Statistica & Applicazioni - Vol. XIII, n. 2, 2015, pp. 143-162

ASYMPTOTIC PROPERTIES OF SOME ESTIMATORS FOR GINI AND ZENGA INEQUALITY MEASURES: A SIMULATION STUDY

Alina Jedrzejczak*

SUMMARY

It is well known that unequal income distribution, yielding poverty, stratification and polarization, can be a serious economic and social problem. The reliable inequality analysis of both, total population of households and subpopulations classified by different characteristics, can be a helpful piece of information for economists and social policy- makers. Therefore, it seems especially important to present reliable estimates of income inequality measures for a population of households in different divisions. Among many income inequality measures the Gini index based on the Lorenz curve is the most popular one. Another interesting measure of income inequality is the Zenga index proposed in 1984, based on the relation between income and population quantiles. In the paper some nonparametric estimators of Gini and Zenga inequality measures are presented and analyzed from a point of view of their statistical properties. In particular, the bias, dispersion and normality of the estimators are considered. The Monte Carlo experiments include the cases of heavy-tailed and light-tailed distributions as theoretical models. Finally, the estimators are applied to the data on income distributions in Poland.

Keywords: *Income Distribution, Zenga Income Inequality Index, Gini Inequality Index.*

1. INTRODUCTION

The true values of income inequality coefficients are usually unknown and they can only be estimated on the basis of sample data coming from various household budget surveys. In many applications, however, the estimates of inequality measures are presented without any information about their precision which should be the basis for further statistical inference e.g. hypothesis testing and interval estimation. The problem can be neglected to some extent when we consider the overall population and the sample size is large enough to apply the asymptotic theory; one should be conscious however, that for the heavy-tailed income distributions the sufficient sample size can be very large indeed. For some population divisions (by age, occupation, family type or geographical area) estimators of inequality measures can be seriously biased and their standard errors can be far beyond the values that can be accepted by social policy-makers for making reliable policy decisions. The situation seems even more complicated as inequality coefficients are usually nonlinear sample statistics and their standard errors cannot be easily obtained. It can even be worse

[1] Department of Statistics and Demography - University of Lodz - via Rewolucji 1905r. 41. - 90214 LODZ (e-mail: jedrzej@uni.lodz.pl).

when we consider complex sampling designs instead of simple random sampling with replacement. The methods of variance estimation that can solve this problem include: various replication techniques, Taylor expansion, and parametric procedures based on income distribution models (see Davidson, 2009; Jedrzejczak, 2012). Moreover, also the normality may not be preserved what can be a serious disadvantage in case of interval estimation.

The main objective of the paper is to empirically testify basic statistical properties of some nonparametric estimators proposed for Gini and Zenga inequality coefficients in the presence of various population characteristics and sample sizes. The properties of the parametric estimators of Gini and Zenga (1984) coefficients, based on the Dagum and lognormal models, have been studied by Latorre (1990). The nonparametric estimators considered in the present paper can be used in wider range of applications, including survey sampling approach, as they can account for non-equal probability weighting usually applied in practice. Moreover, these distribution-free statistics can also be applied for small samples, when the underlying income distribution model cannot be assumed.

The problem of developing appropriate measures of income inequality has received considerable attention in the economic literature since the turn of the twentieth century. Among numerous income inequality measures the Gini index based on the Lorenz function, widely known as the ''Lorenz curve'', has become the most popular one. The Gini index can be described by several mathematical representations – each of them can be given its own interpretation and naturally leads to different estimation formulas (see Yitzhaki and Schechtman, 2013). Other important measures of income inequality were proposed by Bonferroni (1930) and Zenga (1984, 2007). The Zenga (1984) index, based on the relation between income and population quantiles, has attracted much attention in the literature during the last few decades due to its outstanding statistical properties. It has proven to present several advantages over the Gini ratio, being easily decomposable and sensitive to changes at every ''point'' of income distribution i.e. detecting all deviations from equality in any part of the distribution with the same sensibility.

First of all, it is necessary to recall the Gini significant contributions to the study of income inequality that are based on two inequality curves and on the corresponding point and synthetic inequality indices; the synthetic indices are arithmetic means of the point indices.

Let $F(y) = \mathrm{P}(Y \leq y)$ denote the cumulative distribution function of a random variable Y, which we assume to be nonnegative throughout the paper. The Lorenz function $L(p)$ can be expressed by the following formula

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

where $\mu = \mathrm{E(Y)}$ denotes the positive and finite expected value of a random variable Y and $F^{-1}(p) = \inf\{x : \mathrm{f}(x) \leq p\}$ stands for its p^{th} quantile.

In 1909 Gini proposed the point index $\delta(p) = \log(1-p)/\log(1-\mathrm{L}(p))$. The

corresponding synthetic index is $\delta = \int_0^1 \delta(p)\, dp$. As the point measure $\delta(p)$ has a drawback that it is not normalized ($\delta(p) \geq 1$), in 1984 Zenga normalized it with $\lambda(p) = (\delta(p) - 1)/\ \delta(p)$. Note that $\lambda(p)$ does not have forced behaviour and in the case of the Pareto model $\lambda(p) = 1/\theta$.

The popular Gini index of inequality was proposed in 1914 and it takes the form

$$G = 2\int_0^1 (p - L(p))dp \tag{2}$$

where: $p = F(y)$ is the cumulative distribution function of income.

Using the definition (2) it can be found that: (see Lerman and Yitzhaki, 1984; Davidson, 2009)

$$G = 2\int_0^1 (p - L(p))dp = \frac{2}{\mu}\int_0^\infty yF(y)dF(y) - 1 \tag{3}$$

According to Gini (1914), the relative inequality (concentration) corresponding to p can be given by:

$$\rho(p) = (p - L(p))/p \tag{4}$$

The concentration measure (2) can be expressed on the basis of the point index (4) as the weighted mean of $\rho(p)$ with the weight p:

$$G = \int_0^1 \frac{p - L(p)}{p} \cdot p\ dp / \int_0^1 p\ dp = 2\int_0^1 (p - L(p)\ dp \tag{5}$$

Unfortunately, the point concentration measures $\rho(p)$ and $2(p - L(p))$ have forced behaviour so their sensitivity to detecting income inequalities depends on p.

The point and synthetic inequality indices proposed by Bonferroni (1930) are $V(p) = \frac{1}{\mu}(\mu - \bar{M}(p))$ and $V = \int_0^1 V(p)\ dp$, respectively, where the expression $\bar{M}(p) = \frac{1}{p}\int_0^p F^{-1}(t)\ dt$ denotes the lower mean. De Vergottini (1940) showed that $\rho(p) = V(p)$ and pointed out that the Gini index G can also be expressed as the weighted mean of $V(p)$ with weight p:

$$G = \int_0^1 \frac{\mu - \bar{M}(p)}{\mu} \cdot p\ dp / \int_0^1 p\ dp \tag{6}$$

An alternative to the Lorenz curve (1), is the concentration curve proposed by Zenga (1984, 1990), defined in terms of quantiles of a size distribution and the corre-

sponding quantiles of the first-moment distribution. This point inequality curve has not a pre-established behaviour being sensitive to changes of inequality in each part (point) of a population.

For a non-negative continuous random variable Y with probability density function $f(y)$, strictly increasing distribution function $F(y)$, and positive finite expectation μ, the first incomplete moment can be defined as $Q(y) = \frac{1}{\mu}\int_0^y tf(t)dt$.

The Zenga point measure of inequality is based on the relation between income and population quantiles

$$Z_p = \frac{y_p^* - y_p}{y_p^*} = 1 - R(p) \tag{7}$$

$$R(p) = \frac{y_p}{y_p^*}$$

where $y_p = F^{-1}(p)$ denotes the population p^{th} quantile and $y_p^* = Q^{-1}(p)$ is the corresponding income quantile. Thus the Zenga approach consists of comparing the abscissas at which $F(y)$ and $Q(y)$ take the same value p.

Zenga synthetic inequality index can be expressed as the area below the Zenga curve (7), and is defined as simple arithmetic mean of point concentration measures $Z_p, p \in < 0, 1 >$:

$$\zeta = \int_0^1 Z_p \, dp \tag{8}$$

It is worth mentioning that the point concentration measure $Z(p)$ has the interesting property of being constant for the lognormal density, so the departure from the straight-line pattern can be utilized as the indication of non-lognormality of a distribution. Zenga (1991) showed that the values of the point concentration function $Z(p)$ and of the synthetic index ζ in the case of a finite population can be easily obtained by using the cograduation table between the variables Y and Y^*, where Y^* takes on the n ordered values with relative weights $q(y_{(i)}) = y_{(i)}/n\bar{y}$. Further interesting results regarding $Z(p)$ and ζ were reported in Zenga (2007). In particular, it has been shown that the partial ordering based on ζ is coherent with the Lorenz ordering (Berti and Rigo, 1995). It has also been noted that the dependence of $Z(p)$ on the inverse functions of $F(y)$ and $Q(y)$ is likely to prevent a wide diffusion of $Z(p)$ and ζ values. Using the concept of cograduation tables, Porro and Zenga (2014) and Arcagni and Zenga (2014) obtained for ζ the decomposition by sources and by subpopulations, respectively. The recent paper by Arnold (2015) has been devoted to the properties of the Bonferroni and Zenga curves, considered in the context of inequality ordering between income distributions.

2. ESTIMATION OF GINI AND ZENGA INEQUALITY MEASURES

As it was mentioned above, numerous formulas have been applied to express the Gini index of inequality. Some of them seem especially convenient to derive Gini index estimators for survey data while some others are rarely used for this purpose.

Suppose that an iid sample of size n has been drawn randomly from a population, and let its empirical distribution function be denoted as $\hat{F}$. The natural plug-in estimator of the Gini index based on (3) can be defined as:

$$\hat{G}_0 = \frac{2}{\bar{y}} \int_0^{\infty} y\hat{F}(y)d\hat{F}(y) - 1 \tag{9}$$

It can be noticed that using (9) different estimates of G can be obtained, depending on how the empirical distribution function is defined (right- or left-continuous). To avoid this ambiguity one would rather consider different Gini index expressions for discrete data.

Let us consider a finite population consisting of n units (households), $y_{(1)} \leq_{...} \leq y_{(i)} \leq ... \leq y_{(n)}$ where $y_{(i)}$ are the ordered values of the variable Y. In this case the Gini index can be given by the formula based on the lower mean $\bar{M}_i$ (see Zenga, 2013)

$$\hat{G}_1 = \frac{2}{n+1} \sum_{i=1}^{n} \frac{\bar{y} - \bar{M}_i(Y)}{\bar{y}} \cdot \frac{i}{n} = \frac{1}{n} \sum_{i=1}^{n} G_i(Y) \tag{10}$$

where $$G_i(Y) = \frac{\bar{y} - \bar{M}_i(Y)}{\bar{y}} \cdot \frac{2i}{n+1} = 2\left(\frac{i}{n} - L\left(\frac{i}{n}\right)\right)\frac{n}{n+1}$$

can be assumed ''Gini point measure of inequality''.

The Gini index is also related to the mean difference with repetitions $\Delta = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$ by the following equation

$$G = \frac{\Delta}{2\bar{y}} \tag{11}$$

It is worth noting that in the formula (11) we preserve neither the connection with the Lorenz curve nor with the point measure. What is more, the introduction of the double sum and absolute values makes this expression computationally inconvenient.

In the case of survey data, the Gini index is most frequently estimated using the following formula based on order statistics: (see Sen,1973; Fei, Ranis and Kuo, 1979)

$$\hat{G}_2 = \frac{n+1}{n} - \frac{2}{n^2\bar{y}} \sum_{i=1}^{n} (n+1-i)y_{(i)} \tag{12}$$

With the formula (12) one can easily incorporate unequal survey weights into the

estimation process by replacing the original sample incomes y_i by their corresponding *expanded values* $y_i w_i$:

$$\hat{G}_3 = \frac{2\sum_{i=1}^{n}\left(w_i y_{(i)} \sum_{j=1}^{i} w_j\right) - \sum_{i=1}^{n} w_i y_{(i)}}{\left(\sum_{i=1}^{n} w_i\right)\sum_{i=1}^{n} w_i y_{(i)}} - 1 \tag{13}$$

where: $y_{(i)}$ – household incomes in a non-descending order, w_i- survey weight for i-th economic unit, $\sum_{j=1}^{i} w_j$- rank of i-th economic unit in n-element sample.

Note that the formula (12) can also be derived from (11) by substituting the numerator of Δ with the relation $S = 2\sum_{i=1}^{n} y_{(i)}(2i - n - 1)$. Thus, we return to the ordered values on which are based both- the lower means $\bar{M}_i$ and the (empirical) Gini point index: $\rho(i/n) = (i/n - L(i/n))/(i/n)$.

To obtain the estimator of the Zenga synthetic inequality index ζ one can consider the transformation of the variates proposed in Zenga (1984). This seminal paper gave birth to the point and synthetic inequality measures (7) and (8) and can also be helpful to study their statistical properties. After the transformation $p = Q(y)$ the formula (7) takes the form (see Zenga, 1984; equations 2.4, 2.5):

$$\zeta = 1 - \int_0^{\infty} \frac{F^{-1}(Q(y))}{y} q(y)\, dy = 1 - \int_0^{\infty} \frac{F^{-1}(Q(y))}{E(Y)} f(y)\, dy \tag{14}$$

where $q(y) = \dfrac{\mathrm{y}}{\mathrm{E(Y)}} f(y)$ is the density of income values (*densità delle quote di reddito*); note that integrating $q(y)$ between 0 and y one can obtain $Q(y)$.

The commonly used nonparametric estimator of the Zenga index (8) introduced by Aly and Hervas (1999) is obviously connected with the right-side of the result (14) and can be expressed by the following equation:

$$\hat{\zeta} = 1 - \frac{1}{n\bar{y}} \left\{ y_{1:n} + \sum_{j=1}^{n-1} y \left\langle \frac{\sum_{i=1}^{j} y_{i:n}}{\bar{y}} \right\rangle_{:n} \right\} \tag{15}$$

where: $y_{i:n} - i^{\text{th}}$ order statistics in n-element sample, $\bar{y}$ – sample arithmetic mean, <x> is the smallest integer $\geq$ x. For large samples it has been proven that $\sqrt{n}(\hat{\zeta} - \zeta) \to N(0, \sigma_{\zeta}^2(F))$, so the estimator is consistent and asymptotically normally distributed.

In order to illustrate the calculations necessary to obtain (15), we will utilize the data reported in Porro and Zenga (2014) for the decomposition by subpopulations of ζ: 5, 11, 14, 20, 2, 11, 23, 24, 42, 48 (see Table 1).

TABLE 1. - *Outline of the calculations for the formula (15)*

j	$y_{(j)}$	$\sum_{i=1}^{j} y_{i:n}$	$(\sum_{i=1}^{j} y_{i:n})/\bar{y}$	$\left\langle \frac{\sum_{i=1}^{j} y_{i:n}}{\bar{y}} \right\rangle :n$	$y\left\langle \frac{\sum_{i=1}^{j} y_{i:n}}{\bar{y}} \right\rangle :n$
1	2	2	0.1	1	2
2	5	7	0.35	1	2
3	11	18	0.9	1	2
4	11	29	1.45	2	5
5	14	43	2.15	3	11
6	20	63	3.15	4	11
7	23	86	4.3	5	14
8	24	110	5.5	6	20
9	42	152	7.6	8	24
10	48	200	10	10	48
	200				91

Finally we obtain $\hat{\zeta} = 1 - (1/200)(2 + 91) = 0,525$, the result that is obviously different from the value 0.3996 obtained in Porro and Zenga (2014) by means of a cograduation table. The reason is that the Aly and Hervas formula is based on $nQ(y)$ rather than $q(y)$. The difference is likely to diminish when the sample size increases.

In the survey sampling context, when the design weights w_i based on inverse inclusion probabilities are given for each sampling unit, we can modify the formula (15) by putting the expanded income values instead of the original ones and obtain

$$\hat{\zeta} = 1 - \frac{1}{\sum_{j=1}^{n} w_j y_j} \left\{ w_1 y_{1:n} + \sum_{j=1}^{n-1} w_j y \left\langle \frac{\sum_{i=1}^{j} w_i y_{i:n}}{\bar{y}} \right\rangle :n \right\} \tag{16}$$

Both the estimators given by the formulas (12) and (15) are consistent and asymptotically normal random variables with large–sample variances derived by Davidson (2009) and Aly and Hervas (1999), respectively. Pollastri (1987), Dancelli (1990) and Latorre (1990) studied the properties of the Z curve and the index ζ and showed that they have several favorable properties relative to other measures. Latorre (1990) studied the statistical properties of parametric estimates of ζ an G under several models of income distributions. Aly and Hervas (1999) conducted Monte Carlo experiments to evaluate the performance of the estimator (15) where they employed

the lognormal and Pareto models of income distributions. In the paper the results of simulation studies based on the Dagum and the lognormal models are presented for both Gini and Zenga parametric estimators. We provide and contrast the results of Monte Carlo experiments concerning unbiasedness, dispersion, and normality of the estimators, the properties that are necessary for reliable statistical inference.

3. RESULTS

A simulation study has been conducted to verify large sample properties of the estimators for Gini and Zenga (1984) inequality coefficients given by the formulas (12) and (15). Both estimators are known to be consistent but it would be interesting to assess the sample sizes necessary to achieve their asymptotic unbiasedness and normality. The dispersion of sampling distributions will also be evaluated for different sample sizes, as it can be useful for making inference about the two indices.

In the experiment two different probability distributions were used as population models:

– two-parameter lognormal distribution,

– three-parameter Dagum distribution.

The Dagum model (known also as the Burr type-III distribution) is a heavy-tailed probability distribution incorporated into income distribution analysis by Dagum (1977) on the basis of the observed patterns of stable regularity of the income elasticity of cumulative distribution function. It has proved sufficient goodness-of-fit in many applications, including wage and income distributions in Poland by different divisions. It is a flexible distribution that can be unimodal or zeromodal, depending on parameters. Thus it can approximate income distributions, which are usually unimodal, and wealth distributions, that are zeromodal (for details see Dagum, 1977; Kleiber and Kotz, 2003).

Contrary to the Dagum model, the lognormal distribution is a light-tailed positively asymmetrical probability distribution which has frequently been applied for at least the last 60 years as an appropriate model of income and wage size distributions. It is still being used for various income data, especially for transition-economies, mainly for its simplicity and economic interpretation of parameters.

The parameters of both theoretical distributions were established on the basis of real income data coming from the Polish HBS and administrative registers, comprising large variety of subpopulations differing in the level of income inequality. As the result, several separate population types have been established for the purpose of the experiment, differing from each other not only in the underlying distribution model but also in the inequality level. The sample sizes were fixed for each variant as $n = 100, 200, 300, 400, 500, 1000, 2000, 3000, 5000, 7000$. The number of repetitions of Monte Carlo experiment was $N = 10\,000$.

The results of the experiments are presented in Tables 2 and 3 and on Figures 1-10. Table 2 summarizes basic statistical characteristics of the empirical distributions of Gini and Zenga inequality coefficients, assuming the lognormal distribution. Be-

sides the Gini index estimator $\hat{G}$ given by (12), a bias-corrected estimator $\tilde{G}$ proposed by Davidson (2010) has been considered. In the Table 3 the corresponding results for the Dagum distribution as a population model have been presented.

The detailed analysis of the outcome of the experiments allows us to formulate several conclusions concerning statistical properties of the estimators:

– the distributions of the Zenga index estimator are slightly more dispersed and more skew than the corresponding distributions of the Gini index. It can also be noticed that the asymmetry and dispersion for both estimators are obviously much higher when the Dagum model is taken into consideration,

– the asymmetry and the dispersion of both estimators increase when the concentration of the underlying distribution model is higher,

– the estimator $\hat{G}$ of the Gini index presents lower absolute bias and lower sampling variance comparing to the Zenga index estimator $\hat{\zeta}$, what is more evident for the Dagum model.

– the bias of $\hat{G}$ is negative, while for the estimator $\hat{\zeta}$ certain overestimation has been observed in most cases, which can be considered more convenient in practice. As the bias of inequality estimators tends to diminish together with increasing distribution inequality, the negative bias of the Gini index is becoming more serious for highly unequal populations (Figure 1; Figure 3).

– the bias of $\hat{G}$, even for the Dagum model with high concentration level and for very small samples, does not exceed 1% of the true parameters. The bias-corrected estimator $\tilde{G}$ behaves well for the Dagum model but has the tendency to slightly overestimate the Gini index, especially when the right tail of the distribution model is not so heavy (Table 2),

– the bias of the estimator $\hat{\zeta}$, as observed for the Dagum model, is positive for smaller concentration of the general population and tends to decrease as the concentration level increases, reaching negative values for high concentration ($\zeta = 0,44$).

TABLE 2. - *Characteristics of empirical distributions of the estimators $\hat{G}$ and $\hat{\zeta}$ under lognormal model*

n	Gini index estimator				Zenga index estimator		
	Expected value		Standard deviation	Coeff. of skewness	Expected value	Standard deviation	Coeff. of skewness
	$E(\hat{G})$	$E(\tilde{G})$	$D(\hat{G})$	$\gamma_1(\hat{G})$	$E(\hat{\zeta})$	$D(\hat{\zeta})$	$\gamma_1(\hat{\zeta})$
Population I **G=0.3286**					**Z=0.3023**		
100	0.3280	0.3313	0.0246	0.1699	0.3099	0.0403	0.3185
500	0.3284	0.3291	0.0110	0.0821	0.3045	0.0183	0.1519
1000	0.3286	0.3289	0.0078	0.0733	0.3038	0.0130	0.1311
2000	0.3286	0.3288	0.0055	0.0802	0.3032	0.0092	0.1158
3000	0.3286	0.3287	0.0045	0.0390	0.3029	0.0075	0.0576
5000	0.3286	0.3287	0.0035	0.0388	0.3027	0.0057	0.0564
Population II **G=0.3512**					**Z= 0.3395**		
100	0.3504	0.3540	0.0264	0.1851	0.3466	0.0445	0.3123
500	0.3509	0.3516	0.0119	0.0878	0.3417	0.0203	0.1541
1000	0.3511	0.3515	0.0084	0.0727	0.3410	0.0144	0.1337
2000	0.3512	0.3513	0.0060	0.0727	0.3404	0.0102	0.1192
3000	0.3512	0.3513	0.0048	0.0383	0.3401	0.0083	0.0569
5000	0.3512	0.3513	0.0037	0.0382	0.3399	0.0064	0.0546
Population III **G=0.4041**					**Z=0.4302**		
100	0.4029	0.4070	0.0308	0.2281	0.4357	0.0537	0.2812
500	0.4040	0.4046	0.0139	0.0877	0.4322	0.0249	0.1580
1000	0.4041	0.4044	0.0099	0.0925	0.4317	0.0178	0.1393
2000	0.4041	0.4043	0.0070	0.0920	0.4312	0.0126	0.1268
3000	0.4041	0.4042	0.0057	0.0394	0.4309	0.0103	0.0575
5000	0.4041	0.4042	0.0044	0.0378	0.4307	0.0079	0.0532

TABLE 3. - *Characteristics of empirical distributions of the estimators* $\hat{G}$ *and* $\hat{\zeta}$ *under Dagum type-I model*

n	Gini index estimator				Zenga index estimator		
	Expected value		Standard deviation	Coeff. of skewness	Expected value	Standard deviation	Coeff. of skewness
	$E(\hat{G})$	$E(\tilde{G})$	$D(\hat{G})$	$\gamma_1(\hat{G})$	$E(\hat{\zeta})$	$D(\hat{\zeta})$	$\gamma_1(\hat{\zeta})$
Population **G=0.3132**					**Z=0.2907**		
100	0.3122	0.3154	0.0283	0.6260	0.2957	0.0476	0.7319
500	0.3127	0.3133	0.0130	0.3363	0.2919	0.0228	0.5084
1000	0.3129	0.3132	0.0092	0.2309	0.2915	0.0163	0.3765
2000	0.3130	0.3132	0.0064	0.1366	0.2911	0.0115	0.2947
3000	0.3131	0.3132	0.0053	0.1561	0.2910	0.0096	0.2739
5000	0.3132	0.3132	0.0041	0.0944	0.2910	0.0074	0.1953
Population I **G= 0.3514**					**Z=0.3540**		
100	0.3493	0.3528	0.0359	1.0154	0.3548	0.0616	0.8497
1000	0.3509	0.3513	0.0121	0.4626	0.3538	0.0227	0.5640
2000	0.3511	0.3513	0.0085	0.3410	0.3538	0.0163	0.4557
3000	0.3512	0.3513	0.0071	0.3411	0.3538	0.0136	0.4629
5000	0.3513	0.3514	0.0055	0.2362	0.3538	0.0107	0.3515
Population II **G= 0.4006**					**Z=0.4410**		
100	0.3970	0.4010	0.0430	1.1764	0.4354	0.0735	0.8697
500	0.3992	0.4002	0.0212	1.0174	0.4379	0.0391	0.8215
1000	0.3998	0.4002	0.0152	0.7418	0.4388	0.0289	0.6927
2000	0.4001	0.4003	0.0109	0.6264	0.4393	0.0211	0.6451
3000	0.4002	0.4004	0.0090	0.6276	0.4395	0.0177	0.6500
5000	0.4004	0.4005	0.0070	0.4553	0.4398	0.0139	0.5178

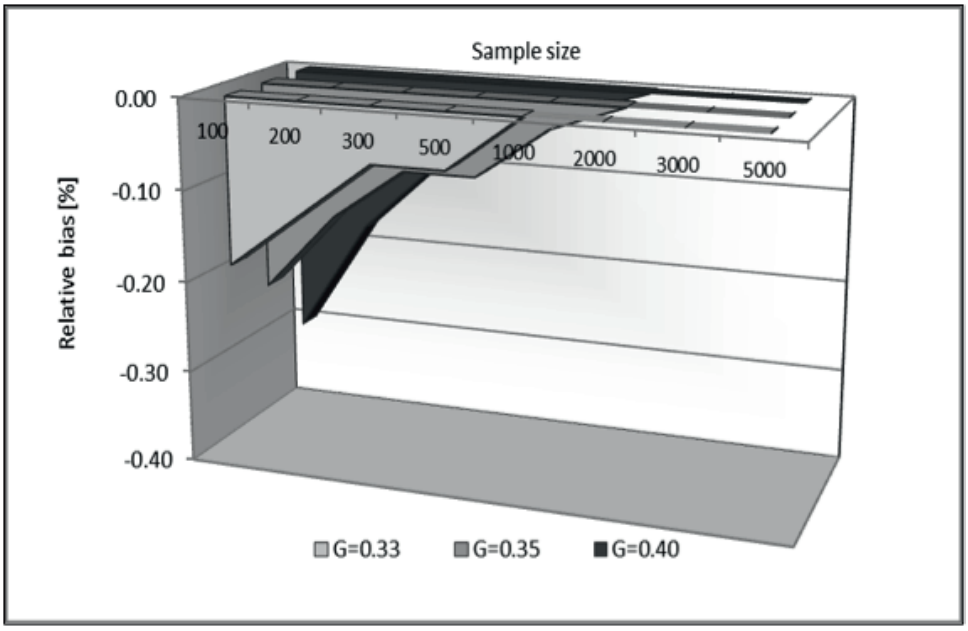

FIGURE 1. - *Relative bias of* $\hat{G}$ *for lognormal distribution*

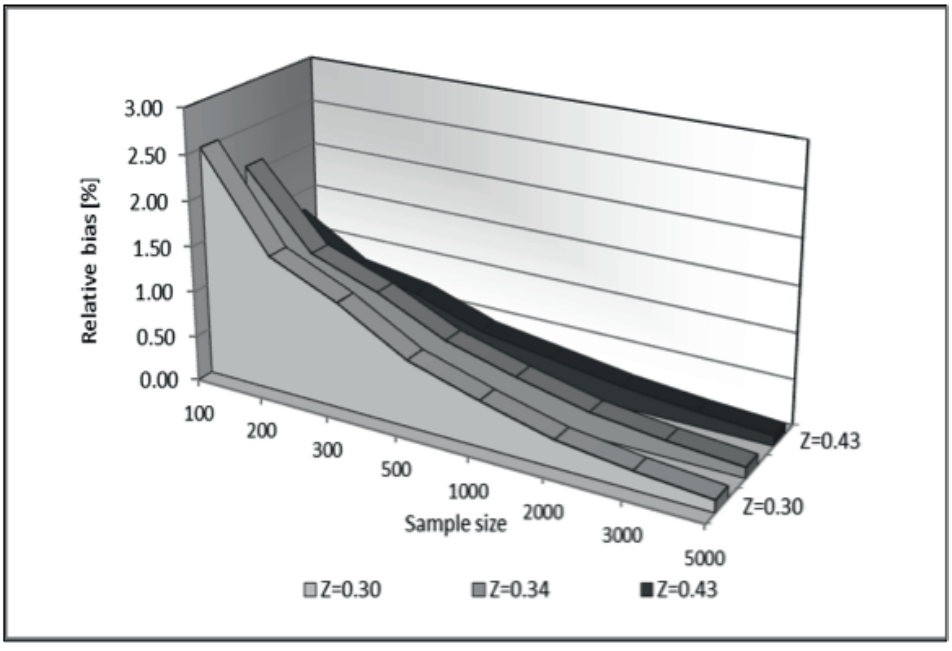

FIGURE 2. - *Relative bias of* $\hat{\zeta}$ *for Dagum distribution*

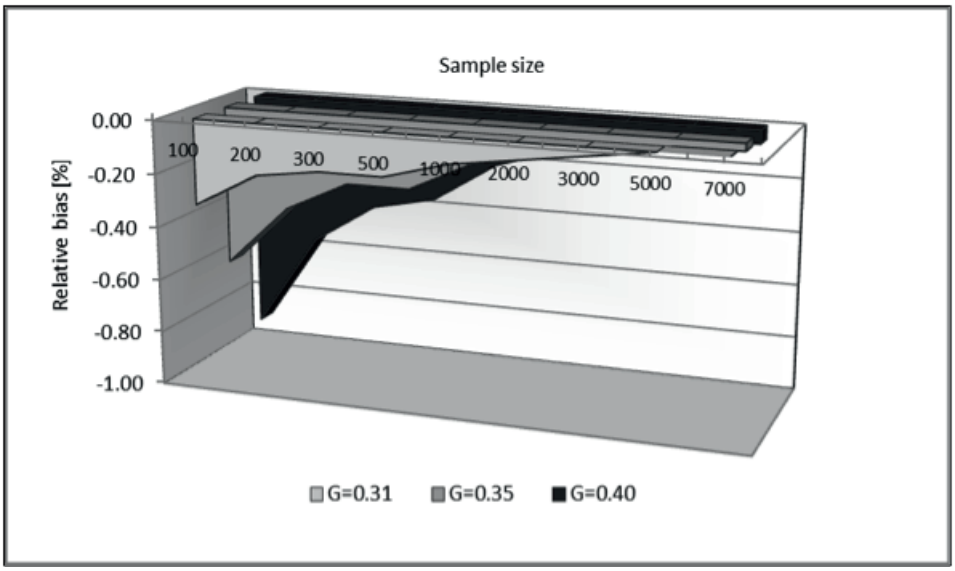

FIGURE 3. - *Relative bias of* $\hat{G}$ *for lognormal distribution*

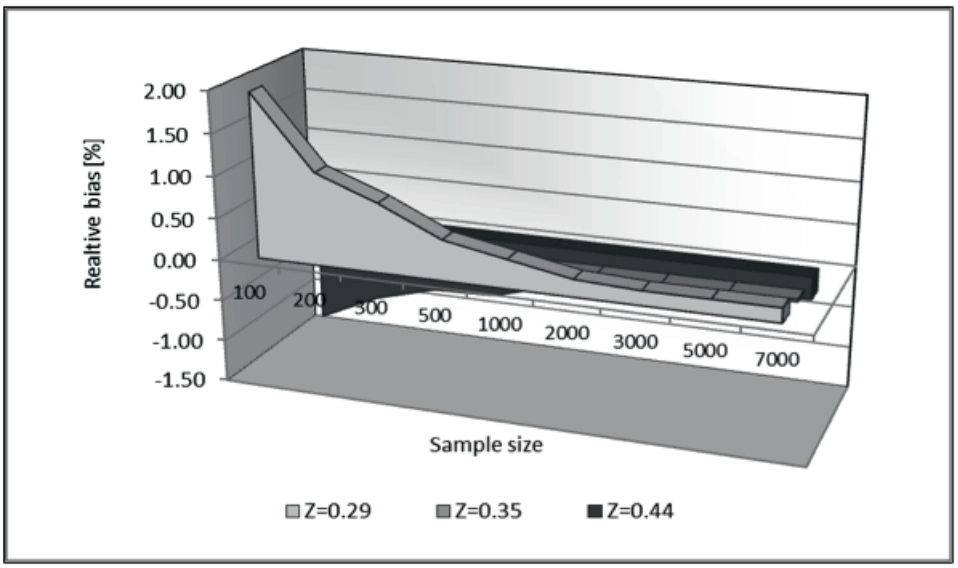

FIGURE 4. - *Relative bias of* $\hat{\zeta}$ *for Dagum distribution*

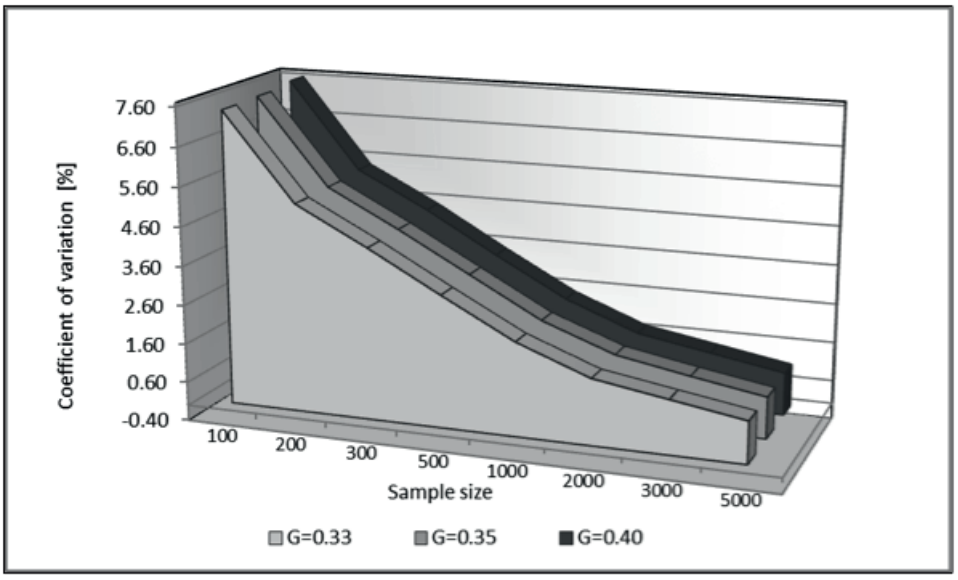

FIGURE 5. - *Coefficient of variation of* $\hat{G}$ *(lognormal model)*

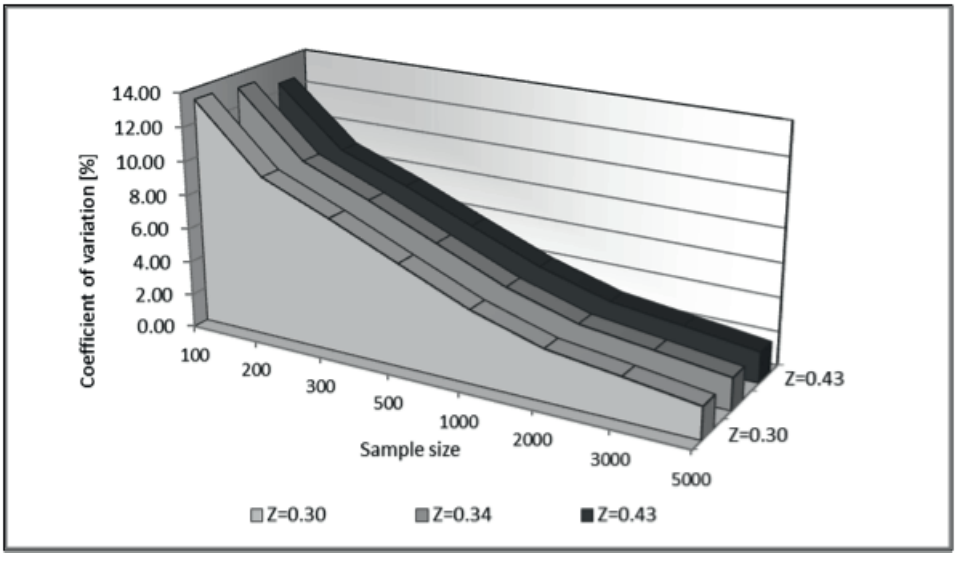

FIGURE 6. - *Coefficient of variation of* $\hat{\zeta}$ *(lognormal model)*

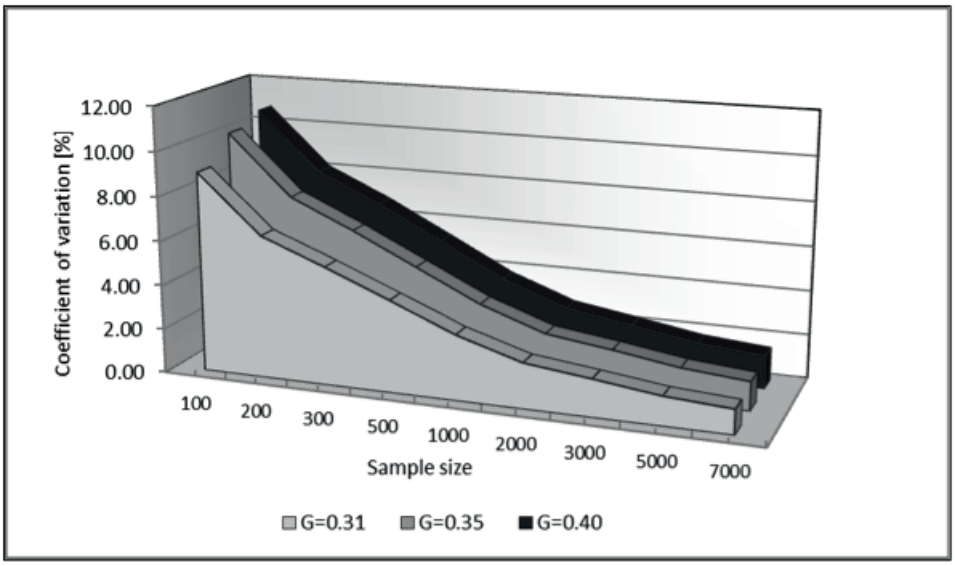

FIGURE 7. - *Coefficient of variation of* $\hat{G}$ *(Dagum model)*

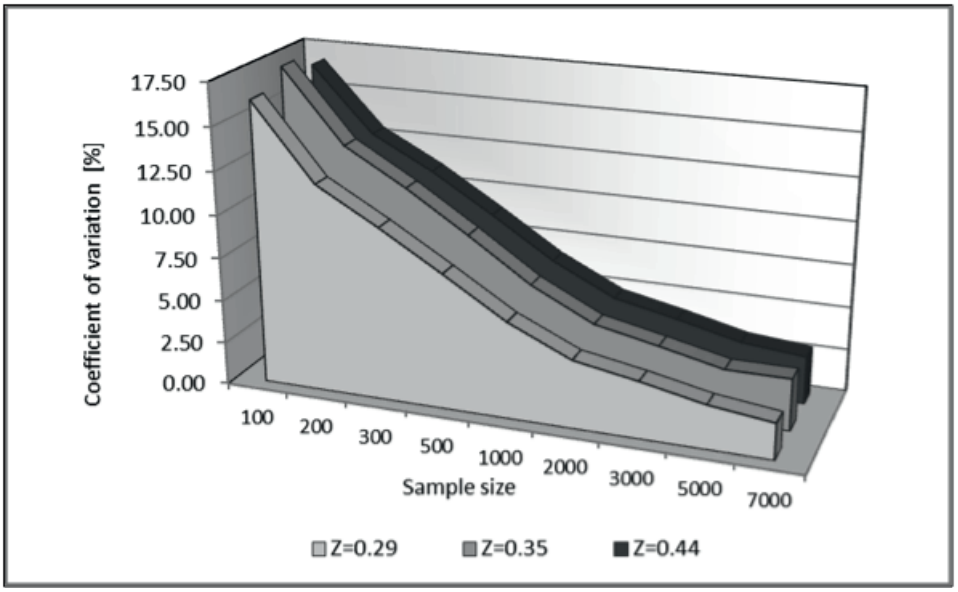

FIGURE 8. - *Coefficient of variation of* $\hat{\zeta}$ *(Dagum model)*

Figures 9 and 10 show the relative frequency histograms for Gini and Zenga estimators obtained on the basis of 10 000 repetitions of the experiment for sample sizes equal to 200. The histograms are accompanied by fitted normal density curves and the values of their corresponding overlap measures (OLM), presenting the goodness-of-fit of the empirical distributions with the assumed theoretical ones. The measure of distributions similarity OLM was first proposed by the Polish economist, demographer and statistician Egon Vielrose in 1960, and represents the ''common part'' of a pair of distributions. In the study it has been applied for the number of intervals k=20.

As it can easily be noticed, the normality of the estimators may not be preserved for smaller samples, especially for the Zenga statistic presenting higher asymmetry (Tables 2 and 3) and for the Dagum distribution as a population model (Table 3). According to the experiments assuming the Dagum distribution as a population model, the goodness-of-fit is satisfactory ($OLM > 0{,}95$) only for sample sizes $n = 2000$ and more. For the lognormal model, the distributions of both estimators tend to the normal one for relatively small samples $(n = 200)$ as it can be seen in the Figures 9 and 10.

The estimators of Gini and Zenga inequality measures which were discussed in the paper have recently been applied to the estimation of income inequality in Poland by family-type (Jedrzejczak and Kubacki, 2013). The basis for the calculations was micro-data coming from the Polish Household Budget Survey (HBS) conducted by the Central Statistical Office of Poland, with an annual sample size exceeding 37 000 of households. The Household Budget Survey plays an important role in the analysis of the living standards of the population in Poland. It is the basic source of information on the revenues, expenditure, quantitative food consumption and other aspects of the living conditions of particular groups of the population. For the purpose of this study, the overall sample of households has been divided into six family types, established on the basis of the number of children (Table 4). The survey is based on the random sampling method which allows for the generalisation of the results to the whole population of households within a margin of an error. The adopted sampling design was a geographically stratified and two-stage one with different selection probability at the first stage. Standard errors of $\hat{G}$ and $\hat{\zeta}$ have been estimated on the basis of replication techniques as well as using the parametric approach proposed by Latorre (1990), assuming the Dagum type-I model. The Dagum model parameters have been estimated by means of the Maximum Likelihood method.

The results presented in the Table 4 confirm the regularities obtained within the Monte Carlo experiments (Tables 2 and 3). For some subpopulations (groups of families with more than 3 children) the effective sample sizes are too small to obtain high accuracy of income inequality estimates and their relative standard errors rSE exceed 10%. To increase the precision some model-based methods using auxiliary information coming from administrative registers can be adopted, as described in (Jedrzejczak and Kubacki, 2013).

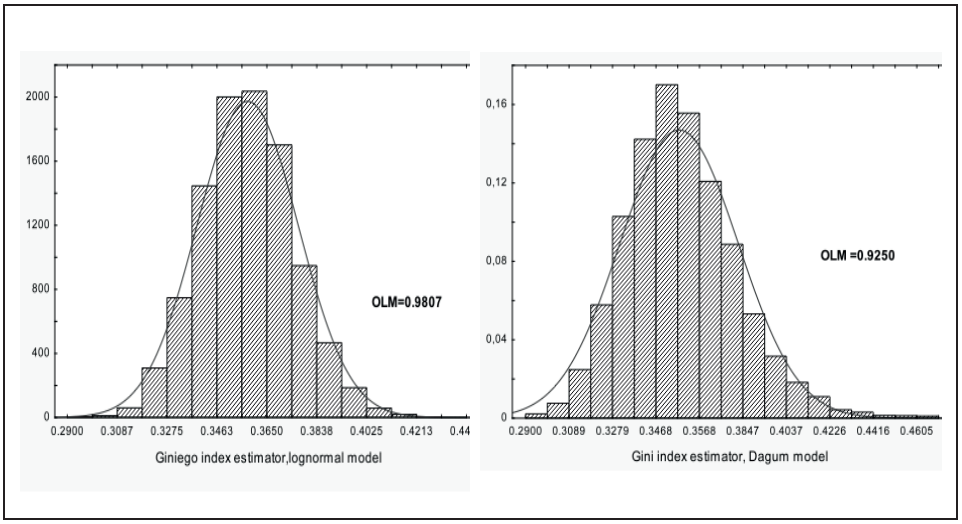

FIGURE 9. - *Empirical distributions of Gini index estimator under Dagum and lognormal models*

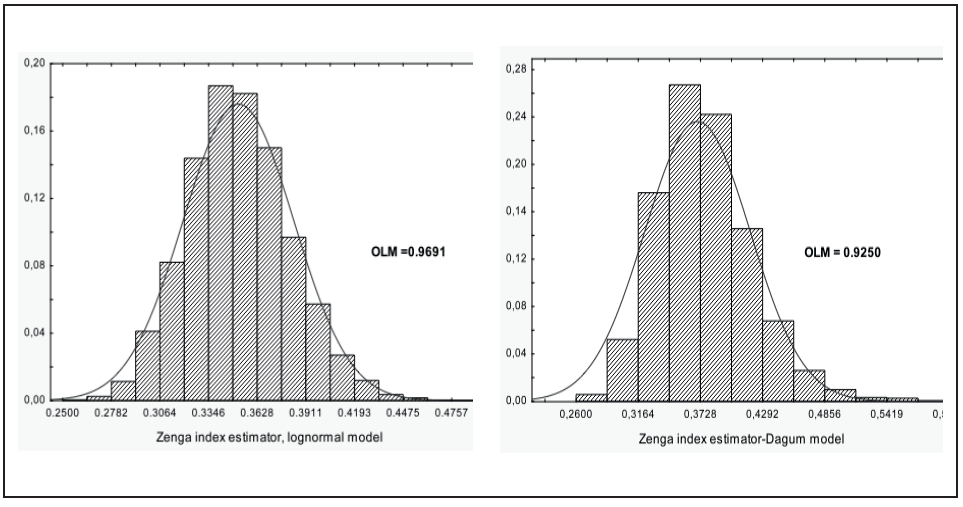

FIGURE 10. - *Empirical distributions of Zenga index estimator under Dagum and lognormal models*

TABLE 4. - *Estimation results for income distributions in Poland based on the micro-data coming from Household Budget Survey*

No of children	n	Dagum model parameters			Central tendency*		Zenga point inequality			Synthetic inequality (rSE)	
		λ	β	δ	Mean	Median	$p = 0.1$	$p = 0.5$	$p = .95$	$\hat{G}$	$\hat{\zeta}$
0	23985	23.80	0.697	3.039	2.89	2.38	0.34	0.35	0.39	0.36 (0.004)	0.37 (0.007)
1	6946	96.17	0.728	3.448	3.81	3.29	0.28	0.24	0.36	0.32 (0.009)	0.30 (0.016)
2	4658	34.33	1.044	3.036	3.93	3.27	0.26	0.27	0.46	0.33 (0.014)	0.31 (0.021)
3	1273	20.45	1.226	3.021	3.58	2.97	0.23	0.26	0.46	0.32 (0.031)	0.30 (0.059)
4	296	40.32	0.989	3.288	3.59	3.06	0.22	0.21	0.29	0.30 (0.057)	0.28 (0.125)
5 ...	144	77.21	0.823	3.549	3.60	3.15	0.25	0.27	0.39	0.29 (0.072)	0.26 (0.137)
Total	37302	37.14	0.703	3.122	3.23	2.70	0.37	0.32	0.41	0.35 (0.003)	0.36 (0.005)

Source: author's calculations on the basis of the Polish HBS, GUS 2009
*Net family income in thous. PLN

4. CONCLUSION

Summing up it is worth to point out that when comparing different inequality measures one should take into considerations the following properties:

- the interpretation of the point and of the synthetic measures.

For example the value $G = 0.35$ understood as the 35% of the value that we have in the case of maximum inequality is not very informative as we refer to the abstractive situation of $G = 1$. It is possible to understand more if we say that, on average, the lower mean is the 65% of the mean of the whole country.

- the behaviour of the point measure (whether it presents a forced behaviour or not),
- the sensitivity of the indices around the set of values that we usually have in the empirical distributions.
- the mean value of an income distribution is very important. Consequently for the estimation of the parameters of a model it seems useful to put the condition that the estimated mean be equal to the empirical (sample mean). This suggest also to use models for which the expectation is equal to one parameter.

- the suitability to the decomposition (for the Zenga index both the source and group decompositions are available when for the Gini index the clear-cut group decompositions can be difficult).
- the inferential procedures including estimation and hypothesis testing.

The Monte Carlo simulations which were carried out within the study confirmed that the estimators of Gini and Zenga (1984) indices are both consistent, and also provided comprehensive information on minimum sample sizes necessary to fulfill estimation quality requirements. The results of the experiments can be useful in many practical applications in the field of income distribution and income inequality, especially in small area statistics where reliable estimates based on small samples are required. It has been revealed for Gini and Zenga estimators that the measure of total sampling error, which can be obtained as the sum of their sampling variances and squared sampling biases (Mean Squared Error), increases when the concentration of the underlying distribution model is higher. Thus the sample sizes for highly unequal subpopulations should be relatively higher in order to ensure appropriate precision.

Assuming the Dagum distribution as an appropriate income distribution model, confidence intervals for inequality measures (especially for subpopulations) should be based on the bootstrap methods rather than the classical approach based on the asymptotic normal distribution. To complete the analysis, similar experiments concerning the properties of relevant variance estimators should be considered. It would also be interesting to broaden the spectrum of populations to cover the distributions of income components.

REFERENCES

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

Arcagni A., Zenga M. (2014). Decomposition by sources of the ζ inequality index. *Proceedings of S.I.S*, Cagliari 2014.

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

Berti P., Rigo P. (1995). A Note on Zenga Concentration Index. *Journal of the Italian Statistical Society*, **4**, 397-404.

Dagum C. (1977). A New Model of Personal Income Distribution. Specification and Estimation. *Economie Appliquée*, **XXX(3)**, 413-436.

Dancelli L. (1990). On the behavior of the Z(p) concentration curve. In C. Dagum and M. Zenga (Eds.), *Income and Wealth Distributions. Income Inequality and Poverty* (pp. 111-127). Springer-Verlag, Berlin.

Davidson R. (2009). Reliable Inference for the Gini Index. *Journal of Econometrics*, **150(1)**, 30-40.

Fei J., Ranis G., Kuo S. (1979). Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics*, **92**, 17-53.

Jedrzejczak A. (2012). Estimation of Concentration Measures and Their Standard Errors for Income Distributions in Poland. *International Advances in Economic Research*, **18(3)**, 287-297.

Jedrzejczak A., Kubacki J. (2013). Estimation of Income Inequality and the Poverty Rate in Poland by Region and Family Type. *Statistics in Transition*, **14(3)**, 259-378.

Kleiber C., Kotz S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, Hoboken New Jersey.

Latorre G. (1990). Asymptotic distributions of indices of concentration: Empirical verification and applications, In C. Dagum and M. Zenga (Eds.), *Income and Wealth Distributions, Income Inequality and Poverty* (pp. 149-169). Springer-Verlag, Berlin.

Lerman R.I., Yitzhaki S. (1984). A Note on the Calculation and Interpretation of the Gini Index. *Economic Letters*, **15**, 363-369.

Pollastri A. (1987). Le curve di concentrazione L_{p }e Z_{p }nella distribuzione log-normale generalizzata. *Giornale degli Economisti ed Annali di Economia*, **46**, 639-663.

Porro F., Zenga M. (2014). The decomposition by subgroups of the inequality curve Z(p) and the inequality index ζ. *Riunione società italiana di statistica*, Cagliari 2014 contribution paper.

Sen A. (1976). Poverty- an Ordinal Approach to Measurement. *Econometrica*, **44**, 219-231.

Vielrose E. (1960). *Rozklad dochodów wedlug wielkosci*. Polskie Wydawnictwo Gospodarcze, Warszawa.

Yitzhaki S., Schechtman E. (2013). *The Gini Methodology*. Springer, New York.

Zenga M. (1984). Proposta per un indice di concentrazione basato sui rapporti fra quantili di popolazione e quantili reddito. *Giornale Degli Economisti ed Annali di Economia*, **48**, 301-326.

Zenga M. (1990). Concentration Curves and Concentration Indices Derived from Them, In C. Dagum and M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty* (pp.94-110). Springer-Verlag, Berlin.

Zenga M. (1991a). L'indice Z(*p*) come misura della concentrazione locale. *Giornale degli Economisti e Annali di economia*, **3-4**, 151-161.