Asymptotic Properties of Some Estimators for Income Inequality Measures- a Simulation Study

Alina Jedrzejczak,
Faculty of Economics and Sociology, University of Lodz, Poland
jedrzej@uni.lodz.pl

Abstract

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. Among many income inequality measures the Gini index based on the Lorenz curve is the most popular. Another interesting measure of income inequality is the Zenga index, 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, efficiency 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, income inequality

1. Introduction

Measures of inequality (called also concentration coefficients) are widely used to study income, welfare and poverty issues. They can also be helpful to analyze the efficiency of a tax policy or to measure the level of social stratification and polarization. They are most frequently applied to dynamic comparisons (comparing inequality across time). Among many inequality measures, the Gini and Zenga coefficients are of greatest importance. The Gini coefficient is the most widely used measure of income inequality, mainly because of its clear economic interpretation. The Zenga „point concentration" measure based on the Zenga curve has recently received some attention in the literature.

The Gini inequality index, based on the Lorenz curve, can be expressed as follows:

$$G = 2\int_0^1 (p - L(p))dp \qquad (1)$$

where: $p = F(y)$ is a cumulative distribution function of income, $L(p)$- the Lorenz function given by the following formula:

$$L(p) = \mu^{-1}\int_0^p F^{-1}(t)dt, \qquad (2)$$

where μ denotes the expected value of a random variable Y and $F^{-1}(p)$ is the distribution p^{th} quantile. Using the definition (1) it can be found that: (see: Davidson, 2010)

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

Suppose that an iid sample of size n is drawn randomly from the population, and let its empirical distribution function be denoted as $\hat{F}$. The natural plug-in estimator of G can be defined as:

$$\hat{G}_0 = \frac{2}{\hat{\mu}}\int_0^\infty y\hat{F}(y)d\hat{F}(y) - 1 \qquad (4)$$

It can be noticed that using (4) different estimates of G can be obtained, depending on how the empirical distribution function is defined (right- or left-continuous). To avoid the ambiguity one can estimate the value of the Gini index from discrete data using the formula based on order statistics: (see: Sen ,1973; Fei, Ranis and Kuo, 1979)

$$\hat{G} = \frac{2\sum_{i=1}^{n} y_{(i)} i - n\bar{y}}{n\sum_{i=1}^{n} y_{(i)}} - 1 \qquad (5)$$

where: $y_{(i)}$ – household incomes in a non- descending order,
i - rank of i-th economic unit in n-element sample.

An alternative to the Lorenz curve (2), is the concentration curve proposed by Zenga (1984, 1990), defined in terms of quantiles of the size distribution and the corresponding quantiles of the first-moment distribution. It is called "*point concentration measure*", being sensitive to changes of inequality in each part (point) of a population.

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

$$Z_p = [y_p^* - y_p]/y_p^*, \qquad (6)$$

where y_p denotes the population p^{th} quantile and y_p^* is the corresponding income quantile defined as follows:

$$y_p^* = Q^{-1}(p). \qquad (7)$$

The function $Q(p)$, called *first-moment distribution function*, can be interpreted as cumulative income share related to the mean income. Thus the Zenga approach consists of comparing the abscissas at which $F(p)$ and $Q(p)$ take the same value p.

Zenga synthetic inequality index can be expressed as the area below the Zenga curve (6), and is defined as simple arithmetic mean of point concentration measures Z_p, $p \in <0,1>$:

$$Z = \int_0^1 Z_p dp \qquad (8)$$

The commonly used nonparametric estimator of the Zenga index (8) was introduced by Aly and Hervas (1999) and can be expressed by the following equation:

$$\hat{Z} = 1 - \frac{1}{n\bar{y}} \left\{ y_{1:n} + \sum_{j=1}^{n-1} y_{\left\langle \frac{\sum_{i=1}^{j} y_{i:n}}{\bar{y}} \right\rangle :n} \right\} \qquad (9)$$

where: $y_{i:n}$ – i-th order statistics in n-element sample,
$\bar{y}$ – sample arithmetic mean.

2. Results

A simulation study has been conducted to investigate large sample properties of the estimators for Gini and Zenga inequality coefficients given by the formulas (5) and (9). In particular, the bias, efficiency and normality of the estimators were considered.

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

- two-parameter lognormal distribution,
- three-parameter Dagum distribution.

The lognormal distribution, being a classical model of income and wage size distributions, has frequently been applied for at least the last 60 years. It is still being used for various income data, especially for transition-economies. The Dagum model (known also as Burr type-III distribution) is, contrary to the lognormal, a heavy-tailed distribution. It has proved sufficient goodness-of-fit in many applications. 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, Kotz, 2003)

The parameters of both theoretical distributions mentioned above were established on the basis of real income data coming from Polish HBS, comprising large variety of subpopulations differing in the level of income inequality. The sample sizes were fixed as n=100, 200, 300, 400, 500, 1000, 2000, 3000, 5000, 7000. The number of repetitions of Monte Carlo experiment was N=10000.

The results of the calculations are presented in tables 1 and 2. Table 1 summarizes basic statistical characteristics of the empirical distributions of Gini and Zenga inequality coefficients, assuming the lognormal distribution. Besides the Gini index estimator given by (5), a bias-corrected estimator $\tilde{G}$ proposed by Davidson (2010) was considered. In the table 2 the results for the Dagum distribution as a population model are presented.

Table 1. Characteristics of empirical distributions of Gini and Zenga inequality measures under lognormal model

Sample size	Gini index estimators				Zenga index estimator		
	Expected value $\hat{G}$	Expected value $\tilde{G}$	Standard deviation	Coeff. of skewness	Expected value	Standard deviation	Coeff. of skewness
Population A		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 B		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 C		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

Source: author’s calculations

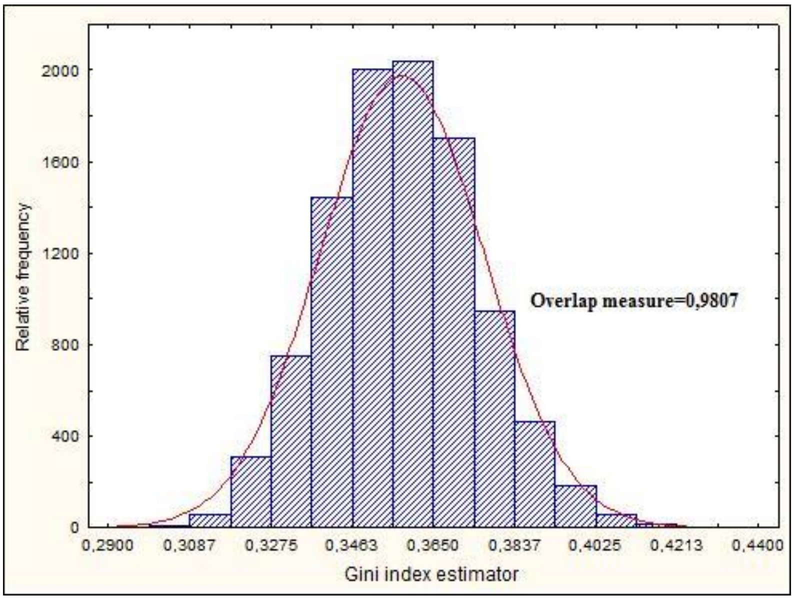
The estimators of the Gini index present generally smaller bias, variance and asymmetry than does the estimator of the Zenga index. The estimator $\hat{G}$ underestimates the true value of the Gini index, while $\hat{Z}$ tends to overestimate the true value of Z. It is worth mentioning that even for small samples, high concentration level and heavy-tailed Dagum density as underlying income distribution model, the relative bias of c is less than 1%. Bias-corrected estimator $\tilde{G}$ truly reduces the bias for the Dagum model, while for the lognormal overestimates the value in many cases.

Table 2. Characteristics of empirical distributions of Gini and Zenga inequality measures under Dagum type-I model

Sample size	Gini index estimators				Zenga index estimator		
	Expected value		Standard deviation	Coeff. of skewness	Expected value	Standard deviation	Coeff. of skewness
	$\hat{G}$	$\tilde{G}$					
Population A :			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 B:			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 C:			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

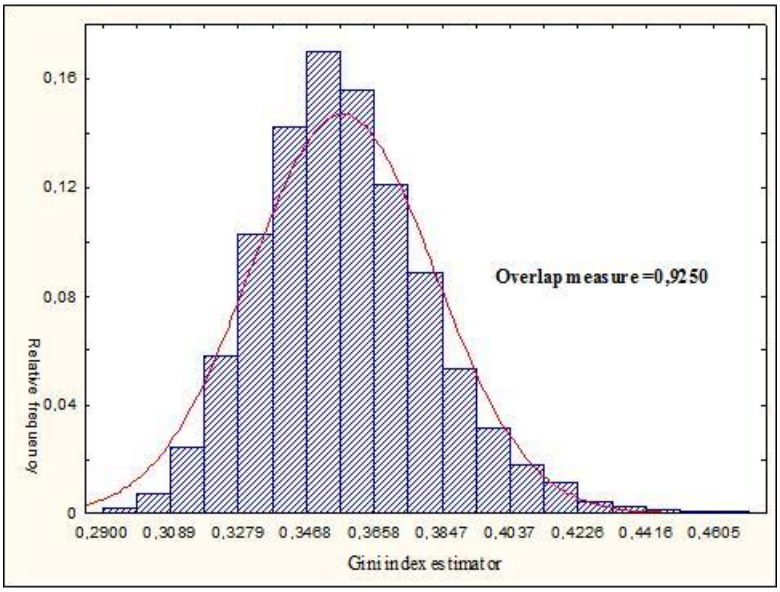
Source: author's calculations

Fig. 1. Empirical distribution of Gini index estimator under lognormal model (n=200)



Source: author's calculations

Fig. 2. Empirical distribution of Gini index estimator under Dagum model (n=200)



Source: author's calculations

For both inequality measures, the consistency with the normal distribution is high (more than 95%), even for relatively small samples, when the underlying income distribution model is lognormal. For the Dagum model and *n*<2000, the discrepancies can be considered significant, especially when the concentration level increases. Figures 1 and 2 show relative frequency histograms obtained on the basis of 10 000 repetitions of the experiment for the estimator $\hat{G}$ and sample sizes equal to 200. The histograms are accompanied by fitted normal density curves and their corresponding overlap measures .

3. Conclusion

The results of the simulation study 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. Assuming the Dagum distribution as an appropriate income distribution model, confidence intervals for inequality measures (especially for subpopulations) should be based on bootstrap methods rather than the classical approach based on 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.

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

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

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

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

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, w: Income and Wealth Distribution, Inequality and Poverty, Springer -Verlag, Berlin, 94–110.