Estimation of Quantile Ratios of the Dagum Distribution

Wojciech Zieliński[1], Alina Jędrzejczak[2], Dorota Pekasiewicz[3]

Abstract

Inequality measures based on ratios of quantiles are frequently applied in economic research, especially to the a-nalysis of income distributions in different divisions. Simple quantile ratios or quantile dispersion ratios, which can be considered supplementary to the popular Gini and Zenga indices, are often applied in comparison of incomes for various subpopulations or to assess inequality changes over time. They have the advantage of being focused on extremal income groups that are especially interesting from the point of view of economic inequality and polarization. In the paper a confidence interval for such measures, assuming the Dagum distribution, is constructed. The ends of the confidence interval depend on an unknown shape parameter of the underlying income distribution model. In applications this parameter must be estimated from the data. The constructed confidence interval, applied to decile and quintile ratios, was implemented to the analysis of income inequaity in Poland. The quantile-based inequality measures have been estimated for the Polish macro-regions (NUTS1) and for the whole country, on the basis of micro-data coming from the Household Budget Survey 2015.

Keywords: *Dagum distribution, income inequality, quantile ratios*

JEL Classification: C13, C18

DOI: 10.14659/SEMF.2018.01.61

1 Introduction

Distribution quantiles of a random variable *X* or estimators of these quantiles have frequently been applied to the construction of numerous inequality indices and indicators. Among them the most popular are quintile and decile share ratios (see e.g.: Panek, 2011) mainly for their simplicity and straightforward economic interpretation The income quintile share ratio is calculated as the ratio of income received by the 20% of the population with the highest income to that received by the 20% of the population with the lowest income or as the ratio of the top quintile to the bottom quintile. Similarly we can define decile share ratios. More sophisticated measures of income inequality have been constructed using ratios (or differences) between population and income quantiles. Probably the first of such measures was the Holme's coefficient standardized by Bortkiewicz, which is based on the quantiles of

[1]Corresponding author: Warsaw University of Life Sciences, Department of Econometrics and Statistics, Nowoursynowska 159 St., 02-776 Warsaw, Poland, wojciech_zielinski@sggw.pl

[2] University of Łódź, Institute of Statistics and Demography, Department of Statistical Methods, 41/43 Rewolucji 1905 St., 90-214 Łódź, Poland, jedrzej@uni.lodz.pl.

[3] University of Łódź, Institute of Statistics and Demography, Department of Statistical Methods, 41/43 Rewolucji 1905 St., 90-214 Łódź, Poland, pekasiewicz@uni.lodz.pl.

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. The quantile based inequality measures were found more sensitive to changes of income inequality in particular parts of income distribution, especially in the tails.

Quantile-based inequality measures are traditionally estimated using the classical quantile estimator based on a relevant order statistic. Their estimators can also be obtained in the parametric approach based on a theoretical income distribution model. The parametric estimators of quantile ratios which are based on income distribution models with sensible stochastic or empirical foundations, have the advantage of being robust to irregularities coming from imperfect data collection methods. Moreover, based on these models, the confidence intervals can be derived to provide information about how close the point estimate is to the true parameter with the margin of error. The main objective of this study was to provide a confidence interval of quantile ratio for the three-parameter Dagum distribution.

The second section of this paper is devoted to the non-parametric point estimators of quantile ratios. The third part introduces a parametric estimator of quantile ratios based on the Dagum distribution and finally the confidence interval for this parameter is derived. In the last part of the paper we present the application of the proposed estimation methods to income inequality analysis based on the Polish Household Budget Survey (HBS) data.

2 Point estimators of quantile ratios

Let X be a continuous random variable with a distribution function F and let $\gamma_p = F^{-1}(p)$ be the p-quantile of the random variable X, where $p \in (0,1)$. If F is continuous and strictly increasing distribution function, the p^{th} quantile is uniquely determined.

Among estimators of quantiles γ_p we can distinguish the standard estimator, Huang-Brill estimator, Harrel-Davis estimator and Bernstein estimator, to name only a few (Huang and Brill, 1999; Harrell and Davis, 1982; Zieliński, 2006).

An application of these estimators to the evaluation of quantiles and quantile ratios has recently been presented in Jędrzejczak and Pekasiewicz (2017).

In what follows we apply the well-known estimator of the quantile γ_p:

$$\hat{\gamma}_p = X_{\lfloor pn \rfloor + 1:n}, \tag{1}$$

where $X_{1:n} \le X_{2:n} \le ... \le X_{n:n}$ is the ordered sample of a random sample $X_1, X_2, ..., X_n$ and $\lfloor x \rfloor$ denotes the greatest integer not greater than *x*.

In this case the estimator of the quantile ratio $r_{\alpha,\beta} = \frac{\gamma_\beta}{\gamma_\alpha}$, where $0 < \alpha < \beta < 1$, has the following form:

$$\hat{r}_{\alpha,\beta} = \frac{X_{\lfloor \beta n \rfloor + 1:n}}{X_{\lfloor \alpha n \rfloor + 1:n}}. \tag{2}$$

3 Point and interval estimators of quantile ratios for the Dagum distribution

In our considerations we confine ourselves to the Dagum distribution, i.e. throughout the paper it will be assumed that the distribution of the population income is the Dagum one. As it was mentioned above, the Dagum distribution fits population income quite well for many countries all over the world. It is based on both empirical and stochastic foundations, similarly to the Pareto model (Dagum, 1977).

The probability density function of the Dagum distribution is given by (Kleiber and Kotz, 2003):

$$f_{a,v,\lambda}(x) = \frac{av}{\lambda}\left(\frac{x}{\lambda}\right)^{av-1}\left(1+\left(\frac{x}{\lambda}\right)^{v}\right)^{-a-1} \quad \text{for } x > 0, \tag{3}$$

where $a, v, \lambda > 0$. Its cumulative distribution function equals:

$$F_{a,v,\lambda}(x) = \left(1+\left(\frac{x}{\lambda}\right)^{-v}\right)^{-a} \quad \text{for } x > 0 \tag{4}$$

and the quantile function is

$$Q_{a,v,\lambda}(q) = \lambda\left(q^{-\frac{1}{a}} - 1\right)^{-\frac{1}{v}} \quad \text{for } 0 < q < 1, \tag{5}$$

so the ratio of quantiles of the Dagum distribution has the following form:

$$r_{\alpha,\beta} = \frac{\gamma_\beta}{\gamma_\alpha} = \left(\frac{\beta^{-\frac{1}{a}} - 1}{\alpha^{-\frac{1}{a}} - 1}\right)^{-\frac{1}{v}} \quad \text{for } 0 < \alpha < \beta < 1. \tag{6}$$

The problem lies in constructing a confidence interval at the confidence level δ for a ratio of quantiles $r_{\alpha,\beta}$ based on $\hat{r}_{\alpha,\beta}$.

For “large” sample sizes, i.e. when $n \to \infty$, it is known that $\hat{r}_{\alpha,\beta}$, defined by the formula (2), is a consistent estimator of $r_{\alpha,\beta}$ (David and Nagaraja, 2003; Serfling, 1980).

Let $Y_i = \ln X_i$ and γ_α^Y, γ_β^Y denote the quantiles of *Y*. We have (David and Nagaraja, 2003; Serfling, 1980):

$$\sqrt{n}\begin{bmatrix} Y_{\lfloor n\alpha \rfloor+1:n} - \gamma_\alpha^Y \\ Y_{\lfloor n\beta \rfloor+1:n} - \gamma_\beta^Y \end{bmatrix} \to N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\alpha(1-\alpha)}{\left(f_Y\left(\gamma_\alpha^Y\right)\right)^2} & \frac{\alpha(1-\beta)}{f_Y\left(\gamma_\alpha^Y\right) f_Y\left(\gamma_\beta^Y\right)} \\ \frac{\alpha(1-\beta)}{f_Y\left(\gamma_\alpha^Y\right) f_Y\left(\gamma_\beta^Y\right)} & \frac{\beta(1-\beta)}{\left(f_Y\left(\gamma_\beta^Y\right)\right)^2} \end{bmatrix}\right), \tag{7}$$

where $f_Y(\cdot)$ is the probability density function of *Y*. Hence

$$\sqrt{n}\left[\left(Y_{\lfloor n\beta \rfloor+1:n} - Y_{\lfloor n\alpha \rfloor+1:n}\right) - \left(\gamma_\beta^Y - \gamma_\alpha^Y\right)\right] \to N\left(0, \sigma^2\right), \tag{8}$$

where

$$\sigma^2 = \frac{\beta(1-\beta)}{\left(f_Y\left(\gamma_\beta^Y\right)\right)^2} + \frac{\alpha(1-\alpha)}{\left(f_Y\left(\gamma_\alpha^Y\right)\right)^2} - 2\frac{\alpha(1-\beta)}{f_Y\left(\gamma_\alpha^Y\right) f_Y\left(\gamma_\beta^Y\right)}. \tag{9}$$

Since $Y_{\lfloor np \rfloor+1:n} = \ln X_{\lfloor np \rfloor+1:n}$ and in Dagum distribution $\gamma_p^Y = \ln \gamma_p$ we have:

$$\sqrt{n}\left(\ln\frac{X_{\lfloor n\beta \rfloor+1:n}}{X_{\lfloor n\alpha \rfloor+1:n}} - \ln\frac{\gamma_\beta}{\gamma_\alpha}\right) \to N\left(0, \sigma^2\right) \tag{10}$$

and applying Delta method (Greene, 2003) with $g(t) = e^t$ we obtain:

$$\sqrt{n}\left(\frac{X_{\lfloor n\beta \rfloor+1:n}}{X_{\lfloor n\alpha \rfloor+1:n}} - \frac{\gamma_\beta}{\gamma_\alpha}\right) \to \left(\frac{\gamma_\beta}{\gamma_\alpha}\right) N\left(0, \sigma^2\right), \tag{11}$$

where

$$\sigma^2 = \frac{1}{(av)^2}\left(\frac{1-\beta}{\beta}\frac{1}{\left(1-\beta^{\frac{1}{a}}\right)^2} + \frac{1-\alpha}{\alpha}\frac{1}{\left(1-\alpha^{\frac{1}{a}}\right)^2} - 2\frac{1-\beta}{\beta}\frac{1}{\left(1-\alpha^{\frac{1}{a}}\right)\left(1-\beta^{\frac{1}{a}}\right)}\right). \tag{12}$$

The parameter *v* can be determined using the parameter *a* and the quantile ratio $r_{\alpha,\beta}$ of the Dagum distribution in the following way: $v = \ln\left(\frac{\alpha^{-\frac{1}{a}} - 1}{\beta^{-\frac{1}{a}} - 1}\right)\left(\ln r_{\alpha,\beta}\right)^{-1}$.

Hence

$$\sigma^2 = \left(\ln r_{\alpha,\beta}\right)^2 \omega^2(a), \tag{13}$$

where

$$\omega^2(a)=\left(a\ln\left(\frac{\alpha^{-\frac{1}{a}}-1}{\beta^{-\frac{1}{a}}-1}\right)\right)^{-2}\left(\frac{1-\beta}{\beta}\frac{1}{\left(1-\beta^{\frac{1}{a}}\right)^2}+\frac{1-\alpha}{\alpha}\frac{1}{\left(1-\alpha^{\frac{1}{a}}\right)^2}-2\frac{1-\beta}{\beta}\frac{1}{\left(1-\alpha^{\frac{1}{a}}\right)\left(1-\beta^{\frac{1}{a}}\right)}\right). \tag{14}$$

The confidence interval for the quantile ratio of the Dagum distribution has the following form:

$$P_{r,a}\left\{\sqrt{n}\left|\frac{\hat{r}_{\alpha,\beta}-r_{\alpha,\beta}}{\omega(a)r_{\alpha,\beta}\ln r_{\alpha,\beta}}\right|\le u_{(1+\delta)/2}\right\}=\delta, \tag{15}$$

where δ is a given confidence level and $u_{(1+\delta)/2}$ is the quantile of N(0,1) distribution. Solving the above inequality with respect to $r_{\alpha,\beta}$ we obtain:

$$\left(\frac{\hat{r}_{\alpha,\beta}z_+(a)}{W\left(\hat{r}_{\alpha,\beta}z_+(a)e^{z_+(a)}\right)},\frac{\hat{r}_{\alpha,\beta}z_-(a)}{W\left(\hat{r}_{\alpha,\beta}z_-(a)e^{z_-(a)}\right)}\right), \tag{16}$$

where $z_+(a)=\dfrac{\sqrt{n}}{u_{(1+\delta)/2}\omega(a)}$, $z_-(a)=\dfrac{\sqrt{n}}{u_{(1-\delta)/2}\omega(a)}$ and *W* is the *W* Lambert function.

Note that the ends of the confidence interval depend on an unknown shape parameter *a*. This parameter should be estimated from the data. As the estimation techniques we can choose for example: Maximum Likelihood Method, Method of Moments or L-moments, Methods of Ordinary Least-Squares or Weighted Least-Squares and the methods based on percentiles (Dey et al., 2017).

4 Application of quantile ratios of the Dagum distribution to the Polish income data

The inequality measures based on deciles and quintiles have been applied to income inequality analysis in Poland by macro-region (NUTS1). The calculations were based on the micro data coming from the Household Budget Survey 2015 conducted by the Central Statistical Office of Poland. The variable of interest was household available income, the basic income category of HBS sample.

To adjust the available income for differences in family size, we adopted the recent OECD equivalence scale, where the household income was divided by the square root of relevant household size. Basic characteristics of the HBS sample by macro-region are presented in Table 1.

Table 2 contains the results of fitting the Dagum distribution to the empirical data. In Fig. 1-2 there are histograms accompanied by fitted Dagum density curves describing income distributions in selected macro-regions.

Analysing the results presented in Fig. 1 and 2 one can observe very high consistency of the empirical distributions with the theoretical ones for the selected *Central* and *Southern* macroregions. Similar results were obtained for the other macro-regions. It can also be 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 Table 2. The overlap coefficient also known as the coefficient of distribution similarity was proposed by Vielrose in 1960, and represents the "common part" of the empirical and theoretical distributions.

Table 1. Numerical characteristics of equivalent income in macroregions.

Macroregion	Number of households	Minimum	Maximum	Average	Standard Deviation
Central	8058	3.78	69047.65	2763.75	2248.36
Southern	7465	88.00	26400.00	2358.24	1224.19
Eastern	6207	1.77	52702.31	2120.09	1502.48
North-western	5608	36.69	105846.64	2378.83	1909.83
South-western	3914	7.85	28394.00	2428.96	1383.25
Northern	5608	10.00	67370.00	2299.07	1826.83
Poland	36860	1.77	105846.64	2408.43	1760.20

Table 2. Approximation of equivalent income distributions for macroregions.

Macroregion	Dagum distribution parameters			Overlap measure
	a	v	λ	
Central	0.817	3.175	2512.444	0.989
Southern	0.888	3.953	2223.571	0.994
Eastern	0.775	3.677	2041.918	0.993
North-western	0.829	3.871	2260.756	0.996
South-western	0.697	4.051	2494.583	0.995
Northern	0.825	3.527	2138.206	0.993
Poland	0.821	3.588	2261.110	0.996

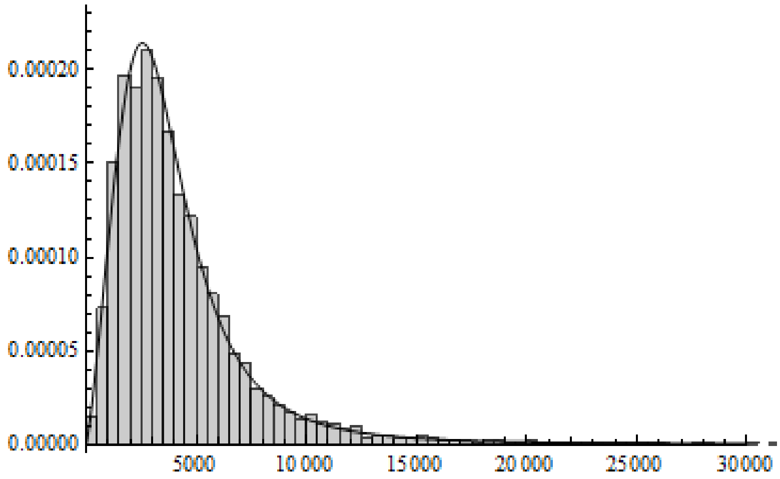

Fig. 1. Equivalent income distribution for Central Macroregion and fitted Dagum density.

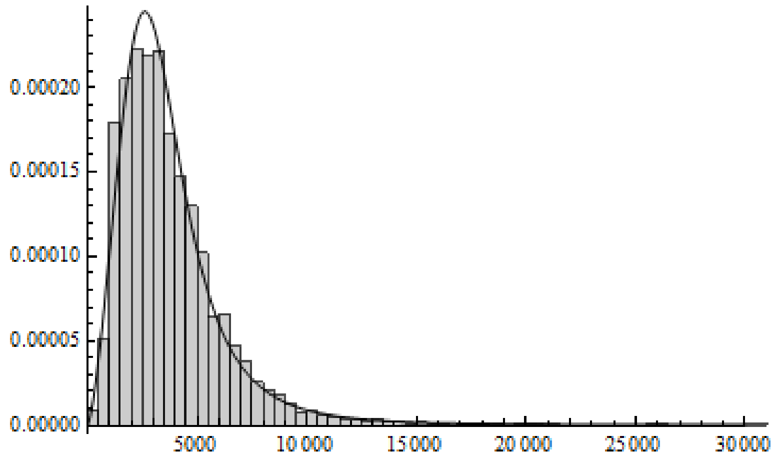

Fig. 2. Income distribution for Southern Macroregion and fitted Dagum density.

Based on the confidence interval for quantile ratios which has been proposed in the section 3 (see: eq. (16)), we constructed the confidence intervals for the income quintile ratio:

$$W_{20:20} = \frac{\gamma_{0.8}}{\gamma_{0.2}}, \tag{17}$$

and the income decile ratio:

$$W_{10:10} = \frac{\gamma_{0.9}}{\gamma_{0.1}}, \tag{18}$$

where $\gamma_{0.8}$, $\gamma_{0.2}$ are quintiles and $\gamma_{0.9}$, $\gamma_{0.1}$ are deciles.

The shape parameter *a* has been estimated by the maximum likelihood method which is the most popular one due to its good asymptotic properties.

The basic results of the inequality analysis have been outlined in Table 3. The estimated values of quintile and decile ratios indicate the Central macroregion as the one with the highest income inequality level. It is especially visible for the decile share ratio which shows the “social distance” between the rich and the poor of about 4.3 times $\gamma_{0.1}$ (more precisely, it is a number from the interval (4.1673, 4.5685)).

Table 3. Confidence intervals of quintile and decile share ratios.

Macro-region	$\hat{\mathbf{W}}_{20:20}$	**Confidence Interval of $\mathbf{W}_{20:20}$**	$\hat{\mathbf{W}}_{10:10}$	**Confidence Interval of $\mathbf{W}_{10:10}$**
Central	2.4787	(2.4252, 2.5376)	4.3039	(4.1673, 4.5685)
Southern	2.1319	(2.0917, 2.1760)	3.2186	(3.1324, 3.3143)
Eastern	2.2615	(2.2115, 2.3171)	3.5354	(3.4243, 3.6606)
North-western	2.1751	(2.1282, 2.2270)	3.2982	(3.1985, 3.4104)
South-western	2.2010	(2.1433, 2.2665)	3.4080	(3.2803, 3.5557)
Northern	2.3209	(2.2690, 2.3811)	3.7146	(3.5914, 3.8541)
Poland	2.2863	(2.2646, 2.3089)	3.6293	(3.5800, 3.6811)

5 Conclusions

In the paper an asymptotic confidence interval for a ratio of quantiles of the Dagum distribution was constructed. The ends of this confidence interval depend on the shape parameter *a* of the Dagum distribution. This parameter can be estimated by the maximum likelihood method, which was shown to be efficient for large sample sizes.

The proposed method of the estimation of quantile ratios can be applied to income inequality analysis when a household or personal income follows the Dagum distribution. The confidence interval constructed above is symmetrical in the following sense: the risks of underestimation as well as overestimation of the true population parameter are the same. It can be a useful tool for social-policy makers interested in the evaluation of current income inequality level with the acceptable margin of error.

Income quintile and decile ratios considered in the paper have the advantage of being focused on extremal income groups that are especially interesting from the point of view of economic inequality and polarization. They can assess the “social distance” between rich and poor groups of income receivers. The empirical analysis revealed substantial discrepancies between extremal income groups in the macro-regions of Poland. Income distribution in the most affluent *central* region turned out to be the most unequal what is especially visible in the extremal decile groups. All the distributions under consideration presented very high consistency with the Dagum model.

References

Dagum, C. (1977). A New Model of Personal Income Distribution: Specification and Estimation. *Economie Appliquee, 30,* 413-437.

David, H. A. & Nagaraja, H. N. (2003). *Order Statistics, Third Edition*. John Wiley & Sons.

Dey, S., Al-Zahrani, B. & Basloom, S. (2017). Dagum Distribution: Properties and Different Methods of Estimation. *International Journal of Statistics and Probability, 6,* 74-92.

Greene, W. H. (2003). *Econometric Analysis (5th ed.)*. Prentice Hall.

Harrell, F. E. & Davis C. E. (1982). A New Distribution-Free Quantile Estimator. *Biometrika*, *69*, 635–640.

Huang M. L. & Brill P. H. (1999). A Level Crossing Quantile Estimation Method. *Statistics & Probability Letters, 45*, 111–119.

Jędrzejczak, A. & Pekasiewicz, D. (2017). Analysis of the properties of selected inequality measures based on quantiles with the application to the Polish income data. In: Papież, M. & and Śmiech, S. *11th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings. Cracow: Foundation of the Cracow,* 113-122.

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

Panek, T. (2011). *Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru*. Szkoła Główna Handlowa, Warszawa.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Vielrose E. (1960). *Rozkład dochodów według wielkości,* Polskie Wydawnictwo Gospodarcze, Warszawa.

Zieliński, R. (2006). Small-Sample Quantile Estimators in a Large Nonparametric Model. *Communications in Statistics Theory and Methods*, 35, 1223–1241.