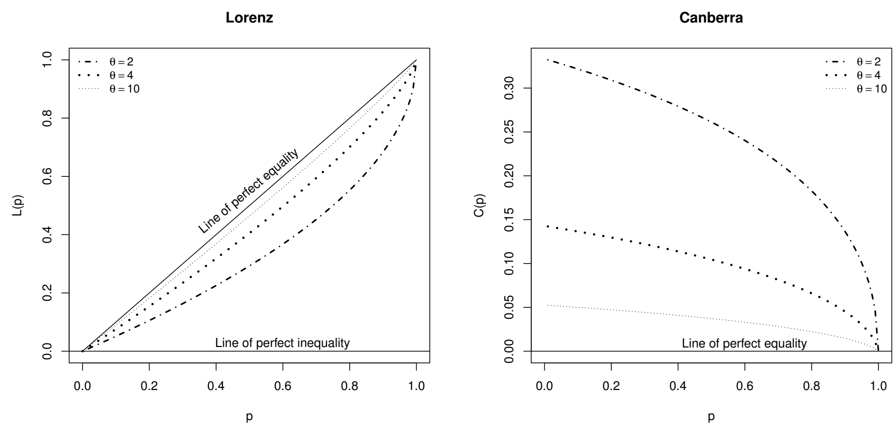

Figure 4. Lorenz and Canberra curves of Pareto distribution

in the Pareto model for Canberra curve in (5.1), we have

- $\lim_{p\to 0^+} C(p) = \frac{1}{2\theta-1}$, $\lim_{p\to 1^-} C(p) = 0$, and therefore, as $p \to 0$, the $C(p)$ curve tends to a value which depends only on the parameter θ. Also as $p \to 1$, the $C(p)$ curve has a force behavior and it tends to 1.
- $C'(p) > 0, \quad \theta > 1$, consequently $C(p)$ curve is increasing for $p \in (0, 1]$.
- $C''(p) < 0, \quad \theta > 1$, consequently $C(p)$ curve is concave for $p \in (0, 1]$.

5.3. The uniform model. Consider the random variable X with uniform distribution $F_X(x) = \frac{x-a}{b-a} I_{(a<x<b)}$, where $0 < a < b$ are parameters distribution, then the $L(p)$ curve is

$$L(p) = \frac{2ap + (b-a)p^2}{a+b}, \qquad p \in [0, 1],$$

and therefore the $C(p)$ curve is

$$C(p) = \frac{(1-p)(b-a)}{p(b-a) + (3a+b)}, \qquad p \in (0, 1]. \tag{5.2}$$

Here, the role of the two parameters a and b for the uniform model in relation to Lorenz and Canberra curves has been analyzed. In Figure 5 some $L(p)$ and $C(p)$ curves are shown with different values of a and $b = 10$. In Figure 6, the distribution parameter a is fixed in 2, and the value of b changes. It is evident that, if the value of the distribution parameter b is fixed then a is an inverse inequality indicator, and if a is fixed, then b is a direct indicator.

It is notable that in the uniform model for Canberra curve in (5.2), we have

- $\lim_{p\to 0^+} C(p) = \frac{b-a}{3a+b}$ and $\lim_{p\to 1^-} C(p) = 0$, and therefore, as $p \to 0^+$, $C(p)$ curve tends to a value which depends on the two parameters.
- $C'(p) = \frac{2(b^2-a^2)}{(ap-bp-3a-b)^2} > 0$, consequently $C(p)$ curve is increasing for $p \in (0, 1)$.
- $C''(p) = \frac{4(b-a)^2(a+b)}{(ap-bp-3a-b)^2} > 0$, consequently $C(p)$ curve is convex for $p \in (0, 1)$.

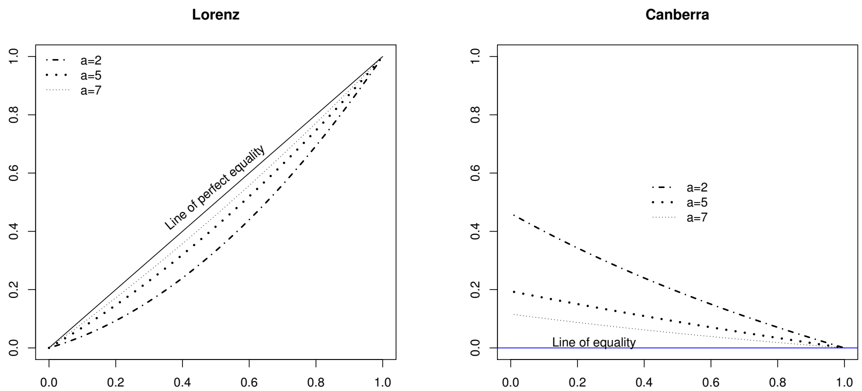

Figure 5. Lorenz and Canberra curves for uniform model with $b = 10$ and different values of a

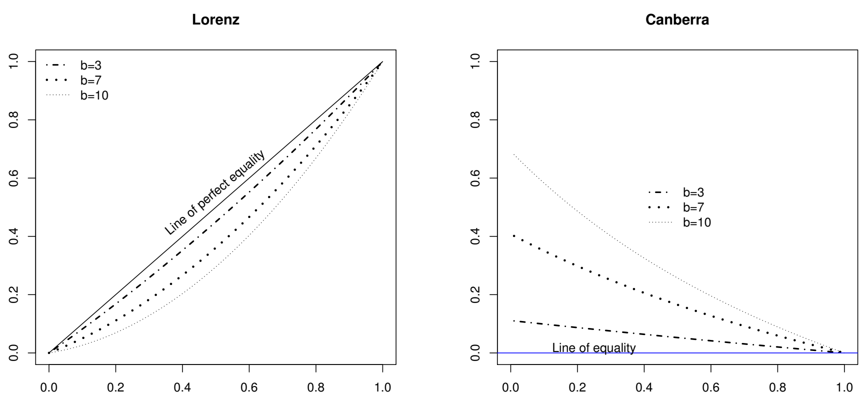

Figure 6. Lorenz and Canberra curves for uniform model with $a = 2$ and different values of b

5.4. The Singh Madalla model. Let X be a non-negative random income variable with Singh Madalla distribution and corresponding cdf,

$$F(x) = 1 - \frac{1}{[1 + (\frac{x}{b})^a]^q},$$

where $a, b, q > 0$. If $q > \frac{1}{a}$, then using Expression (2.1), the Lorenz curve of X is

$$\begin{aligned} L(p) &= \frac{1}{\mu}\int_0^p b[(1-y)^{-\frac{1}{q}}-1]^{\frac{1}{a}}dy \\ &= \frac{bq}{\mu}\int_0^z t^{\frac{1}{a}}(1-t)^{q-\frac{1}{a}-1}dt \\ &= I_z(1+\frac{1}{a}, q-\frac{1}{a}), \qquad p \in [0,1], \end{aligned}$$

where $z = 1-(1-p)^{\frac{1}{q}}$ and $I_x(a,b)$ denotes the incomplete beta function ratio defined as

$$I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}.$$

Therefore, $C(p)$ curve is

$$C(p) = \frac{p - I_z(1+\frac{1}{a}, q-\frac{1}{a})}{p + I_z(1+\frac{1}{a}, q-\frac{1}{a})}, \qquad p \in (0,1].$$

Here, the role of the parameter q for the Singh Madalla model in relation to Lorenz and Canberra curves has been analyzed. We note that since b is the scale parameter in the Singh Madalla model, the inequality indices and curves are invariant to it, and thus our results will not be affected by the choice. In Figure 7, some $L(p)$ and $C(p)$ curves are shown for different values of q using $a = 2$. It is evident that, if the value of the distribution parameter a is fixed then q is an inverse inequality indicator. In similar way, if q is fixed, then a is an indirect indicator too.

It is noteworthy that when the inequality curve is concave/convex, the lowest and the highest incomes are the most frequent and there is a tendency for polarization, in which case $C(p)$ is convex/concave. That is, unlike what occurs with Lorenz curve, the shape of the Canberra curve yields more information on the associated distribution.

6. Sampling estimator of Canberra index

Let $X_1, \ldots, X_n$ be a random sample of size n from an cdf F. Suppose $X_{1:n} \leq \ldots \leq X_{i:n} \leq \ldots \leq X_{n:n}$ be the corresponding order statistics. We can define the Canberra estimator by plug in empirical cdf of F ($\hat{F}_n$) instead of F in (3.1) and (3.2) as

$$\hat{C}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{\bar{X} - \bar{X}_{i:n}}{\bar{X} + \bar{X}_{i:n}}, \qquad i = 1, 2, \ldots, n,$$

where $\bar{X}_{i:n}$ is the partial ordered mean and $\bar{X}$ denoting the sample mean of $X_1, \ldots, X_n$.

A nice property of this distance estimator is that for large sample size, the Canberra estimator is asymptotically normal. In many income distribution studies, however, involving fairly large samples, the distribution of Canberra estimator is normal distribution without regard of what distribution each of the income variables follow. It can be noted that $C(p)$ curve is a function of F based on (3.2) and (3.3). Thus, for more emphesize, we show $C(F)$ instead of $C(p)$ in the following theorem.

6.1. Theorem. *Let X be a random variable belonging to $\mathcal{L}$ class with $E(|X|^{2+\alpha}) < \infty$ for $\alpha \geq 1$, the estimator of Canberra index is asymptotically normal.*

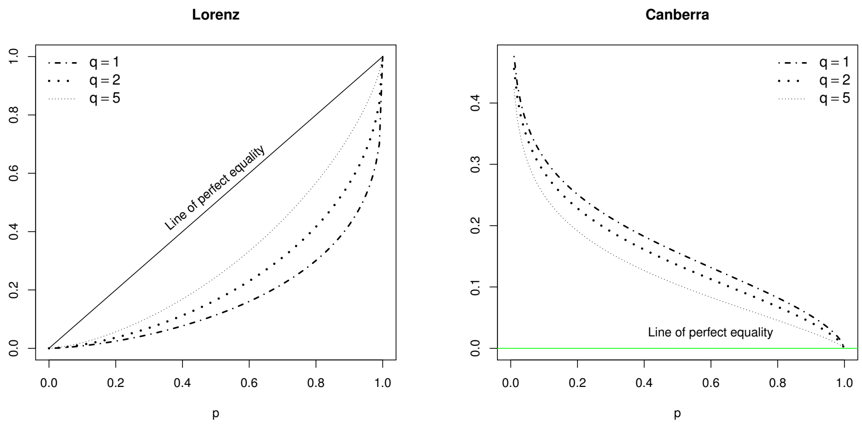

Figure 7. Lorenz and Canberra curves for the Singh Madalla model with $a = 2$ and different values of q

Proof. Under mild conditions on cdf of X such as continuity and differentiability, the Canberra estimator may be represented as (Hoeffding [7])

$$\hat{C}_n = C + \frac{1}{n}\sum_{i=1}^{n} h_C(X_i) + o(n^{\frac{-1}{2}}),$$

where $h_C(X_i)$ denotes the influence function evaluated at the point X_i, i.e.

$$h_C(X_i) = \lim_{\lambda\to 0} \frac{C(F + \lambda(\delta_{X_i} - F)) - C(F)}{\lambda},$$

and δ_X denotes the distribution with unit mass at X. It follows that the Canberra inequality estimator has normal distribution asymptotically. Then

$$\sqrt{n}(\hat{C}_n - C) \xrightarrow{d} N(0, \sigma_C^2),$$

where $\xrightarrow{d}$ denotes convergence in distribution and $\sigma_C^2 = Var(h_C(X))$.

□

7. Simulation results

In an effort to gauge the actual performance of the inferential procedures, we develop a simulation study. We consider the Singh Madalla distribution (with cdf $F(x) = 1 - \frac{1}{[1+(\frac{x}{b})^a]^q}$), the fitted distribution to real income data based on 2013 nominal family income. The data can be obtained from the Census Population Reports which are available online ‖. The shape parameter values of the fitted distribution given by the maximum likelihood estimates are $a = 1.4803, q = 4.0415$ and scale parameter equal to $b = 185.2766$. We remind that b as a scale parameter of the Singh Madalla model does not affect the inequality at all.

In this section, we study by simulation to what extent the Canberra estimator proposed here give reliable inference. First, in order to see whether the asymptotic normality

‖http://www.cesus.gov/hhes/www/income/data/incpovhlth assumption yields a good approximation, simulations were undertaken with drawings from the fitted distribution to real data. The true value of the Canberra index for this distribution (C_0) is easily computed to be 0.4327. In Figure 8, graphs are shown of the empirical distribution function of 10,000 realisations of the statistic $\tau_C = \frac{\hat{C}-C_0}{\hat{\sigma}_{\hat{C}}}$. It can be noted that the estimation of the standard error was obtained using bootstraping method. For sample sizes $n = 50$ and 100 the graph of the standard normal cdf is also given as a benchmark in Figure 8.

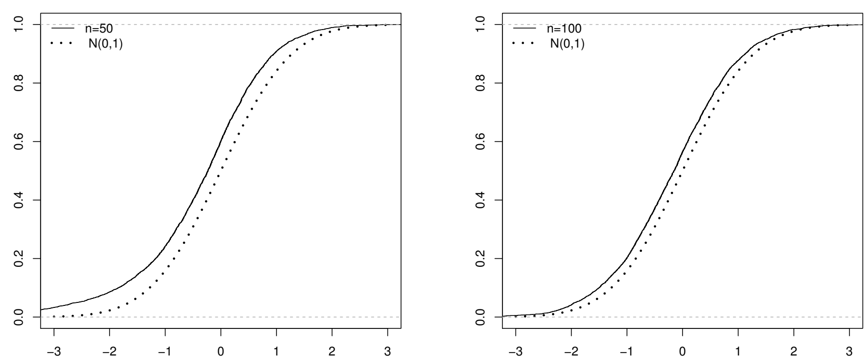

Figure 8. Distribution of Canberra standardized statistic as a function of sample size

It can be seen that, the Canberra estimator is consistent and its asymptotic standard normal is good.

Here, we carried out a simulation study to investigate the behavior of the Canberra index in terms of bias, mean square errors (MSE) and the great absolute deviation of its distribution from the normal distribution function using the Kolmogorov distance under the fitted distribution to real data with mentioned parameters. To find these summary statistics, 10,000 estimate of Canberra index is obtained by taking the sample size n= 10, 20, 30, 50, 70 and 100. The results are shown in Table 2.

Table 2. Summary statistics of the Canberra measure in Singh Madalla model

n	Bias	MSE	Divergence from normal
10	-0.08212663	0.0117186263	0.3987576
20	-0.04392713	0.0050472391	0.2923378
30	-0.03033756	0.0031951613	0.2441087
50	-0.01851632	0.0018111669	0.1976690
70	-0.01283985	0.0012556791	0.1765563
100	-0.01017360	0.0008928519	0.1332017

From Table 2, it is observed that bias, MSE and the great absolute deviation from normal distribution for the Canberra index decrease with increasing n. Also, there is an underestimate for discussed measure in all of the chosen values of sample size.

8. Application to real data

In this section, some empirical Lorenz and Canberra curves for the income distribution of the data coming from the Polish Household Budget Survey (HBS) for the years 2006 and 2008 have been presented.**

The left panel of Figure 9 depicts the Lorenz curves for the considered income distributions. The curves point out a medium-high level of inequality, but actually no relevant difference over years is highlighted. It is difficult to obtain more details about the dynamic evolution of the inequality. In the right panel of Figure 9, the corresponding $C(p)$ curves are drawn. By an analysis of these curves, some dissimilarities among the years can be found. The behavior of the $C(p)$ curves over years clearly shows that the inequality for lower incomes (related to the low values of p, from 0 up to about 0.40) increased from 2006 to 2008, while for higher incomes (related to the high values of p, from 0.40 up to 1) the situation is different: for such kind of incomes, the inequality decreased in the considered time range. Remarks of such kind are difficult to be observed in the Lorenz curve. Thus the Canberra curve seems to be more sensitive to low incomes that the traditional Lorens curve.

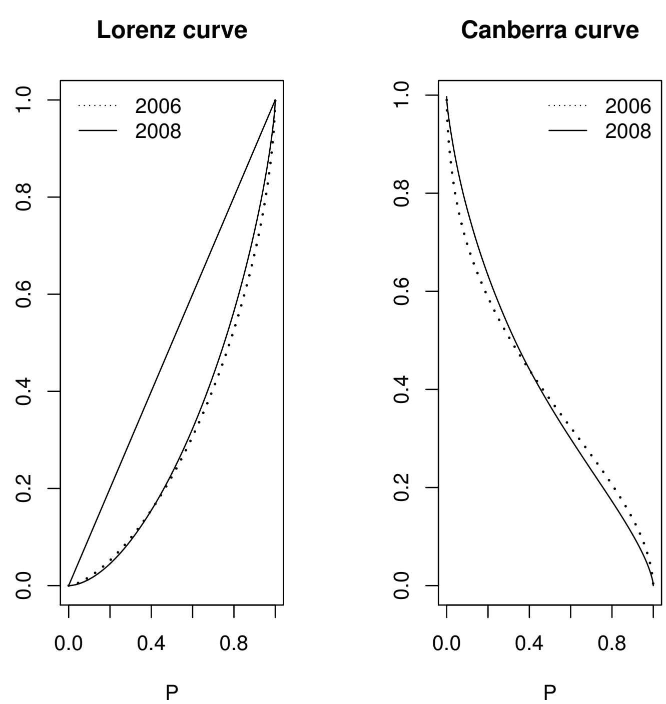

Figure 9. Lorenz curves (left) and Canberra curves (right) for the real income data

**see Jedrzejczak [8]

9. Conclusion

In this note, some important properties for the Canberra inequality index have been discussed. This measure fulfills the most common required properties to an inequality measure. Comparison of Lorenz and Canberra curves highlights main results. Canberra curve can be considered more explanatory and flexible than Lorenz curve. All these characteristics play an essential role, especially in economics and applied statistics. Another important point is about inequality indicators. For the analyzed income distributions, all the direct (respectively inverse) inequality indicators for Lorenz curve are direct (respectively inverse) inequality indicators for Canberra curve and vice versa: this feature implies the consistency between the two curves, although they approach the inequality measuring in two different methods. For these reasons, the Canberra and the related index seem to be a valid alternative to the well-known Lorenz curve and the concentration indices derived from it. This index and its underlying curve display interesting graphical interpretations and it is sensitive to low incomes.

Acknowledgement

The authors are grateful to the anonymous referees for their careful reading and useful comments to improve the quality of the paper.

References

[1] Arcagni, A. and Porro, F. *The graphical reprensentation of inequality*, Revista Colombiana de Estadistica **37** (2), 419-436, 2014.

[2] Arnold, B. C. *On Zenga and Bonferroni curves*, Metron **73** (1), 25-30, 2015.

[3] Bonferroni, C. E. *Elementi di statistica generale*, Università commerciale L. Bocconi, 1961.

[4] Cha, S. H. *Comprehensive survey on distance measures between probability density function*, International Journal of Mathematical Models and Methods in Applied Sciences **4**, 343-356, 2007.

[5] Chakravarty, S. R. *Extended Gini indices of inequality*, International Economic Review **29**, 147-156, 1988.

[6] Gini, C. *Variabilita e mutabilita. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E & Salvemini, T)*, Rome: Libreria Eredi Virgilio Veschi **1**, 1912.

[7] Hoeffding, W. *A class of statistics with asymptotically normal distribution*, Journal of Annals of Mathematical Statistics **19**, 293-325, 1948.

[8] Jedrzejczak, A. *Estimation of concentration measures and their standard errors for income distributions in Poland*, International Advanced Economic Research **18**, 287–297, 2012.

[9] Kakwani, N. *On a class of poverty measures*, Journal of Econometric Society **16**, 437-446, 1980.

[10] Kleiber, C. and Kotz, S. *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley and Sons, New Jersey, Hoboken, 2003.

[11] Lance, G. N. and Williams, W. T. *Mixed Data Classificatory Programs I Agglomerative Systems*, Statistical size distributions in economics and actuarial sciences **1**, 15-20, 1967.

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

[13] Mehran, F. *Linear measures of income inequality*, Journal of Econometric Society **15**, 805-809, 1976.

[14] Pietra, G. *Delle relazioni fra indici di variabilita note I e II*, Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti **74** (2), 775-804, 1915.

[15] Polisicchio, M. and Porro, F. *A comparison between Lorenz $L(p)$ curve and Zenga $I(p)$ curve*, Statistica Applicata **21** (3-4), 289-301, 2009.

[16] Sarabia, J. M. Parametric Lorenz curves: models and applications. *In Modeling Income Distributions and Lorenz Curves*, Springer, New York, 167-190, 2008.

[17] Subramanian, S. *On a distance function-based inequality measure in the spirit of the Bonferroni and Gini Indices*, WIDER Working Paper **62**, 2012.

[18] Zenga, M. *Tendenza alla massima ed alla minima concentrazione per variabili casuali continue*, Statistica **44** (4), 619-640, 1984.

[19] Zenga, M. *Inequality curve and inequality index based on the ratios between lower and upper arithmetic means*, Statistica Applicazioni **5** (1), 3-28, 2007.