ARGUMENTA
OECONOMICA
CRACOVIENSIA

No 18 • 2018
ISSN 1642-168X
AOC, 2018; 18: 51–67
https://doi.org/10.15678/AOC.2018.1803

Alina Jędrzejczak
Dorota Pekasiewicz

PROPERTIES OF SELECTED INEQUALITY MEASURES BASED ON QUANTILES AND THEIR APPLICATION TO THE ANALYSIS OF INCOME DISTRIBUTION IN POLAND BY MACROREGION

Abstract

Quantiles of income distributions are often applied to the estimation of various inequality, poverty and wealth characteristics. They are traditionally estimated using the classical quantile estimator based on a relevant order statistic. The main objective of the paper is to compare the classical, Huang-Brill and Bernstein estimators for these measures from the point of view of their statistical properties. Several Monte Carlo experiments were conducted to assess biases and mean squared errors of income distribution characteristics for different sample sizes under the lognormal or Dagum type-I models. The results of these experiments are used to estimate inequality, poverty and wealth measures in Poland by macroregion on the basis of micro data originating from the Household Budget Survey 2014.

Keywords: income distribution, inequality, poverty, wealth, quantile estimator.
JEL Classification: C13, C15.

1. Introduction

Statistical measures based on quantiles are frequently applied to the analysis of income distribution as they comprise many popular inequality

Alina Jędrzejczak, University of Łódź, Institute of Statistics and Demography, Department of Statistical Methods, Rewolucji 1905 41/43, 90-214 Łódź, Poland, e-mail: jedrzej@uni.lodz.pl

Dorota Pekasiewicz, University of Łódź, Institute of Statistics and Demography, Department of Statistical Methods, Rewolucji 1905 41/43, 90-214 Łódź, Poland, e-mail: pekasiewicz@uni.lodz.pl

and poverty indices and indicators. Simple dispersion ratios, defined as the ratios of the income of the richest quantile over that of the poorest quantile, usually utilise deciles and quintiles, but in principle, any quantile of income distribution can be used. A version of the decile dispersion ratio using the ratio of the 10th over the 40th percentile which has recently become popular is the so-called Palma ratio. Another popular inequality measure based on deciles is the coefficient of maximum equalisation, also known as the Schutz index or the Pietra ratio. Contrary to the well-known Gini ratio, the quantile-based dispersion ratios are focused on income differences located in the tails of the distribution rather than in the middle groups. They can be used in a supplementary way to overcome the shortcoming of the Gini index, namely, that it is proportionally oversensitive to changes in the middle of the distribution. More sophisticated measures of income inequality have been constructed using differences (or ratios) between population and income quantiles. Probably the first such measure was the Holme's coefficient standardised by Bortkiewicz, which is based on the quantiles of order 0.5. The concentration curve and corresponding synthetic concentration coefficient proposed by Zenga are also defined in terms of quantiles of a size distribution and the corresponding quantiles of the first--moment distribution.

Quantile-based inequality measures are traditionally estimated using the classical quantile estimator based on a relevant order statistic. In many applications these estimates are presented without any information about their precision, which must be a basis for further statistical inference, e.g. statistical hypothesis testing and interval estimation. The problem can be neglected to some extent if we consider the overall population or sample size large enough to apply the asymptotic theory; one should be conscious however, that for 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) these simple methods have been shown to be seriously biased, and the estimation errors were found to be far beyond the values that can be accepted by social policy-makers for making reliable policy decisions (Jędrzejczak 2015).

This paper addresses the problem of statistical properties of the estimators of popular inequality measures based on quantiles. After a brief description of such measures (section 2), selected quantile estimators are introduced (section 3). Section 4 comprises the results of Monte Carlo experiments conducted to assess biases and mean squared errors of quantile estimators and their functions. In the last part of the paper (section 5) we present the application of quantile-based inequality, poverty and wealth indices to Polish Household Budget Survey (HBS) data divided by macroregions.

2. Selected Statistical Inequality Measures Based on Quantiles

Distribution quantiles of a random variable *X*, which is identified with a household or personal income, or the estimators of these quantiles, have been applied in the construction of simple inequality indices such as the quintile dispersion ratio and decile dispersion ratio (for details, see Panek 2011).

The quintile dispersion ratio has the following form:

$$W^{(1)}_{20:20} = \frac{Q_{0.8}}{Q_{0.2}}, \tag{1}$$

where $Q_{0.8}, Q_{0.2}$ are quintiles, respectively, the fourth and the first.

The quintile dispersion ratio can also be defined as the ratio of the sum of incomes of the richest 20% of the population to the sum of incomes of the poorest 20%:

$$W^{(2)}_{20:20} = \frac{\sum_{i \in GK_5} x_i}{\sum_{i \in GK_1} x_i}, \tag{2}$$

where GK_j is *j*-th quintile group.

The measure (2) can be interpreted as the ratio of the average income of the richest 20% of the population to the average income of the poorest 20% of the population and is usually calculated on the basis of equivalised income.

Similar ratios can also be calculated for other quantiles, for instance deciles or percentiles (95th and 5th) of income distributions. Using the first and ninth decile we can obtain the following decile dispersion ratio:

$$W^{(1)}_{10:10} = \frac{Q_{0.9}}{Q_{0.1}}, \tag{3}$$

where $Q_{0.9}, Q_{0.1}$ are deciles, respectively, the ninth and the first:
and

$$W^{(2)}_{10:10} = \frac{\sum_{i \in GD_{10}} x_i}{\sum_{i \in GD_{10}} x_i}, \tag{4}$$

where GD_j is *j*-th decile group.

The reciprocal of the decile dispersion ratio defined by (4) takes values from the interval (0,1) and is called the dispersion index for the end portions of the distribution:

$$K_{1:10} = \frac{\sum_{i \in GD_1} x_i}{\sum_{i \in GD_{10}} x_i} = \frac{1}{W_{10:10}}. \tag{5}$$

If the index $K_{1:10}$ is closer to 1, the inequality is lower (mean incomes in the extremal decile groups are the same).

A popular inequality measure based on income shares received by subsequent decile groups is the coefficient of maximum equalisation, also known as the Schutz index or the Pietra ratio:

$$E = \sum_{j \in I} 100\left(S_j - \frac{1}{10}\right), \tag{6}$$

for $S_j > 0.1$ and $S_j = \frac{\sum_{i \in GD_j} x_i}{\sum_{i=1}^{n} x_i}$, where S_j is income share of the j-th decile group in the total income.

The measure (6) can be interpreted as the portion of the total income that would have to be redistributed (taken from the richer half of the population and given to the poorer half) for there to be income equality.

During a thorough income distribution analysis the problem of inequality measurement is usually interrelated with the estimation of poverty indices. To obtain reliable poverty characteristics it becomes crucial to define and estimate the poverty threshold z_u. There are numerous definitions of this threshold, taking into consideration an absolute or relative approach. The relative poverty line utilised by Eurostat is $z_u = 0.6M_{0.5}$, where $M_{0.5}$ is the median of a random variable X.

On the basis of the poverty line, the popular head-count ratio (at-risk-of--poverty rate) can be determined: $W_{zg.ub.} = F(z_u)$, where F is the distribution function of X.

The poverty threshold and head-count ratio can be estimated using the following estimators:

$$\hat{z}_u = 0.6Me \tag{7}$$

and

$$\hat{W}_{zg.ub.} = \frac{\#\{X_i \le 0.6Me\}}{n}, \tag{8}$$

where $X_1,\ X_2,...,X_n$ is a random sample and *Me* is the median estimator established on the basis of the random sample.

Wealth indices, concentrated on the upper part of income distribution, are utilised to measure the share of the best-off in a population of households. Among others, a wealth line can be defined as $z_b = 3M_{0.5}$ (Brzeziński 2014, Peichl, Schaefer & Scheicher 2008) and the wealth index based on it is given by: $W_b = 1 - F(z_b)$. These measures can be estimated using the following formulas:

$$\hat{z}_b = 3M_{0.5}, \tag{9}$$

and

$$\hat{W}_b = \frac{\#\{X_i > 3Me\}}{n}, \tag{10}$$

where $X_1,\ X_2,...,X_n$ is a random sample and *Me* is the median estimator.

Examples of more sophisticated inequality measures, focused on each and every part of an income distribution, are the Gini and Zenga indices. The popular Gini index based on the Lorenz curve is not considered in this paper. The synthetic Zenga index is based on the concentration curve that can be considered a point concentration measure, as it is sensitive to changes at every “point” of the income distribution. The Zenga point measure of inequality is based on the relation between income and population quantiles (Zenga 1990, Jędrzejczak 2012, Greselin, Pasquazzi & Zitikis 2013, Arcagni 2016):

$$Z_p = \frac{x_p^* - x_p}{x_p^*} = 1 - \frac{x_p}{x_p^*}, \tag{11}$$

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

The Zenga synthetic inequality index is defined as simple arithmetic mean of point concentration measures $Z_p,\ p \in \langle 0,\ 1 \rangle$.

3. Quantile Estimators and Their Properties

Let X be a continuous random variable with distribution function F and let $Q_p = F^{-1}(p)$ be the p-quantile of the random variable X, where $p \in (0,\ 1)$.

If F is a continuous and strictly increasing distribution function, the pth quantile always exists and is uniquely determined.

The well-known estimator of the quantile Q_p is the statistic:

$$\hat{Q}_p = F_n^{-1}(p) = \inf\{x: F_n(x) \geq p\}, \tag{12}$$

where $F_n(x)$ is the empirical distribution obtaining on the basis of a n-element random sample $X_1, X_2, ..., X_n$.

The problem of quantile estimation has a very long history. In the subject literature numerous nonparametric (distribution-free) quantile estimators have been presented. Their particular expressions depend on the underlying empirical distribution function definition.

The classical quantile estimator obtained for the distribution $F_n(x) = \dfrac{card\{1 \leq j \leq n: \; x_i \leq x\}}{n}$ for $x \in R$ is defined by the following formula:

$$\hat{Q}_p = \begin{cases} X^n_{(np)} & \text{for } np \in N, \\ X^{(n)}_{([np]+1)}, & \text{for } np \notin N, \end{cases} \tag{13}$$

where $X^{(n)}_{(k)}$ is an order statistic of rank k.

Among other estimators of quantiles, Q_p we can mention the standard estimator, Huang-Brill estimator, Harrell-Davis estimator and Bernstein estimator, to name just a few (Huang & Brill 1999, Harrell & Davis 1982).

By means of the empirical distribution *level crossing*, which has the following form:

$$F_n(x) = \sum_{i=1}^{n} w_{n,i} I_{(-\infty, x\rangle}(x^{(n)}_{(i)}), \tag{14}$$

where $w_{n,i} = \begin{cases} \dfrac{1}{2}\left[1 - \dfrac{n-2}{\sqrt{n(n-1)}}\right] & \text{for } i = 1, \; n, \\ \dfrac{1}{\sqrt{n(n-1)}} & \text{for } i = 2, \; 3, \; ..., n-1, \end{cases}$

we obtain the Huang-Brill estimator of the *pth* quantile Q_p:

$$\hat{Q}_p^{HB} = X^{(n)}_{(q)}, \tag{15}$$

where

$$q = \left[\sqrt{n(n-1)}\left(p - \frac{1}{2}\left[1 - \frac{n-2}{\sqrt{n(n-1)}}\right]\right)\right] + 2. \tag{16}$$

It can easily be noticed that for $p = 0.5$ the estimator of the quantile $Q_{0.5}$ is the order statistic $X^{(n)}_{\left(\left[\frac{n}{2}\right]+1\right)}$.

Another interesting quantile estimator is the Bernstein estimator given by:

$$\hat{Q}_p^{Brs} = \sum_{i=1}^{n} \left[\binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \right] X^{(n)}_{(i)}. \tag{17}$$

More examples of quantile estimators can be found in the papers of Pekasiewicz (2015) and Zieliński (2006).

4. Analysis of Monte Carlo Experiments

The main objective of the Monto Carlo experiments conducted in the study was to assess the properties of selected estimators of quantiles. We were especially interested in their biases and sampling variances, i.e. the components of their sampling errors. The following estimators have been taken into consideration: the classical quantile estimator (13), Huang-Brill estimator (15) and Bernstein estimator (17). The estimators presenting the best performance were further applied to evaluate the quantile-based inequality measures for income distributions in Poland by macroregion.

In the experiments two different probability distributions were utilised as population models: lognormal distribution, $LG(\mu, \sigma)$, defined by the following density function $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$, $x > 0$ and Dagum distribution $D(\delta, a, b)$, known also as the Burr type-III distribution, with the density function of the form (Kleiber & Kotz 2003) $f(x) = ab^{-a\delta}\delta x^{a\delta-1}\left(1+\left(\frac{x}{b}\right)^a\right)^{-\delta-1}, x > 0$.

The sets of parameters of both theoretical distributions were established on the basis of real income data originating from the Polish HBS and administrative registers, comprising a large variety of subpopulations differing in the level of income inequality, which have been observed over the last two decades. The sample sizes were fixed for each variant as $n = 500$, $n = 1000$, $n = 2000$. The number of repetitions of the Monte Carlo experiment was $N = 20{,}000$. The simulated sample spaces were used to assess, for each estimator, its empirical bias and standard error.

Tables 1 and 2 present the results of the calculations for three quantile estimators: classical, Huang-Brill and Bernstein, for sample sizes 500 and 1000.

Table 1. Properties of Selected Quantile Estimators for Sample Size $n = 500$

Distribution	p	$\hat{Q}_p$		$\hat{Q}_p^{HB}$		$\hat{Q}_p^{Brs}$	
		BIAS	*RMSE*	*BIAS*	*RMSE*	*BIAS*	*RMSE*
LG(8, 0.6)	0.1	–0.215	4.587	–0.489	4.639	0.283	4.419
	0.2	–0.122	3.839	–0.333	3.833	0.194	3.739
	0.3	–0.100	3.535	–0.285	3.550	0.153	3.445
	0.7	0.230	3.574	–0.118	3.548	0.103	3.471
	0.8	–0.161	3.824	–0.158	3.856	0.095	3.728
	0.9	–0.319	4.582	–0.306	4.600	0.077	4.427
LG(8.3, 0.8)	0.1	–0.270	6.095	0.768	6.276	0.448	5.883.
	0.2	–0.150	5.071	0.450	5.140	0.297	4.928
	0.3	–0.089	4.715	0.382	4.756	0.271	4.614
	0.7	0.314	4.754	–0.151	4.703	0.176	4.619
	0.8	–0.158	5.070	–0.195	5.113	0.225	4.955
	0.9	–0.316	6.077	–0.329	6.120	0.259	5.900
D(0.7, 3.6, 3800)	0.1	–0.280	5.558	0.564	5.534	0.279	5.332
	0.2	–0.174	3.957	0.341	3.969	0.105	3.841
	0.3	–0.133	3.298	0.177	3.298	0.073	3.216
	0.7	0.167	2.927	–0.104	2.924	0.065	2.846
	0.8	–0.127	3.247	–0.102	3.234	0.097	3.165
	0.9	–0.203	4.240	–0.212	4.254	0.196	4.128
D(0.7, 2.8, 3800)	0.1	–0.315	7.041	–0.737	7.174	0.433	6.782
	0.2	–0.181	5.065	0.437	5.146	0.213	4.918
	0.3	–0.092	4.228	0.272	4.283	0.184	4.133
	0.7	0.241	3.766	–0.124	3.748	0.118	3.662
	0.8	–0.127	4.159	–0.187	4.138	–0.186	4.061
	0.9	–0.342	5.428	–0.279	5.482	–0.218	5.274

Source: authors' own calculations in Mathematica.

In particular, the tables show the relative biases and relative root mean squared errors of these estimators obtained for predefined population models – lognormal and Dagum – differing across the experiments in the overall inequality level. Similar experiments for the Gini and Zenga ratios were reported in Jędrzejczak (2015).

Analysing the results of the calculations it becomes obvious that the Bernstein estimator performs better than its competitors – its root mean squared errors (*RMSE*) are much smaller than those observed for the other quantile estimators and its relative biases (*BIAS*) are also smaller, especially when the quantiles of higher orders are taken into account.

Table 2. Properties of Selected Quantile Estimators for Sample Sizes $n = 1000$

Distribution	p	$\hat{Q}_p$		$\hat{Q}_p^{HB}$		$\hat{Q}_p^{Brs}$	
		BIAS	*RMSE*	*BIAS*	*RMSE*	*BIAS*	*RMSE*
LG(8, 0.6)	0.1	–0.087	3.240	0.254	3.248	0.132	3.165
	0.2	–0.079	2.718	0.139	2.726	0.108	2.669
	0.3	–0.039	2.504	0.133	2.511	0.095	2.481
	0.7	0.089	2.528	–0.082	2.521	0.042	2.469
	0.8	–0.077	2.712	–0.077	2.712	0.047	2.680
	0.9	–0.131	3.245	–0.131	3.245	0.041	3.169
LG(8.3, 0.8)	0.1	–0.097	4.350	0.359	4.373	0.302	4.220
	0.2	–0.088	3.581	0.195	3.592	0.177	3.571
	0.3	–0.057	3.336	0.176	3.346	0.134	3.271
	0.7	0.169	3.338	–0.061	3.324	0.108	3.280
	0.8	–0.099	3.620	–0.099	3.620	0.070	3.510
	0.9	–0.116	4.339	–0.116	4.339	0.089	4.208
D(0.7, 3.6, 3800)	0.1	–0.182	3.923	0.313	3.916	0.086	3.803
	0.2	–0.068	2.800	0.141	2.776	0.069	2.741
	0.3	–0.105	2.349	0.114	2.346	0.000	2.303
	0.7	0.010	2.054	–0.080	2.049	0.043	2.013
	0.8	–0.085	2.298	–0.078	2.287	0.032	2.256
	0.9	–0.083	2.984	0.116	2.991	0.121	2.915
D(0.7, 2.8, 3800)	0.1	–0.156	5.073	0.368	5.069	0.221	4.493
	0.2	–0.112	3.580	0.232	3.589	0.082	3.509
	0.3	–0.080	3.015	0.144	2.991	0.062	2.958
	0.7	0.137	2.652	–0.063	2.681	0.073	2.599
	0.8	–0.084	2.956	–0.077	2.935	0.069	2.900
	0.9	–0.133	3.846	–0.112	3.848	0.147	3.774

Source: authors' own calculations in Mathematica.

The bias and *RMSE* of the Huang-Brill estimator are similar to the respective values for the classical quantile estimator. It is worth noting that for all cases biases are rather negligible, so the total errors are dominated by sampling variances. In general, the estimation errors are higher for extremal quantile orders, for the heavy-tailed Dagum model, and they also tend to increase as income inequality increases.

Table 3. Properties of the Quintile Dispersion Ratio Based on Quantile Estimators

Distribution	n	Quintile Dispersion Ratio					
		$\hat{W}^{(1)}_{20:20}$ (stand.)		$\hat{W}^{(1)}_{20:20}$ (Huang-Brill)		$\hat{W}^{(1)}_{20:20}$ (Bernstein)	
		BIAS	*RMSE*	*BIAS*	*RMSE*	*BIAS*	*RMSE*
LG(8.0, 0.6)	500	0.113	4.724	–0.387	4.725	0.054	4.532
	1000	0.070	3.335	–0.194	3.329	0.038	3.239
LG(8.1, 0.7)	500	0.063	5.484	–0.388	5.453	–0.014	5.265
	1000	0.011	3.868	–0.157	3.859	–0.017	3.756
LG(8.3, 0.8)	500	0.134	6.252	–0.549	6.189	0.042	6.004
	1000	0.068	4.444	–0.231	4.473	0.027	4.308
D(0.7, 3.6, 3800)	500	0.156	4.454	–0.262	4.442	0.091	4.275
	1000	0.072	3.154	–0.137	3.120	0.046	3.068
D(0.8, 3.0, 3200)	500	0.174	4.978	–0.323	4.997	0.125	4.794
	1000	0.073	3.521	–0.155	3.492	0.053	3.417
D(0.7, 2.8, 3800)	500	0.235	5.753	–0.287	5.691	0.136	5.505
	1000	0.080	4.044	0.213	4.040	0.046	3.930

Source: authors' own calculations in Mathematica.

Table 4. Properties of the Decile Dispersion Ratio Based on Quantile Estimators

Distribution	n	Decile Dispersion Ratio					
		$\hat{W}^{(1)}_{10:10}$ (stand.)		$\hat{W}^{(1)}_{10:10}$ (Huang-Brill)		$\hat{W}^{(1)}_{10:10}$ (Bernstein)	
		BIAS	*RMSE*	*BIAS*	*RMSE*	*BIAS*	*RMSE*
LG(8.0, 0.6)	500	0.126	6.174	–0.631	6.096	0.021	5.882
	1000	0.065	4.327	–0.324	4.304	0.017	4.191
LG(8.1, 0.7)	500	0.124	7.197	–0.630	7.104	0.013	6.868
	1000	0.084	5.088	–0.273	5.028	0.019	4.926
LG(8.3, 0.8)	500	0.186	8.134	–0.773	8.124	0.029	7.758
	1000	0.124	5.815	–0.352	5.766	0.037	5.615
D(0.7, 3.6, 3800)	500	0.353	6.671	–0.439	6.589	0.181	6.344
	1000	0.162	4.702	–0.211	4.651	0.082	4.543
D(0.8, 3.0, 3200)	500	0.347	7.354	–0.493	7.402	0.234	7.002
	1000	0.097	5.179	–0.266	5.193	0.039	5.009
D(0.7, 2.8, 3800)	500	0.554	8.598	–0.551	8.470	0.283	8.181
	1000	0.181	6.003	–0.298	5.948	0.066	5.800

Source: authors' own calculations in Mathematica.

The next step of the experiment was to study basic statistical properties of the estimators of income inequality measures: $W^{(1)}_{10:10}$ and $W^{(1)}_{20:20}$ given by the formulas (1) and (3). These estimators can be obtained as functions of the subsequent quantile estimators mentioned above. The properties of quintile and decile dispersion ratios are illustrated in Tables 3 and 4. All the values are presented as percentages relative to their corresponding population parameters.

Analysing the results of the calculations presented in Tables 3 and 4 it becomes obvious that the estimators of quintile and decile dispersion ratios based on the Bernstein quantile estimator outperform the estimators based on the classical and Huang-Brill estimators of quantiles. For the Bernstein estimator, the biases and mean squared errors turned out to be substantially smaller for most cases.

5. Application of Inequality Measures to the Analysis of Income Distribution in Poland

The inequality measures based on deciles and quintiles, as well as the Zenga indices, have been applied to income inequality analysis in Poland by macroregion (NUTS1), based on the HBS 2014 sample. They include the decile and quintile dispersion ratios, the reciprocal of the decile dispersion ratio *K*, the coefficient of maximum equalisation *E* and the synthetic Zenga index Z. To obtain reliable estimates of these coefficients we used the Bernstein quantile estimator, which turned out to have the highest precision (Tables 1 and 2).

Table 5. Numerical Characteristics of Available Income in Macroregions

Macroregion	Number of households	Minimum	Maximum	Average	Standard deviation
Central	8046	11.00	155017.49	4240.21	3790.53
Southern	7433	12.50	37152.00	3634.03	2179.59
Eastern	6246	10.00	84032.90	3461.45	2876.23
North-Western	5658	3.00	43493.45	3772.15	2611.00
South-Western	3971	1.67	37200.00	3591.07	2337.83
Northern	5575	9.00	126739.54	3646.44	3225.72
Poland	36929	1.67	155017.49	3755.33	2959.95

Source: authors' calculations based on the HBS 2014 sample.

Basic characteristics of the HBS sample, divided by macroregion, are presented in Table 5. Table 6 shows the results of the approximation of the empirical income distributions by means of the Dagum model using the maximum likelihood method. Additionally, in Figure 1 there are histograms and fitted Dagum density curves describing income distributions in Poland by macroregion.

Table 6. Approximation of Income Distributions in NUTS1 by Means of the Dagum Model

Macroregion	Dagum distribution parameters			Overlap measure
	δ	a	b	
Central	0.790	2.8044	3839.630	0.982
Southern	0.669	3.618	3800.167	0.970
Eastern	0.756	3.051	3286.467	0.971
North-Western	0.743	3.233	3687.076	0.964
South-Western	0.722	3.301	3587.800	0.970
Northern	0.718	3.158	3544.934	0.979
Poland	0.747	3.125	3611.017	0.975

Source: authors' calculations based on the HBS 2014 sample.

Analysing the outcomes of the approximation presented in Figure 1 one can observe very high consistency between the empirical distributions and the theoretical ones. This is also confirmed by the values of a goodness-of-fit measure (the overlap coefficient) calculated for each region and the whole country and presented in the last column of the Table 6.

Table 7. Estimated Inequality Measures for Macroregions

Macroregion	$W^{(1)}_{20:20}$	$W^{(2)}_{20:20}$	$W^{(1)}_{10:10}$	$W^{(2)}_{10:10}$	$K_{1:10}$	E	Zenga
Central	3.049	6.939	5.494	12.085	0.083	26.491	0.386
Southern	2.595	4.962	4.283	7.577	0.132	21.667	0.269
Eastern	2.904	6.147	4.927	9.908	0.101	24.740	0.348
North-Western	2.750	5.577	4.742	8.614	0.116	23.221	0.308
South-Western	2.789	5.375	4.536	8.172	0.122	23.017	0.295
Northern	2.828	6.039	4.814	9.841	0.102	24.412	0.347
Poland	2.819	5.916	4.843	9.526	0.105	22.000	0.338

Source: authors' calculations based on the HBS 2014 sample.

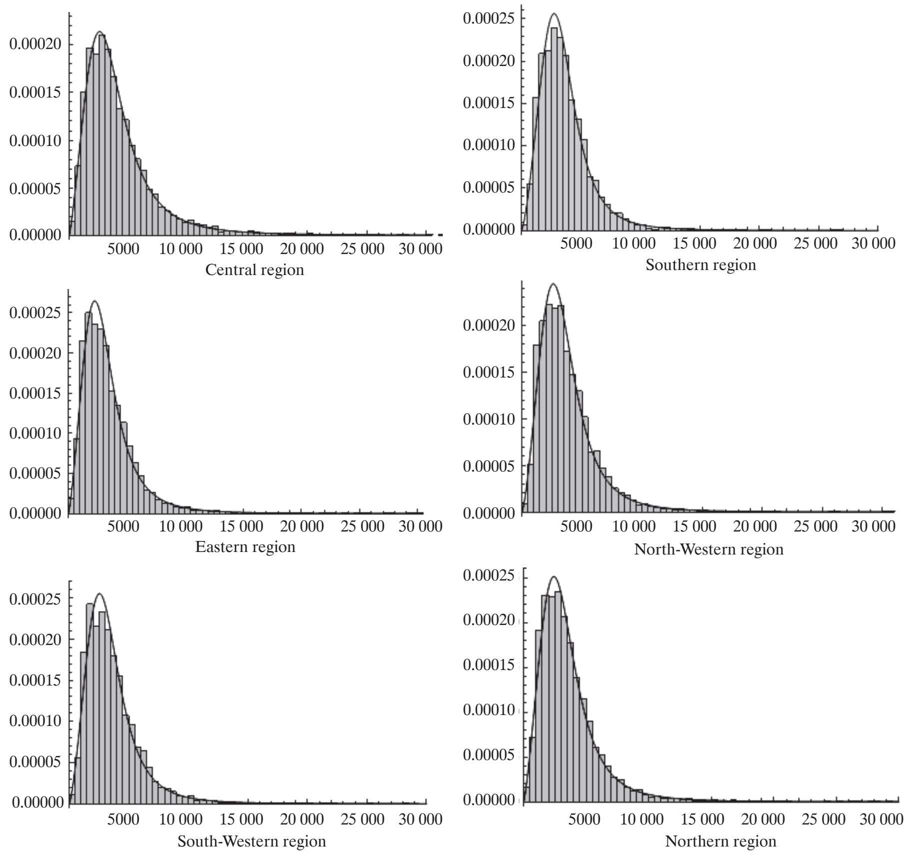

Fig. 1. Income Distributions for NUTS1 and Fitting by Means of the Dagum Model

Source: authors' elaborations in R.

The estimated values of inequality measures such as the decile and quintile dispersion ratios, the reciprocal of the decile dispersion ratio *K* and the synthetic Zenga index *Z*, obtained on the basis of implementation of the Bernstein estimator, are given in Table 7. The indexed values of selected inequality measures from Table 7 have been used to order Polish macroregions by inequality level, as is demonstrated in Figure 2. They also show the differentiation of income inequality across regions.

The estimated values of quintile and decile share ratios, as well as the values of synthetic Zenga inequality measures, indicate the Central macroregion as the one with the highest income inequality level. This is particularly evident for extremal income groups, e.g. the income of the richest 10% of households is 12 times bigger than the income of the poorest 10% ($W^{(2)}_{10:10}$ = 12.085). On the other hand, the lowest values of all inequality measures (except for the K index) have been observed for the Southern macroregion. Three macroregions: Central, Eastern and Northern present income inequality above the national level, while in the remaining three: North-Western, South-Western and Southern it was found to be substantially lower than for the whole country (Figure 2). In general, 22% of the total income of Polish households would have to be redistributed from the richer to the poorer groups for there to be income equality (E = 22%).

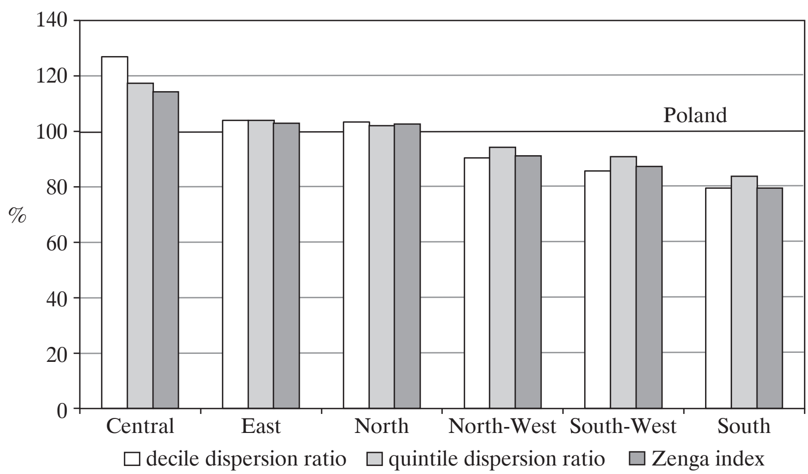

Fig. 2. Selected Inequality Measures for Macroregions (Poland = 100)

Source: authors' elaborations.

The relative poverty threshold established as 60% of equivalent national median income, and the relative wealth line established as the median estimated by means of the Bernstein estimator, are equal to 1181.85 PLN and 5909.23 PLN, respectively. Estimates of the poverty index (head-count ratio (8)) and wealth index (9) for each macroregion based on these thresholds are presented in Table 8. Also contained in the table are the poverty thresholds and wealth lines estimated separately for each macroregion. The indexed values of poverty and wealth ratios (Poland = 100%) are presented in Figure 3.

Table 8. Estimated Poverty and Wealth Measures for Macroregions

Macroregion	Poverty line	Head-count ratio	Wealth line	Wealth index
Central	1394.94	12.73	6974.68	5.42
Southern	1247.39	12.04	6236.93	1.52
Eastern	1085.24	20.12	5426.19	1.68
North-Western	1242.65	12.99	6213.26	1.63
South-Western	1211.82	12.49	6059.09	1.81
Northern	1204.10	16.72	6020.48	2.26
Poland	1181.85	14.46	5909.23	2.56

Source: authors' calculations based on the HBS 2014 sample.

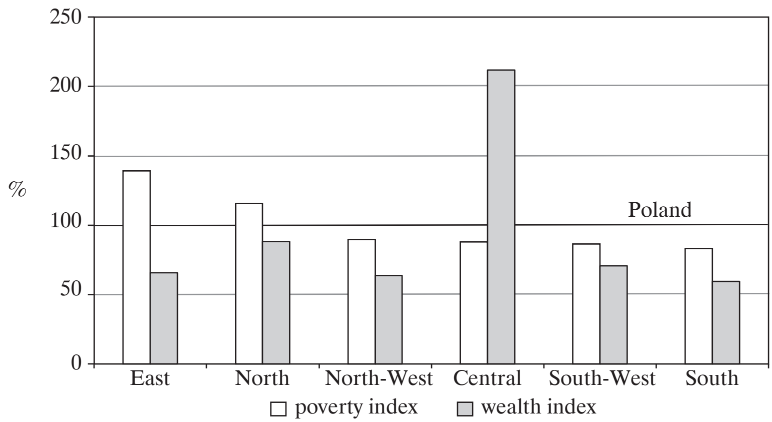

Fig. 3. Poverty and Wealth Measures for Macroregions (Poland = 100)

Source: authors' elaborations.

It is worth noting that the ordering of Polish macroregions by poverty rates is different from the ordering by inequality levels – for some regions (Central) relatively high income inequality does not coincide with high poverty rates, and inversely, relatively low inequality does not always induce low poverty rates (North-Western region). On the other hand, for highly unequal distributions (Central, Eastern), one can observe a large discrepancy between poverty and wealth rates (Figure 3), indicating different within--region inequality patterns – a large amount of inequality due to extremely low income groups (the case of the Eastern region) or extremely high income groups (the Central region).

6. Conclusion

Analysis of income and wage distribution is strictly connected with the estimation of inequality and poverty measures based on quantiles. Therefore, for income data usually originating from sample surveys, it becomes crucial to use the quantile estimators that present satisfying statistical properties. In this paper, the Huang-Brill and Bernstein estimators have been proposed and analysed from the point of view of their sampling errors under several income distribution models. In the simulation studies the properties of these estimators have been compared with the classical one, which is most often applied in practice. The results of the calculations reveal that the Bernstein estimator performs better than its competitors – its root mean squared error (*RMSE*) is much smaller than the one observed for the other quantile estimators and its relative bias (*BIAS*) is also smaller, especially when the quantiles of higher orders are taken into account. Consequently, the Bernstein estimator has been applied to the estimation of various inequality measures for NUTS1 regions in Poland.

The reliable quantile estimators, as well as various inequality, poverty and wealth measures based on them, enabled us to analyse income distributions in Poland by macroregion. The analysis revealed substantial discrepancies between regions in Poland, which can be the basis for further analysis by economists and social-policy makers.

Bibliography

Arcagni, A. (2016) "On the Decomposition by Sources of the Zenga 1984 Point and Synthetic Inequality Indexes". *Statistical Methods & Applications* 26 (1): 113–33, https://doi.org/10.1007/s10260-016-0360.0.

Brzeziński, M. (2014) "Statistical Inference for Richness Measures". *Applied Economics* 46 (14): 1599–1608, https://doi.org/10.1080/00036846.2014.880106.

Greselin, F., Pasquazzi, L. and Zitikis, R. (2013) "Contrasting the Gini and Zenga Indices of Economic Inequality". *Journal of Applied Statistics* 40 (2): 282–97, https://doi.org/10.1080/02664763.2012.740627.

Harrell, F. E. and Davis, C. E. (1982) "A New Distribution-Free Quantile Estimator". *Biometrika* 69: 635–40, https://doi.org/10.1093/biomet/69.3.635.

Huang, M. L. and Brill, P. H. (1999) "A Level Crossing Quantile Estimation Method". *Statistics & Probability Letters* 45: 111–19, https://doi.org/10.1016/s0167-7152(99)00049-8.

Jędrzejczak, A. (2012) "Estimation of Standard Errors of Selected Income Concentration Measures on the Basis of Polish HBS". *International Advances in Economic Research* 18 (3): 287–97.

Jędrzejczak, A. (2015) “Asymptotic Properties of Some Estimators for Gini and Zenga Inequality Measures: a Simulation Study”. *Statistica & Applicazioni* 13: 143–62.

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

Panek, T. (2011) *Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru* [Poverty, social exclusion, and inequality. The theory and practice of measurement]. Warsaw: SGH.

Peichl, A., Schaefer, T. and Scheicher, Ch. (2008) “Measuring Richness and Poverty: A Micro Data Application to Europe and Germany”. IZA Discussion Paper 3790, http://ftp.iza.org/dp3790.pdf (accessed: 13 June 2018).

Pekasiewicz, D. (2015) *Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach społeczno-ekonomicznych* [Order statistics in estimation procedures and their application in socio-economic research]. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Zieliński, R. (2006) “Small-Sample Quantile Estimators in a Large Nonparametric Model”. *Communications in Statistics Theory and Methods* 35: 1223–41, https://doi.org/10.1080/03610920600692656.

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

Abstract

Własności wybranych miar nierównomierności opartych na kwantylach i ich zastosowanie w analizach rozkładów dochodów według makroregionów w Polsce

Kwantyle rozkładu dochodów są wykorzystywane do szacowania różnorodnych miar nierówności, analiz ubóstwa i bogactwa gospodarstw domowych. Najczęściej są one szacowane z użyciem klasycznego estymatora, będącego statystyką pozycyjną odpowiedniej rangi. Głównym celem pracy jest porównanie własności klasycznego estymatora kwantyla z własnościami estymatorów zaproponowanych przez M.L. Huanga i P.H. Brilla oraz Bernsteina. W celu zbadania obciążeń i błędów średniokwadratowych estymatorów kwantyli i miar nierówności opartych na kwantylach przeprowadzono eksperymenty Monte Carlo, rozważając różne liczebności prób i różne rozkłady. W pracy przedstawiono wyniki badań dla populacji o rozkładach lognormalnym i Daguma, które najczęściej charakteryzującą dochody gospodarstw domowych. Wyniki eksperymentów symulacyjnych wskazują, że spośród rozważanych estymatorów najlepsze własności ma estymator Bernsteina, dlatego został on wykorzystany do oszacowania miar nierówności dochodowych, ubóstwa i bogactwa w Polsce w 2014 r. z uwzględnieniem podziału kraju na makroregiony. Analizy przeprowadzono na podstawie danych pochodzących z badania budżetów gospodarstw domowych prowadzonego przez Główny Urząd Statystyczny.

Słowa kluczowe: rozkład dochodu, nierówność, ubóstwo, bogactwo, estymator kwantyla.