Metrika
DOI 10.1007/s00184-011-0369-1

Inference by linearization for Zenga's new inequality index: a comparison with the Gini index

Matti Langel · Yves Tillé

Received: 23 March 2011

Abstract Zenga's new inequality curve and index are two recent tools for measuring inequality. Proposed in 2007, they should thus not be mistaken for anterior measures suggested by the same author. This paper focuses on the new measures only, which are hereafter referred to simply as the Zenga curve and Zenga index. The Zenga curve $Z(\alpha)$ involves the ratio of the mean income of the $100\alpha\%$ poorest to that of the $100(1-\alpha)\%$ richest. The Zenga index can also be expressed by means of the Lorenz Curve and some of its properties make it an interesting alternative to the Gini index. Like most other inequality measures, inference on the Zenga index is not straightforward. Some research on its properties and on estimation has already been conducted but inference in the sampling framework is still needed. In this paper, we propose an estimator and variance estimator for the Zenga index when estimated from a complex sampling design. The proposed variance estimator is based on linearization techniques and more specifically on the direct approach presented by Demnati and Rao. The quality of the resulting estimators are evaluated in Monte Carlo simulation studies on real sets of income data. Finally, the advantages of the Zenga index relative to the Gini index are discussed.

Keywords Inequality · Sampling · Influence function · Gini · Variance estimation

1 Introduction

Research on inequality measures can be conducted in different directions. One direction consists in improving methodology on broadly used statistics such as the

M. Langel (✉) · Y. Tillé
University of Neuchâtel, Pierre à Mazel 7,
2000 Neuchâtel, Switzerland
e-mail: matti.langel@unine.ch

Published online: 17 September 2011

Gini index or entropy measures, while another direction focuses on proposing new inequality measures and places emphasis on the corresponding advantages. It is a fact that the level of income inequality in a population is often accounted for by using the Gini index. Many discussions concerning the latter measure have arisen in statistical and economic literature and a lot of competing inequality measures have been proposed. Some, like the Atkinson index, the Theil index or the Quintile Share Ratio, have been known and used for decades. This paper focuses on finite population inference for a very recent measure, Zenga's new inequality index (Zenga 2007) which is seen as a potential alternative to the Gini index. This new inequality index should not be mistaken for anterior measures proposed some years ago by the same author (Zenga 1984) which are also often referred to as Zenga indices in the literature. For the sake of simplicity, Zenga's new inequality curve and index are hereafter denoted by the terms Zenga curve and Zenga index and respectively expressed by $Z(\alpha)$ and Z.

Like the Gini index, the Zenga index can be expressed by means of the Lorenz curve. However, it is also associated with a new inequality curve, the Zenga curve which provides interesting and direct interpretations on inequality. The paper is structured as follows. In the next section, the index and the curve are defined and an estimator allowing for complex sampling designs is derived. Section 3 is dedicated to variance estimation. Linearization techniques are used to propose a variance estimator for the Zenga index. Some simulations on real data sets are then run in Sect. 4 to apply our theoretical results, while Sect. 5 focuses on comparisons with the Gini index and on the advantages of the Zenga index. The paper ends with some concluding remarks.

2 The Zenga index and Zenga curve

2.1 Definition and notation

Some inequality indices are synthetic values based on an underlying curve or function. The most obvious example is the Gini index and the underlying Lorenz curve. Although the Gini index is the main inequality measure, it does not have unanimous support from statisticians and practitioners and thus has prompted research on other curves and synthetic indices. Zenga (1984) had already proposed two curves as well as the inequality measures λ and ξ. The ξ index and its underlying curve have drawn particular attention (for a review see Zenga 2007), but according to the author, it has not been widely used because it requires estimation of the quantile function as well as of the inverse of the incomplete first moment.

Zenga (2007) has presented a new alternative to the Gini index and other existing inequality measures and curves. Although literature on the Zenga index is not as plentiful as on the Gini index, it has drawn increasing attention in the scientific community. The literature includes some publications on the properties of the index (Maffenini and Polisicchio 2010), inference and applications (Polisicchio 2008; Greselin et al. 2009, 2010) as well as subgroup decomposition (Radaelli 2008, 2010). The literature on the Zenga index and curve also focuses greatly on its advantages compared to the Gini index. Some of these features are described below.

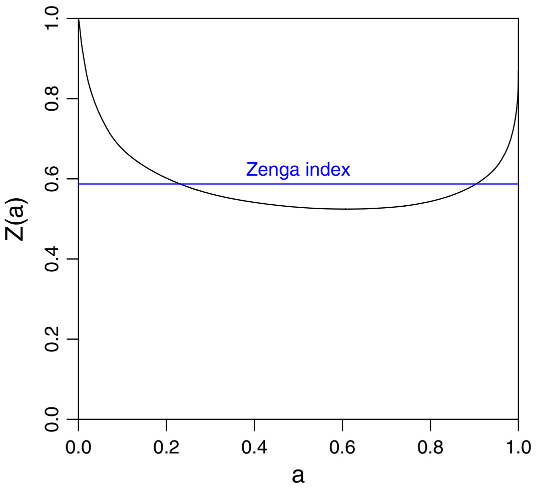


Fig. 1 Example of a Zenga curve and index for a synthetic data set generated from real Austrian EU-SILC survey data (see Sect. 4.1 for more details)

Consider a continuous strictly increasing cumulative distribution function $F(y)$, also, let us denote Q_α, the quantile of order α, such that $F(Q_\alpha) = \alpha$. The quantile function can be written as the inverse of the cumulative distribution function $Q_\alpha = F^{-1}(\alpha)$. The Zenga curve $Z(\alpha)$ is the ratio of the mean income of the poorest $100\alpha\%$ in the distribution to that of the rest of the distribution, namely the $100(1-\alpha)\%$ richest. It is defined by

$$Z(\alpha) = 1 - \frac{L(\alpha)}{\alpha} \cdot \frac{1-\alpha}{1-L(\alpha)},$$

where $0 \leq \alpha \leq 1$ and $L(\alpha)$ is the quantile share or Lorenz curve (Lorenz 1905; Gastwirth 1972; Cowell 1977; Kovacevic and Binder 1997; Langel and Tillé 2011b), which is a central tool of inequality theory and is defined by

$$L(\alpha) = \frac{\int_0^{Q_\alpha} u dF(u)}{\int_0^{\infty} u dF(u)}.$$

The Zenga index, which can be written

$$Z = \int\limits_0^1 Z(\alpha) d\alpha,$$

can thus be defined, like the Gini index, in terms of the Lorenz curve. Figure 1 shows how the index can be plotted together with the Zenga curve and interpreted as a mean level of inequality.

2.2 The Zenga index in finite population

Let U denote a finite population of N identifiable units $u_1, \ldots, u_k, \ldots, u_N$. For the sake of simplicity, we will hereafter denote unit u_k by its identifier k. Associated with each unit k is the value y_k of some characteristic of interest, for example income. To lighten the notation, we will assume with no loss of generality that all y_k's are distinct and sorted. Let us define

$$Y = \sum_{\ell \in U} y_\ell, \tag{2.1}$$

$$Y_k = \sum_{\ell \in U} y_\ell \mathbb{1}[\ell \leq k], \tag{2.2}$$

where $\mathbb{1}(A) = 1$ if A is true and 0 otherwise. As suggested in Langel and Tillé (2011b), let us also denote partial sum $Y(\alpha)$, the sum of incomes up to quantile α by

$$Y(\alpha) = Y_{k-1} + y_k[\alpha N - (k-1)], \tag{2.3}$$

where the value of k is such that $\alpha N < k \leq \alpha N + 1$. With Expression (2.3), the finite population quantile share can be defined by

$$L(\alpha) = \frac{Y(\alpha)}{Y}.$$

The Zenga index for a population of size N is then

$$Z = \sum_{k \in U} Z_k, \tag{2.4}$$

where

$$\begin{aligned}
Z_k &= \int\limits_{(k-1)/N}^{k/N} 1 - \frac{Y(\alpha)}{\alpha} \cdot \frac{1-\alpha}{Y - Y(\alpha)} d\alpha, \\
&= \frac{1}{N} - \int\limits_{(k-1)/N}^{k/N} \frac{Y(\alpha)}{\alpha} \cdot \frac{1-\alpha}{Y - Y(\alpha)} d\alpha, \\
&= \frac{(k-1)y_k - Y_{k-1}}{Y + (k-1)y_k - Y_{k-1}} \log\left(\frac{k}{k-1}\right) \\
&\quad + \left[\frac{Y}{Ny_k} - \frac{Y}{Y + (k-1)y_k - Y_{k-1}}\right] \log\left(\frac{Y - Y_{k-1}}{Y - Y_k}\right).
\end{aligned}$$

We now assume $Y_0 = 0$ and define $A_k = (k-1)y_k - Y_{k-1}$ for all $k \in U$. The above can thus be rewritten

$$Z_k = \begin{cases} \left(\frac{Y}{Ny_1} - 1\right) \log\left(\frac{Y}{Y - Y_1}\right), & \text{if } k = 1, \\ \frac{A_k}{Y + A_k} \log\left(\frac{k}{k-1}\right) + \left[\frac{Y}{Ny_k} - \frac{Y}{Y + A_k}\right] \log\left(\frac{Y - Y_{k-1}}{Y - Y_k}\right), & \text{if } k = 2, \dots, N-1, \\ \left(1 - \frac{Y}{Ny_N}\right) \log\left(\frac{N}{N-1}\right), & \text{if } k = N. \end{cases} \tag{2.5}$$

2.3 An estimator of the Zenga index

A random sample S of size n is drawn from a finite population U of size N from a random sampling design, such that $p(s) = \Pr(S = s)$ is the probability of selecting sample $s \subset U$. The probability for unit $k \in U$ to be included in the sample is written $\pi_k = \Pr(k \in S)$ and d_k denotes the design weight of k such that $d_k = 1/\pi_k$. Note that the design weights d_k are used here for simplicity, and that the following is still valid if the set of weights result from a calibration procedure. Let us also denote

$$D = \sum_{\ell \in S} d_\ell,$$
$$D_k = \sum_{\ell \in S} d_\ell \mathbb{1}[\ell \leq k]$$

and

$$\alpha_k = \frac{D_k}{D}.$$

Expressions (2.1), (2.2) and (2.3) can be respectively estimated from a sample by

$$\widehat{Y} = \sum_{\ell \in S} d_\ell y_\ell,$$
$$\widehat{Y}_k = \sum_{\ell \in S} d_\ell y_\ell \mathbb{1}[\ell \leq k],$$
$$\widehat{Y}(\alpha) = \sum_{\ell \in S} d_\ell y_\ell \mathbb{1}[\ell \leq k-1] + y_k(\alpha D - D_{k-1}) = \widehat{Y}_{k-1} + y_k(\alpha D - D_{k-1}),$$

where k is an integer such that $\alpha_{k-1} < \alpha \leq \alpha_k$. Thus, an estimator for $L(\alpha)$ is

$$\widehat{L}(\alpha) = \frac{\widehat{Y}(\alpha)}{\widehat{Y}}.$$

A natural estimator for the Zenga index is then:

$$\widehat{Z} = \sum_{k \in S} \widehat{Z}_k, \tag{2.6}$$

where

$$\begin{aligned}
\widehat{Z}_k &= \int_{\alpha_{k-1}}^{\alpha_k} 1 - \frac{\widehat{Y}(\alpha)}{\alpha} \cdot \frac{1-\alpha}{\widehat{Y} - \widehat{Y}(\alpha)} d\alpha \\
&= \frac{d_k}{D} - \int_{\alpha_{k-1}}^{\alpha_k} \frac{\widehat{Y}(\alpha)}{\alpha} \cdot \frac{1-\alpha}{\widehat{Y} - \widehat{Y}(\alpha)} d\alpha. \\
&= \frac{D_{k-1} y_k - \widehat{Y}_{k-1}}{\widehat{Y} + D_{k-1} y_k - \widehat{Y}_{k-1}} \log \left(\frac{D_k}{D_{k-1}} \right) \\
&\quad + \left[\frac{\widehat{Y}}{D y_k} - \frac{\widehat{Y}}{\widehat{Y} + D_{k-1} y_k - \widehat{Y}_{k-1}} \right] \log \left(\frac{\widehat{Y} - \widehat{Y}_{k-1}}{\widehat{Y} - \widehat{Y}_k} \right).
\end{aligned}$$

Assuming $\widehat{Y}_0 = 0$, $D_0 = 0$ and defining $\widehat{A}_k = D_{k-1} y_k - \widehat{Y}_{k-1}$ for all $k \in S$, we have:

$$\widehat{Z}_k = \begin{cases}
\left(\frac{\widehat{Y}}{D y_1} - 1 \right) \log \left(\frac{\widehat{Y}}{\widehat{Y} - \widehat{Y}_1} \right), & \text{if } k = 1, \\
\frac{\widehat{A}_k}{\widehat{Y} + \widehat{A}_k} \log \left(\frac{D_k}{D_{k-1}} \right) + \left[\frac{\widehat{Y}}{D y_k} - \frac{\widehat{Y}}{\widehat{Y} + \widehat{A}_k} \right] \log \left(\frac{\widehat{Y} - \widehat{Y}_{k-1}}{\widehat{Y} - \widehat{Y}_k} \right), & \text{if } k = 2, \ldots, n-1, \\
\left(1 - \frac{\widehat{Y}}{D y_n} \right) \log \left(\frac{D_n}{D_{n-1}} \right), & \text{if } k = n.
\end{cases} \tag{2.7}$$

The particular case of inference (point and variance estimator) with non-weighted observations is fully discussed and applied in Greselin et al. (2010).

3 Approximation of the variance by linearization

3.1 Linearization by the Demnati-Rao approach

Linearization regroups a variety of techniques for computing an approximation of the variance of a non-linear statistic $\widehat{\theta}$, an estimator of a function of interest θ. The idea behind these techniques is to find a linearized variable v_ℓ such that

$$\widehat{\theta} - \theta \approx \sum_{\ell \in S} d_\ell v_\ell - \sum_{\ell \in U} v_\ell.$$

The variance of $\sum_{\ell \in S} d_\ell v_\ell$, the weighted sum of the linearized variable v_ℓ, is then used as an approximation of the variance of $\widehat{\theta}$:

$$\operatorname{var}\left(\sum_{\ell \in S} d_\ell v_\ell\right) \approx \operatorname{var}\left(\widehat{\theta}\right).$$

Because the variance of statistic $\widehat{\theta}$ is approximated by the variance of a total, linearization methods can easily provide a variance estimator for all complex sampling designs for which an expression for the variance of a total is known. To compute the values of v_ℓ however, information at the population level is often needed. Thus, v_ℓ is generally replaced by its sample counterpart $\hat{v}_\ell$.

The linearization method was introduced by Woodruff (1971) using Taylor series. Deville (1999) presented a more general method based on influence functions (Hampel 1974; Hampel et al. 1985). In both methods, the linearized variable is computed on the function of interest and is then estimated on the sample. Binder (1996) proposed a direct approach in which the linearized variable is directly computed on the estimator. However, like in Woodruff (1971), it is only adapted for smoothed functions of totals. Demnati and Rao (2004) have proposed yet another direct approach which is of broad application. In the Demnati-Rao approach, we consider weights $a_\ell = d_\ell I_\ell$, for all $\ell \in U$, where

$$I_\ell = \begin{cases} 1 & \text{if } \ell \in S, \\ 0 & \text{if } \ell \notin S. \end{cases}$$

An estimator $\widehat{\theta}$ can be written as a function of the weights a_ℓ: $\widehat{\theta} = f(a_1, a_2, \ldots, a_N)$. The population parameter is obtained by replacing the a_ℓ's by 1's: $\theta = f(1, 1, \ldots, 1)$. By using Taylor series expansion, we can write

$$\widehat{\theta} \approx \theta + \sum_{\ell \in U} (a_\ell - 1) \frac{\partial \widehat{\theta}}{\partial a_\ell}.$$

Thus,

$$\widehat{\theta} - \theta \approx \sum_{\ell \in S} d_\ell \hat{v}_\ell - \sum_{\ell \in U} \hat{v}_\ell,$$

where

$$\hat{v}_\ell = \frac{\partial \widehat{\theta}}{\partial a_\ell} = \frac{\partial \widehat{\theta}}{\partial d_\ell}.$$

In this paper, we use the Demnati and Rao (2004) approach to derive an estimated linearized variable of the Zenga index, and consequently a variance estimator. The estimated linearized variable $\hat{v}_\ell$ is computed directly on the sample and obtained by calculating the partial derivative of the estimator with respect to the weight d_ℓ. Once $\hat{v}_\ell$

is computed, variance estimation is done in the standard framework and usual asymptotic conditions of linearization techniques (Woodruff 1971; Isaki and Fuller 1982; Deville and Särndal 1992; Binder 1996; Kovacevic and Binder 1997; Deville 1999). Note that the design weights d_ℓ are used, but the method holds for calibration weights as well (Demnati and Rao 2004).

3.2 Linearization of the Zenga index

Using the Demnati-Rao approach, the estimated linearized variable $\hat{v}_\ell$ of the Zenga index at y_ℓ can be computed by

$$\hat{v}_\ell = \frac{\partial \widehat{Z}}{\partial d_\ell} = \sum_{k \in S} \frac{\partial \widehat{Z}_k}{\partial d_\ell}. \tag{3.1}$$

Thus, for each sample element ℓ, the partial derivative with respect to d_ℓ of $\widehat{Z}_k$ for all $k \in S$ is computed. Similarly as for point estimation, three cases are derived. We present hereafter the final expressions for $\partial \widehat{Z}_k / \partial d_\ell$ and have included the complete derivation of Expression (3.2) in Appendix A.

$$\frac{\partial \widehat{Z}_k}{\partial d_\ell} = \begin{cases} \dfrac{Dy_\ell - \widehat{Y}}{D^2 y_1} \log\left(\dfrac{\widehat{Y}}{\widehat{Y} - \widehat{Y}_1}\right) + y_\ell \left(\dfrac{\widehat{Y}}{Dy_1} - 1\right)\left[\dfrac{1}{\widehat{Y}} - \dfrac{\mathbb{1}(\ell > 1)}{\widehat{Y} - \widehat{Y}_1}\right], & \text{if } k = 1, \\ \dfrac{\widehat{Y}\,(y_k - y_\ell)\,\mathbb{1}(\ell < k) - \widehat{A}_k y_\ell}{\left(\widehat{Y} + \widehat{A}_k\right)^2} \log\left[\dfrac{D_k\left(\widehat{Y} - \widehat{Y}_{k-1}\right)}{D_{k-1}\left(\widehat{Y} - \widehat{Y}_k\right)}\right] & \\ \quad + \dfrac{\widehat{A}_k}{\widehat{Y} + \widehat{A}_k}\left[\dfrac{\mathbb{1}(\ell \le k)}{D_k} - \dfrac{\mathbb{1}(\ell < k)}{D_{k-1}}\right] + \dfrac{Dy_\ell - \widehat{Y}}{D^2 y_k} \log\left(\dfrac{\widehat{Y} - \widehat{Y}_{k-1}}{\widehat{Y} - \widehat{Y}_k}\right) & \\ \quad + \dfrac{\widehat{Y} y_\ell}{\widehat{Y} - \widehat{Y}_{k-1}}\left[\mathbb{1}(\ell = k) - \dfrac{y_k d_k}{\widehat{Y} - \widehat{Y}_k}\mathbb{1}(\ell > k)\right]\left(\dfrac{1}{Dy_k} - \dfrac{1}{\widehat{Y} + \widehat{A}_k}\right), & \text{if } k = 2, \ldots, n-1, \\ \dfrac{\widehat{Y} - Dy_\ell}{D^2 y_n} \log\left(\dfrac{D}{D_{n-1}}\right) + \left(1 - \dfrac{\widehat{Y}}{Dy_n}\right)\left[\dfrac{1}{D} - \dfrac{\mathbb{1}(\ell < n)}{D_{n-1}}\right], & \text{if } k = n. \end{cases} \tag{3.2}$$

Hence, for example, a variance estimator for the Zenga index under a simple random sampling design without replacement of size n is

$$\widehat{\text{var}}\left(\widehat{Z}\right) = \frac{N(N-n)}{n(n-1)} \sum_{\ell \in S} (\hat{v}_\ell - \bar{v})^2, \tag{3.3}$$

with $\bar{v} = n^{-1} \sum_{\ell \in S} \hat{v}_\ell$.

Table 1 Simulation results (Austrian EU-SILC data, 1,000 replications)

Point estimation		
Z	$\mathrm{E}(\widehat{Z})$	$\mathrm{RB}(\widehat{Z})$
0.5872	0.5870	−0.04%
Variance estimation		
$\widehat{\mathrm{var}}_{sim}\left(\widehat{Z}\right)$	$\mathrm{E}\left[\widehat{\mathrm{var}}_{lin}\left(\widehat{Z}\right)\right]$	$\mathrm{RB}\left(\widehat{\mathrm{var}}_{lin}\right)$
$3.0310 \cdot 10^{-5}$	$2.9811 \cdot 10^{-5}$	−1.65%
Coverage rate of 95% for Z		
	95.9%	

4 Simulation studies

4.1 Synthetic Austrian EU-SILC data

At first, a simulation study is run in the R environment (R Development Core Team 2010) on a synthetic data set generated from original Austrian EU-SILC data. The data is available from the laeken R-package (Alfons et al. 2010) and incorporates 14,824 non-null individual observations from 6,000 households. The simulation design is kept simple: data at the individual level is considered to be the finite population from which random samples of size $n = 3{,}000$ are selected with a simple random sampling design without replacement. One thousand replications are made. In each sample, the Zenga index (Expression 2.6) and its linearization variance (Expression 3.3) are estimated. Results are summarized in Table 1. The relative bias for point and variance estimation are defined respectively by

$$\mathrm{RB}\left(\widehat{Z}\right) = \frac{\mathrm{E}\left(\widehat{Z}\right) - Z}{Z},$$

and

$$\mathrm{RB}\left[\widehat{\mathrm{var}}_{lin}\left(\widehat{Z}\right)\right] = \frac{\mathrm{E}\left[\widehat{\mathrm{var}}_{lin}\left(\widehat{Z}\right)\right] - \widehat{\mathrm{var}}_{sim}\left(\widehat{Z}\right)}{\widehat{\mathrm{var}}_{sim}\left(\widehat{Z}\right)},$$

where $\widehat{\mathrm{var}}_{lin}(\widehat{Z})$ stands for the estimated variance obtained with the linearization technique and $\widehat{\mathrm{var}}_{sim}(\widehat{Z})$ denotes the Monte-Carlo variance computed on the 1000 replications. Results show that point estimation is very successful and that the linearization technique only very slightly underestimates the variance with a relative bias of −1.65%. The coverage rate for a 95% confidence interval is close to the desired level.

4.2 Taxable incomes of Canton of Neuchâtel, Switzerland

In the previous example, extreme observations do not have a large effect on the accuracy of estimation. Our second simulation study is run on real taxable income data in the Canton of Neuchâtel, Switzerland for year 2006. It is composed of all 82,489

Table 2 Simulation results (Neuchâtel data, 1,000 replications)

	Point estimation RB($\widehat{Z}$)	Variance estimation RB ($\widehat{\text{var}}_{lin}$)	Coverage rate CR (95%)
Full data	−0.08%	−6.79%	93.5%
Truncated data	−0.06%	0.22%	94.7%

non-null taxpayers of the canton and includes some extreme observations. The same strategy, design and sample size are used as for the first simulation study in order to allow for comparisons. To account for the extreme observations issue, two sets of simulations are performed: one on the full data set and one on truncated data. In the truncated data, all observations lying above $Q_{0.999}$ are deleted, involving the 83 richest income earners. Truncation of the data reduces the ratio between the largest income and the median income by a factor of 13.2. The results are summarized in Table 2. Note that estimates and true values of the Zenga index and its sampling variance for the Neuchâtel data are not displayed in Table 2 because this data set has been made available to us exclusively for academic purposes. Thus, the quality of estimation for this data is merely summarized by relative biases and coverage rates. For similar reasons, income values in Fig. 2 as well as Zenga and Gini index estimates in Fig. 3 have been masked.

Although point estimation is accounted for in a satisfactory manner in both situations, we can see that the variance is not as well estimated when the most severely extreme observations are part of the population. However, it can be advocated that even in the presence of extreme values, which we believe to be frequent when working on income data, quality of inference for the Zenga index remains reasonable with a relative bias for the variance of −6.79% and a coverage rate of 93.5 for a 95% confidence interval.

5 Comparison with the Gini index

5.1 Definition and properties of the Gini index

Working on a new synthetic inequality index like the Zenga index raises one key question: Do we need yet another inequality measure? The question is not without merit considering the vast collection of already existing inequality measures and the amount of research dedicated to enhancing the quality of inference for these measures. However, by comparing the Zenga index to the leading inequality measure, the Gini index, we try to point out why the Zenga index is a serious and interesting alternative to existing indices. Let us first define the Gini index G by

$$G = 1 - 2\int_0^1 L(\alpha)d\alpha,$$

an estimator of G by

$$\widehat{G} = \frac{2}{D\widehat{Y}} \sum_{k\in S} d_k D_k y_k - \left(1 + \frac{1}{D\widehat{Y}} \sum_{k\in S} d_k^2 y_k\right) = \frac{\sum_{k\in S}\sum_{\ell\in S} d_k d_\ell \,|y_k - y_\ell|}{2D\widehat{Y}},$$

and a linearized variable estimated on the sample (Monti 1991; Langel and Tillé 2011a) by

$$\widehat{u}_k = \frac{1}{D\widehat{Y}}\left[2D_k(y_k - \widehat{\overline{Y}}_k) + \widehat{Y} - Dy_k - \widehat{G}(\widehat{Y} + y_k D)\right],$$

with

$$\widehat{\overline{Y}}_k = \frac{\sum_{\ell\in S} d_\ell y_\ell \mathbb{1}(y_\ell \le y_k)}{D_k}.$$

Both indices have thus in common that they can be defined by means of the Lorenz curve $L(\alpha)$. Also, both measures fulfill the most common properties of the axiomatic approach to inequality theory (Cowell and Kuga 1981) such as anonymity, scale invariance, population principle or principle of transfers (Zenga 2007). Moreover, Radaelli (2010) proposed a subgroup decomposition of the Zenga index which is closely related to the decomposition of the Gini index (Dagum 1997).

5.2 Advantages of the Zenga index

The two measures differ however in many ways. One argument in favor of the Zenga index is described by Greselin et al. (2010, p. 3):

> [. . .] the Gini index underestimates comparisons between the very poor and the whole population and emphasizes comparisons which involve almost identical population subgroups [. . .] the Zenga index detects, with the same sensibility, all deviations from equality in any part of the distribution.

A comparative simulation study regrouping 17 different inequality indices (Langel and Tillé 2009) seems to confirm this idea by showing that the Zenga index is one of the most appropriate measures to detect changes at any level of the income distribution and in many different situations. Another argument in favor of the Zenga index concerns interpretation. A lot of intuitive information can indeed be obtained from analyzing the curve itself. For instance, any point measure $Z(\alpha)$ on the curve indicates that the mean income of the $100\alpha\%$ poorest is equal to $[1 - Z(\alpha)]$ times the mean income of the richest $100(1-\alpha)\%$. Moreover, the Zenga index can be plotted alongside the curve and thus, clearly displays the intervals of α where inequality is lower or higher than the mean level of inequality, which is represented by the index itself. Finally, Maffenini and Polisicchio (2010) have shown that when adding an identical positive income to all observations, the effect on the Zenga curve is more intuitive than on the Lorenz curve. Indeed, the Zenga curve shows that, after translation, the level of inequality decreases more heavily for small incomes than for larger ones, whereas the latter intuition is not captured by the Lorenz curve.

5.3 Influence functions and sampling distributions

In statistics, influence functions (Hampel 1974) are mainly used as a tool to study robustness. However, Deville (1999) showed that the linearized variable is an influence function, only very slightly modified from the definition of Hampel (1974) so that it could be used within a finite population framework. Thus, it is possible to study the sensitivity to extreme observations of the statistic of interest simply by analyzing its linearized variable, or influence curve. Unsurprisingly, the influence curve of the Zenga index shows that the statistic is highly sensitive to extreme observations. As a result, inference can be heavily affected by the presence of very large incomes in the sample. Similar results are found in robust statistics regarding the influence function of the Gini index (Monti 1991; Cowell and Victoria-Feser 1996, 2003) and the Quintile Share Ratio (Hulliger and Schoch 2009), which are both unbounded from above.

However, a comparative study with the Gini index reveals an interesting result. Figure 2 displays the influence function of the Gini index alongside that of the Zenga index computed on one sample of size $n = 3{,}000$ from the Neuchâtel simulation study. To allow for comparisons, both influence functions are normalized following the notion of *relative influence function* proposed by Cowell and Flachaire (2007). The estimated linearized variable of the Gini and Zenga indices are thus divided by the value of the respective index estimated on the sample. Figure 2 shows that the Zenga index is significatively less affected by extreme observations than the Gini index. This is a very important advantage of the Zenga index because inference from income data is often confronted with extreme values.

The outcome of this feature is that the sampling distribution of the Zenga index is by far closer to the Normal distribution than that of the Gini index, allowing for the construction of more reliable confidence intervals. This intuition is confirmed by a small simulation study performed to estimate the skewness and excess kurtosis of the sampling distribution of both indices $\widehat{G}$ and $\widehat{Z}$. We have simulated 1000 samples (simple random sampling design without replacement) of size $n = 100$ from the

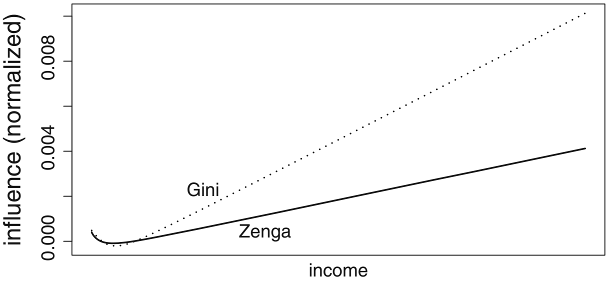
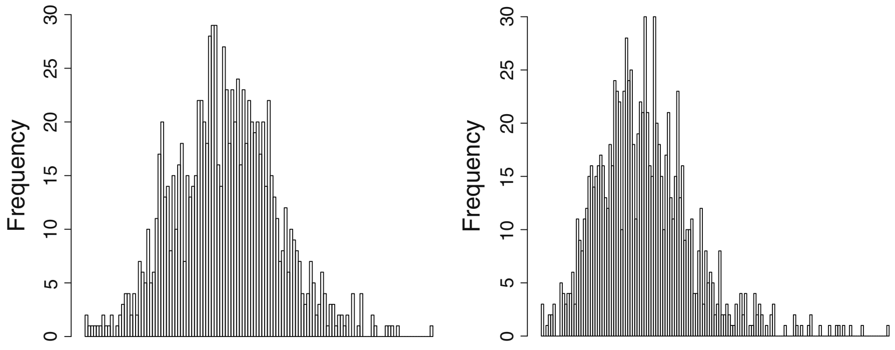


Fig. 2 Normalized influence curves of the Zenga index and Gini index estimated on one sample of size $n = 3{,}000$ drawn from the Neuchâtel data set



(a) Zenga index $\widehat{Z}$ **(b)** Gini index $\widehat{G}$

Fig. 3 Histograms of the distributions of $\widehat{Z}$ and $\widehat{G}$ computed on 1,000 samples of size $n = 100$

Table 3 Skewness and kurtosis: simulation results

	Zenga index $\widehat{Z}$	Gini index $\widehat{G}$
Skewness	0.22	1.15
Excess kurtosis	0.24	2.22

Neuchâtel income data and estimated both indices in each sample. The histograms in Fig. 3 displays the respective sampling distributions of $\widehat{G}$ and $\widehat{Z}$ for this set of simulations. They show that the sampling distribution of the Zenga index is clearly more symmetric than that of the Gini index.

The skewness and excess kurtosis for each index are then estimated on the 1,000 samples. The results, displayed in Table 3, show that unlike what is observed with the Zenga index, the sampling distribution of the Gini index is a serious obstacle in the construction of good confidence intervals around $\widehat{G}$. Indeed, the skewness and excess kurtosis of $\widehat{G}$ are far from the desired level (0 for both statistics).

6 Conclusion

To effectively bring new insights in the study of income inequality a recent measure like the Zenga index needs a general and valid framework for inference in finite populations. In this paper, we have firstly proposed an estimator of the Zenga index which takes the sampling design into account. Secondly, a variance estimator has been presented. The Demnati and Rao linearization technique has been used to derive an estimator that can be applied to samples selected from a complex sampling design. The theoretical results have then been tested successfully in simulation studies. Finally the relevance of the Zenga index has been emphasized by comparing it to the Gini index. It it shown that in addition to having similar properties as the Gini index, the characteristics of the Zenga index facilitate reliable inference. Indeed, in the presence of extreme observations, the sampling distribution of $\widehat{Z}$ is both markedly less skewed and less heavy-tailed than that of the Gini index. Moreover, the Zenga index and its underlying curve display interesting graphical interpretations. We hope that these features can motivate the use of the Zenga index in future research studies and applications.

Acknowledgments The authors are grateful to the Office Cantonal de la Statistique (Canton de Neuchâtel, Switzerland) and especially to Gérard Geiser for the dataset. The authors would also like to thank two anonymous referees for insightful comments and suggestions. This research is supported by grant no. 200021-121604 of the Swiss National Science Foundation.

Appendix A: Linearization

The three cases of Expression (2.7) are linearized separately in order to obtain an estimated linearized variable $\hat{v}_\ell$ such that

$$\hat{v}_\ell = \frac{\partial \widehat{Z}}{\partial d_\ell} = \sum_{k \in S} \frac{\partial \widehat{Z}_k}{\partial d_\ell}. \tag{A.1}$$

A.1 Linearization of $\widehat{Z}_k$ for $k = 2, \ldots, n-1$

First, $\widehat{Z}_k$ for $k = 2, \ldots, n$ is rewritten as the sum of two terms, P_1 and P_2:

$$\begin{aligned}\widehat{Z}_k &= \underbrace{\frac{\widehat{A}_k}{\widehat{Y} + \widehat{A}_k} \log\left(\frac{D_k}{D_{k-1}}\right)}_{P_1} + \underbrace{\left[\frac{\widehat{Y}}{Dy_k} - \frac{\widehat{Y}}{\widehat{Y} + \widehat{A}_k}\right] \log\left(\frac{\widehat{Y} - \widehat{Y}_{k-1}}{\widehat{Y} - \widehat{Y}_k}\right)}_{P_2}. \\ &= P_1 + P_2.\end{aligned}$$

Thus,

$$\frac{\partial \widehat{Z}_k}{\partial d_\ell} = \frac{\partial P_1}{\partial d_\ell} + \frac{\partial P_2}{\partial d_\ell}, \tag{A.2}$$

and the derivation can be split into two separate steps, the linearization of terms P_1 and P_2.

Linearization of term P_1 can be done by computing the partial derivative with respect to d_ℓ. Using differentiation rules, we obtain

$$\begin{aligned}\frac{\partial P_1}{\partial d_\ell} &= \frac{(\widehat{Y} + \widehat{A}_k) \log\left(\frac{D_k}{D_{k-1}}\right) \frac{\partial \widehat{A}_k}{\partial d_\ell} - \widehat{A}_k \log\left(\frac{D_k}{D_{k-1}}\right) \frac{\partial [\widehat{Y} + \widehat{A}_k]}{\partial d_\ell}}{(\widehat{Y} + \widehat{A}_k)^2} \\ &\quad + \frac{D_{k-1}}{D_k} \frac{\partial \left(\frac{D_k}{D_{k-1}}\right)}{\partial d_\ell} \frac{\widehat{A}_k}{\widehat{Y} + \widehat{A}_k}.\end{aligned} \tag{A.3}$$

We now compute the derivatives that are needed in Eq. (A.3):

$$\frac{\partial \widehat{Y}}{\partial d_\ell} = y_\ell, \tag{A.4}$$

$$\frac{\partial \widehat{A}_k}{\partial d_\ell} = (y_k - y_\ell)\, \mathbb{1}(\ell < k), \tag{A.5}$$

$$\frac{\partial\left(\frac{D_k}{D_{k-1}}\right)}{\partial d_\ell}=\frac{D_{k-1}\mathbb{1}(\ell\leq k)-D_k\mathbb{1}(\ell<k)}{D_{k-1}^2},\tag{A.6}$$

and replace them in Expression (A.3) to obtain

$$\begin{aligned}\frac{\partial P_1}{\partial d_\ell}=\frac{\widehat{A}_k}{\widehat{Y}+\widehat{A}_k}\Bigg[&\frac{\widehat{Y}\,(y_k-y_\ell)\,\mathbb{1}(\ell<k)-\widehat{A}_k y_\ell}{\widehat{A}_k(\widehat{Y}+\widehat{A}_k)}\log\left(\frac{D_k}{D_{k-1}}\right)\\&-\frac{\mathbb{1}(\ell<k)}{D_{k-1}}+\frac{\mathbb{1}(\ell\leq k)}{D_k}\Bigg].\end{aligned}\tag{A.7}$$

Similarly, for term P_2:

$$\begin{aligned}\frac{\partial P_2}{\partial d_\ell}=&\left[\frac{\widehat{Y}\frac{\partial[\widehat{Y}+\widehat{A}_k]}{\partial d_\ell}}{(\widehat{Y}+\widehat{A}_k)^2}-\frac{\frac{\partial\widehat{Y}}{\partial d_\ell}}{\widehat{Y}+\widehat{A}_k}+\frac{\frac{\partial\widehat{Y}}{\partial d_\ell}}{Dy_k}-\frac{\widehat{Y}\frac{\partial(Dy_k)}{\partial d_\ell}}{(Dy_k)^2}\right]\log\left(\frac{\widehat{Y}-\widehat{Y}_{k-1}}{\widehat{Y}-\widehat{Y}_k}\right)\\&+\frac{\widehat{Y}-\widehat{Y}_k}{\widehat{Y}-\widehat{Y}_{k-1}}\frac{\partial\left(\frac{\widehat{Y}-\widehat{Y}_{k-1}}{\widehat{Y}-\widehat{Y}_k}\right)}{\partial d_\ell}\left[\frac{\widehat{Y}}{Dy_k}-\frac{\widehat{Y}}{\widehat{Y}+\widehat{A}_k}\right].\end{aligned}\tag{A.8}$$

In addition to Result (A.4), the following derivatives are needed:

$$\frac{\partial(\widehat{Y}+\widehat{A}_k)}{\partial d_\ell}=y_\ell-(y_\ell-y_k)\,\mathbb{1}(\ell<k).\tag{A.9}$$

$$\frac{\partial Dy_k}{\partial d_\ell}=y_k,\tag{A.10}$$

$$\frac{\partial\left(\frac{\widehat{Y}-\widehat{Y}_{k-1}}{\widehat{Y}-\widehat{Y}_k}\right)}{\partial d_\ell}=\frac{y_\ell}{\widehat{Y}-\widehat{Y}_k}\left[\mathbb{1}(\ell=k)-\frac{y_k d_k}{\widehat{Y}-\widehat{Y}_k}\mathbb{1}(\ell>k)\right].\tag{A.11}$$

Results (A.4), (A.9), (A.10) and (A.11) are substituted into Eq. (A.8):

$$\begin{aligned}\frac{\partial P_2}{\partial d_\ell}=&\left[\frac{y_\ell\widehat{A}_k-\widehat{Y}\,(y_\ell-y_k)\,\mathbb{1}(\ell<k)}{(\widehat{Y}+\widehat{A}_k)^2}+\frac{Dy_\ell-\widehat{Y}}{D^2y_k}\right]\log\left(\frac{\widehat{Y}-\widehat{Y}_{k-1}}{\widehat{Y}-\widehat{Y}_k}\right)\\&+\frac{\widehat{Y}y_\ell}{\widehat{Y}-\widehat{Y}_{k-1}}\left[\mathbb{1}(\ell=k)-\frac{y_k d_k}{\widehat{Y}-\widehat{Y}_k}\mathbb{1}(\ell>k)\right]\left(\frac{1}{Dy_k}-\frac{1}{\widehat{Y}+\widehat{A}_k}\right).\end{aligned}\tag{A.12}$$

The final expression for the linearization of $\widehat{Z}_k$ for $k=2,\ldots,n-1$ is obtained by replacing (A.7) and (A.12) in (A.2):

$$\begin{aligned}\frac{\partial\widehat{Z}_k}{\partial d_\ell}&=\frac{\partial P_1}{\partial d_\ell}+\frac{\partial P_2}{\partial d_\ell}\\&=\frac{\widehat{A}_k}{\widehat{Y}+\widehat{A}_k}\left[\frac{\widehat{Y}\,(y_k-y_\ell)\,\mathbb{1}(\ell<k)-\widehat{A}_k y_\ell}{\widehat{A}_k(\widehat{Y}+\widehat{A}_k)}\log\left(\frac{D_k}{D_{k-1}}\right)-\frac{\mathbb{1}(\ell<k)}{D_{k-1}}+\frac{\mathbb{1}(\ell\leq k)}{D_k}\right]\end{aligned}$$

$$+\left[\frac{\widehat{Y}\,(y_k-y_\ell)\,\mathbb{1}(\ell<k)-\widehat{A}_k y_\ell}{(\widehat{Y}+\widehat{A}_k)^2}+\frac{Dy_\ell-\widehat{Y}}{D^2 y_k}\right]\log\left(\frac{\widehat{Y}-\widehat{Y}_{k-1}}{\widehat{Y}-\widehat{Y}_k}\right)$$

$$+\frac{\widehat{Y}y_\ell}{\widehat{Y}-\widehat{Y}_{k-1}}\left[\mathbb{1}(\ell=k)-\frac{y_k d_k}{\widehat{Y}-\widehat{Y}_k}\mathbb{1}(\ell>k)\right]\left(\frac{1}{Dy_k}-\frac{1}{\widehat{Y}+\widehat{A}_k}\right).$$

The latter can be rewritten by

$$\frac{\partial\widehat{Z}_k}{\partial d_\ell}=\frac{\widehat{Y}\,(y_k-y_\ell)\,\mathbb{1}(\ell<k)-\widehat{A}_k y_\ell}{\left(\widehat{Y}+\widehat{A}_k\right)^2}\log\left[\frac{D_k\left(\widehat{Y}-\widehat{Y}_{k-1}\right)}{D_{k-1}\left(\widehat{Y}-\widehat{Y}_k\right)}\right]$$
$$+\frac{\widehat{A}_k}{\widehat{Y}+\widehat{A}_k}\left[\frac{\mathbb{1}(\ell\le k)}{D_k}-\frac{\mathbb{1}(\ell<k)}{D_{k-1}}\right]+\frac{Dy_\ell-\widehat{Y}}{D^2 y_k}\log\left(\frac{\widehat{Y}-\widehat{Y}_{k-1}}{\widehat{Y}-\widehat{Y}_k}\right)$$
$$+\frac{\widehat{Y}y_\ell}{\widehat{Y}-\widehat{Y}_{k-1}}\left[\mathbb{1}(\ell=k)-\frac{y_k d_k}{\widehat{Y}-\widehat{Y}_k}\mathbb{1}(\ell>k)\right]\left(\frac{1}{Dy_k}-\frac{1}{\widehat{Y}+\widehat{A}_k}\right). \qquad (A.13)$$

A.2 Linearization of $\widehat{Z}_1$

The case for $k=1$ can be derived:

$$\frac{\partial\widehat{Z}_1}{\partial d_\ell}=\log\left(\frac{\widehat{Y}}{\widehat{Y}-\widehat{Y}_1}\right)\left[\frac{Dy_1\frac{\partial\widehat{Y}}{\partial d_\ell}-\widehat{Y}\frac{\partial(Dy_1)}{\partial d_\ell}}{(Dy_1)^2}\right]+\left(\frac{\widehat{Y}}{Dy_1}-1\right)\left(\frac{\widehat{Y}-\widehat{Y}_1}{\widehat{Y}}\right)\frac{\partial\left(\frac{\widehat{Y}}{\widehat{Y}-\widehat{Y}_1}\right)}{\partial d_\ell},$$
$$=\frac{Dy_\ell-\widehat{Y}}{D^2 y_1}\log\left(\frac{\widehat{Y}}{\widehat{Y}-\widehat{Y}_1}\right)+y_\ell\left(\frac{\widehat{Y}}{Dy_1}-1\right)\left[\frac{1}{\widehat{Y}}-\frac{\mathbb{1}(\ell>1)}{\widehat{Y}-\widehat{Y}_1}\right]. \qquad (A.14)$$

A.3 Linearization of $\widehat{Z}_n$

Finally the $k=n$ case is also linearized:

$$\frac{\partial\widehat{Z}_n}{\partial d_\ell}=\log\left(\frac{D}{D_{n-1}}\right)\left[\frac{\widehat{Y}\frac{\partial(Dy_n)}{\partial d_\ell}-Dy_n\frac{\partial\widehat{Y}}{\partial d_\ell}}{(Dy_n)^2}\right]+\left(1-\frac{\widehat{Y}}{Dy_n}\right)\frac{D_{n-1}}{D}\frac{\partial\left(\frac{D}{D_{n-1}}\right)}{\partial d_\ell},$$
$$=\frac{\widehat{Y}-Dy_\ell}{D^2 y_n}\log\left(\frac{D}{D_{n-1}}\right)+\left(1-\frac{\widehat{Y}}{Dy_n}\right)\left[\frac{1}{D}-\frac{\mathbb{1}(\ell<n)}{D_{n-1}}\right]. \qquad (A.15)$$

A.4 Linearization of the Zenga index

Finally, by recalling Expression (A.1) and combining Results (A.13), (A.14) and (A.15) into

$$\frac{\partial \widehat{Z}_k}{\partial d_\ell} = \begin{cases} \dfrac{Dy_\ell - \widehat{Y}}{D^2 y_1} \log\left(\dfrac{\widehat{Y}}{\widehat{Y} - \widehat{Y}_1}\right) + y_\ell \left(\dfrac{\widehat{Y}}{Dy_1} - 1\right)\left[\dfrac{1}{\widehat{Y}} - \dfrac{\mathbb{1}(\ell > 1)}{\widehat{Y} - \widehat{Y}_1}\right], & \text{if } k = 1, \\ \dfrac{\widehat{Y}\,(y_k - y_\ell)\,\mathbb{1}(\ell < k) - \widehat{A}_k y_\ell}{\left(\widehat{Y} + \widehat{A}_k\right)^2} \log\left[\dfrac{D_k\left(\widehat{Y} - \widehat{Y}_{k-1}\right)}{D_{k-1}\left(\widehat{Y} - \widehat{Y}_k\right)}\right] & \\ \quad + \dfrac{\widehat{A}_k}{\widehat{Y} + \widehat{A}_k}\left[\dfrac{\mathbb{1}(\ell \le k)}{D_k} - \dfrac{\mathbb{1}(\ell < k)}{D_{k-1}}\right] + \dfrac{Dy_\ell - \widehat{Y}}{D^2 y_k} \log\left(\dfrac{\widehat{Y} - \widehat{Y}_{k-1}}{\widehat{Y} - \widehat{Y}_k}\right) & \\ \quad + \dfrac{\widehat{Y} y_\ell}{\widehat{Y} - \widehat{Y}_{k-1}}\left[\mathbb{1}(\ell = k) - \dfrac{y_k d_k}{\widehat{Y} - \widehat{Y}_k}\mathbb{1}(\ell > k)\right]\left(\dfrac{1}{Dy_k} - \dfrac{1}{\widehat{Y} + \widehat{A}_k}\right), & \text{if } k = 2, \ldots, n-1, \\ \dfrac{\widehat{Y} - Dy_\ell}{D^2 y_n} \log\left(\dfrac{D}{D_{n-1}}\right) + \left(1 - \dfrac{\widehat{Y}}{Dy_n}\right)\left[\dfrac{1}{D} - \dfrac{\mathbb{1}(\ell < n)}{D_{n-1}}\right], & \text{if } k = n, \end{cases}$$

an estimated linearized variable $\hat{v}_\ell$ can now be obtained for all $\ell \in S$, and thus a variance estimator for $\widehat{Z}$, the Zenga index estimated from a sample.

References

Alfons A, Holzer J, Templ M (2010) laeken: Laeken indicators for measuring social cohesion. R package version 0.1.3

Binder DA (1996) Linearization methods for single phase and two-phase samples: a cookbook approach. Surv Methodol 22: 17–22

Cowell FA, Victoria-Feser M-P (2003) Distribution-free inference for welfare indices under complete and incomplete information. J Econ Inequality 1:191–219

Cowell FA (1977) Measuring inequality. Philip Allan, Oxford

Cowell FA, Flachaire E (2007) Income distribution and inequality measurement: the problem of extreme values. J Econom 141: 1044–1072

Cowell FA, Kuga K (1981) Inequality measurement: an axiomatic approach. Eur Econ Rev 15:287–305

Cowell FA, Victoria-Feser M-P (1996) Poverty measurement with contaminated data: a robust approach. Eur Econ Rev 40: 1761–1771

Dagum C (1997) A new approach to the decomposition of the Gini income inequality ratio. Empir Econ 22:515–531

Demnati A, Rao JNK (2004) Linearization variance estimators for survey data (with discussion). Surv Methodol 30:17–34

Deville J-C (1999) Variance estimation for complex statistics and estimators: linearization and residual techniques. Surv Methodol 25:193–204

Deville J-C, Särndal C-E (1992) Calibration estimators in survey sampling. J Am Stat Assoc 87:376–382

Gastwirth JL (1972) The estimation of the Lorenz curve and Gini index. Rev Econ Stat 54:306–316

Greselin F, Pasquazzi L, and Zitikis R (2010) Zenga's new index of economic inequality, its estimation, and an analysis of incomes in Italy. J Probab Stat, ID 718905:1–26

Greselin F, Puri M, Zitikis R (2009) L-functions, processes, and statistics in measuring economic inequality and actuarial risks. Stat Interfaces 2:227–245

Hampel FR (1974) The influence curve and its role in robust estimation. J Am Stat Assoc 69:383–393

Hampel FR, Ronchetti E, Rousseuw PJ, Stahel W (1985) Robust statistics: the approach based on the influence function. Wiley, New York

Hulliger B, Schoch T (2009) Robustification of the quintile share ratio. In: Proceedings of the Colloque sur les méthodes de sondage en l'honneur de Jean-Claude Deville, Neuchâtel, Switzerland

Isaki CT, Fuller WA (1982) Survey design under a regression population model. J Am Stat Assoc 77:89–96

Kovacevic MS, Binder DA (1997) Variance estimation for measures of income inequality and polarization—the estimating equations approach. J Off Stat 13:41–58

Langel M, Tillé Y (2009) An evaluation of the performance of inequality measures for the detection of changes in an income distribution. Technical report, University of Neuchatel

Langel M, Tillé Y (2011a) Corrado Gini, a pioneer in balanced sampling and inequality theory. Metron 69:43–63

Langel M, Tillé Y (2011b) Statistical inference for the quintile share ratio. J Stat Plan Inference 141:2976–2985

Lorenz MO (1905) Methods of measuring the concentration of wealth. Publ Am Stat Assoc 9:209–219

Maffenini W, Polisicchio M (2010) How potential is the I(p) inequality curve in the analysis of empirical distributions. Technical report, Technical report, Universita degli Studi di Milano-Bicocca

Monti AC (1991) The study of the Gini concentration ratio by means of the influence function. Statistica 51:561–577

Polisicchio M (2008) The continuous random variable with uniform point inequality measure I(p). Statistica e Applicazioni 6:137–151

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0

Radaelli P (2008) A subgroups decomposition of Zenga's uniformity and inequality indexes. Statistica e Applicazioni 6:117–136

Radaelli P (2010) On the decomposition by subgroups of the Gini index and Zenga's uniformity and inequality indexes. Int Stat Rev 78:81–101

Woodruff RS (1971) A simple method for approximating the variance of a complicated estimate. J Am Stat Assoc 66:411–414

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 43:301–326

Zenga M (2007) Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. Statistica e Applicazioni 4:3–27