# ON THE PROBLEM WITH NEGATIVE QUANTITIES IN GINI AND CONCENTRATION INDICES CALCULATION WITH APPLICATION TO THE POLISH PERSONAL INCOME TAX

**CYPRIAN KOZYRA**
Wrocław University of Economics, Faculty of Management, Information Systems and Finance,
Department of Statistics,
Komandorska St. 118/120, Wrocław, Poland
e-mail: cyprian.kozyra@ue.wroc.pl

## Abstract

*Since the year 2014 taxpayers using child tax credits have been allowed not only to decrease their tax to zero, but also to receive the amount of overall tax credit from the Polish tax system. In this way Polish personal income tax calculated for the year 2014 can be negative when is analyzed as tax due. Because the sum of negative taxes was equal 2.4 % of the total tax due in the year 2014, negative values are not a marginal problem. Different approaches to calculating Gini and concentration indices with negative data are compared, among others that of Chen, Tsaur and Rhai of 1982, corrected by Berebbi and Silber in 1985 (CTR-BS) and then that of Rafinetti, Silleti and Vernizzi of 2014. Additionally, our own proposals are considered, and pros and cons of all the approaches are compared. The consequences of negative taxes on measures of progressivity of Polish personal income tax are analyzed by means of Gini and concentration indices adapted for negative values.*

***Key words:*** *Gini index, negative tax, absolute Lorenz curve.*

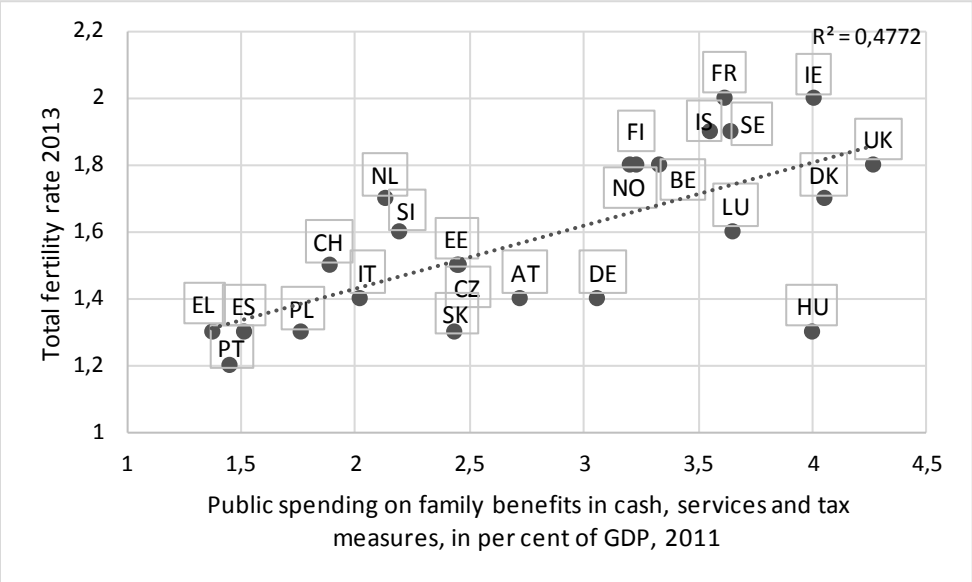
## 1. Introduction

In the year 2007 child tax credits were introduced in the Polish system of Personal Income Tax (PIT) in aim to help families raising children and to encourage families to increase the number of children, improving in this way a difficult demographic situation. Over many years tax due was decreased by child tax credit only to zero, and it cannot be less than zero. Since the year 2014 this rule has been changed, and now families with lower income and tax are allowed not only to decrease their PIT to zero, but they can also receive the rest of overall tax credit from the Polish tax system. According to information from Polish Ministry of Finance, the sum of negative tax was equal to 2.4% of total tax due in the year 2014, so the negative values are not a marginal problem. Taking into account negative data can change some assumptions and possible interpretations of methods analyzing inequality of income and tax, e.g. Gini index of inequality can take values greater than one, that is typically interpreted as the value of the case of maximal inequality, when only nonnegative data are considered. The aim of this paper is to compare different methods of analyzing income and tax with negative values. Especially the influence on measured tax progressivity of these different approaches by means of Kakwani index is considered. The paper is organized as follows: the second section presents Polish PIT and the mechanism of negative tax appearing in the system, the third section reviews different formal approaches to negative data in income and tax analysis, the next section provides description of empirical data with results obtained for this data and the last section contains conclusions.

## 2. Child Tax Credit in the Polish PIT in 2014

As it was stated above, till the year 2014 families with low income very often had no possibility of employment of the whole children tax credit, because their calculated PIT was lower than the potentially subtracted credit. Most of Polish tax payers are paying PIT according to progressive scale of income reduced by social insurance contribution: 0% (tax-free amount) up to 3091 PLN, next 18% up to 85528 PLN and then 32% of surplus. There are many tax allowances, the two most popular are marriage tax splitting (sum of marriage income is divided by two before applying progressive scale and next the calculated tax is multiplied also by two) and children tax credit, which is equal to 1112.04 PLN for the first child (but only if sum of parents' income is not higher than 112000 PLN, otherwise credit is equal to zero), 1112.04 PLN for every second child, 2000.04 for the third child and 2700 PLN for the fourth and each next child. Since the year 2014 PIT can not only be reduced to zero, but also the rest of children tax credit can be received by parents. However, there is a limit of such a refund, which is equal to the sum of social insurance and health insurance contributions, paid by parents. Therefore Polish PIT calculated for the year 2014 can be negative in the strict legal meaning, but cannot be negative in the broader meaning, when PIT and social and health insurances contributions are summed, because income after taxation cannot be greater than the original income. One can avoid playing with negative quantities in this way, analyzing the tax in the broader meaning. But very often analysis of PIT system (e.g. progressivity) is restricted to PIT in the strict sense. Transfers of child tax credit are also treated as negative tax in the reports of Ministry of Finance, because income after taxation can be greater than the taxation base. Figure 1 is produced using the OECD family database[1].

Figure 1: Correlation diagram between public spending on family and total fertility rate




Source: the author based on the OECD family database.

Negative tax in Polish PIT system is different to Friedman negative tax, because the source of refund is the number of children rather than low income. Alternatively, one can perceive this negative tax as a special family benefit, because the reason of its introduction to Polish PIT system was the support of families with children. Negative tax is treated by the system of Polish social aid as a transfer, which increases the final income and can lead to the loss of

[1] Available at http://www.oecd.org/social/family/database.htm (accessed April 20, 2016).

some entitlements to additional benefits. But unambiguously it is strictly dependent on the tax system, because such a benefit is greater than zero if and only if PIT in the legal sense is equal to zero. Therefore the simplest way of such tax-benefit system analysis is including negative PIT to the analysis. Growth of children tax credits over the years (see Mazurek, 2015) and introduction of negative PIT were justified mainly by bad demographic situation in Poland (total fertility rate far below survival limit equal to 2.1) and low financial support of Polish families compared to other European countries. It is shown in Figure 1 as linear regression of total fertility rate (which is defined by OECD as the average number of children born per woman over a lifetime) on lagged public spending on family benefits in OECD countries several years ago (the share in GDP of public spending in Poland seems to be greater now than in the year 2011, but total fertility rate has a similar value as in the year 2013). Increase in fertility rate is expected by government in Poland after steady increase of family benefits.

Table 1: Information from the Polish Ministry of Finance about PIT in 2013 and 2014

| Year | Child tax credits [thous. PLN] | Number of taxpayers using child tax credit [pers.] | Number of declared children [pers.] |
|---|---|---|---|
| 2013 | 5 529 539 | 4 024 017 | 5 825 310 |
| 2014 as in 2013 | 5 588 813 | 4 050 868 | |
| 2014 negative tax because of child tax credit | 1 337 857 | 1 195 698* | |
| 2014 sum | 6 926 670 | 4 050 868 | 6 468 411** |

Note: * all included in above number, ** total number of children, containing children undeclared in previous years because of lack of negative tax.

Source: the author.

The effect of introduction of negative PIT is presented in Table 1, comparing children tax credits in the years 2013 and 2014 according to the reports of Ministry of Finance. One can see that in the year 2014 the number of child tax credit and the number of declared children significantly increased. In the previous year many children were not declared, because of no possibility of additional refund, when PIT was reduced to zero. In the year 2014 overall PIT due was equal to 55 566 791 thousand PLN, so without the negative tax it will be greater by ca. 2.4%. Table 1 was compiled using information from the Polish Ministry of Finance[2].

## 3. Methods of Analyzing Inequality and Concentration Coefficient with Negative Data

Classic analysis of income and tax data usually assumes dealing with only non-negative data, however Lorenz curves for datasets with different amount of negative values were already analyzed by Shutz (1951). The simplest way of avoiding negative data in the analysis is erasing it or coding it bottom to zero, but obviously it influences the results of analysis, because the lower tail of data is not correctly presented in the analyzed distribution. Another simple way is calculating measures of concentration using standard equation. But when the amount of negative data is great, one can see that Gini inequality index, that may be given by formula (1) (see e.g. Rafinetti et al., 2015), can result outside of [0,1] interval:

$$G = \frac{\Delta}{2\mu}, \tag{1}$$

[2] Available at http://www.finanse.mf.gov.pl/documents/766655/2970107/Informacja (accessed April 20, 2016).

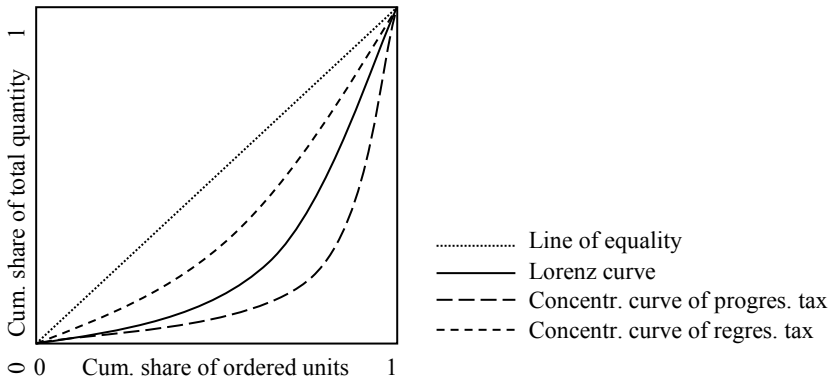
where $\Delta$ is the mean absolute difference, and $\mu$ is mean value of the analyzed quantity. When mean value is tending to zero, Gini index could tend to infinity, but when mean value is negative, Gini index is also negative. Therefore value of Gini inequality index, calculated from classic equation for possible great amount of negative quantities, could take a very big or a negative value, which can be hard to interpret. However, standard Gini index is also calculated in this paper in aim to compare it with other methods.

Many measures used in the analysis of income and tax are calculated as a difference between two Gini or concentration indices, e.g. Kakwani progressivity index, redistribution index and Atkinson-Plotnick tax equity index (see Mazurek, 2015). Because the aim of this paper is the analysis of consequences of negative PIT on tax system progressivity, Kakwani index (Kakwani, 1977) given by formula (2) was selected from among alternative measures of progressivity:

$$\Pi^K = C_{T|X} - G_X , \tag{2}$$

where $G_X$ is Gini index of income before taxation and $C_{T|X}$ is concentration index of PIT, calculated for tax ordered according to non-decreasing income before taxation. Kakwani progressivity index is often presented as in terms of $G_X$ as $\left(\Pi^K / G_X\right)\cdot 100\%$. Graphical meaning of these measures is presented in Figure 2. The difference between Lorenz and concentration curves is that data are sorted non-decreasing for Lorenz curve, but in case of concentration curve data are ordered according to another variable (e.g. tax ordered according to income before taxation). Both Lorenz and concentration curves have cumulated distribution function of quantity in X-axis and cumulated share of quantity in the total quantity in Y-axis. Gini (concentration) index is calculated as twice the area between the equality line and Lorenz (concentration) curve, but when concentration curve lies above the equality line (e.g. when the orders of income and tax are reversed), this area is measured with a negative sign. Gini index is equal to zero only in case of equality, and in case of extreme concentration, when the total quantity is assigned to one unit and the rest of units have value zero, Gini index tends to one for non-negative data.

Figure 2: Lorenz and concentration curves drawn for Kakwani progressivity index



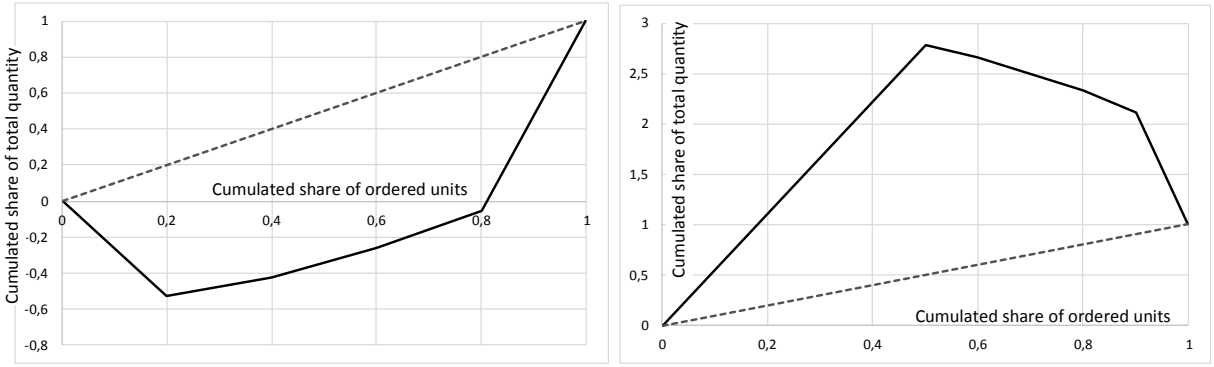

Source: the author based on Kakwani (1977).

The values of Kakwani progressivity index fall within the range $\Pi^K \in [-1-G_X, 1-G_X]$ (Mazurek, 2015). Positive values of the index, $\Pi^K > 0$, represent the progressive tax system ($C_{T|X} > G_X$, tax is more concentrated than income, when e.g. concentration curve for tax lies below Lorenz curve for income before taxation). For the proportional system $\Pi^K = 0$, e.g. when Lorenz curve equals concentration curve. The regressive tax system is described by

negative values of the index, $\Pi^K < 0$, when e.g. concentration curve lies above Lorenz curve. The values of the Kakwani progressivity index can be interpreted as the percentage of overall tax burden that has been transferred from low-income groups to high-income groups, as a result of the progression of the tax system (Mazurek, 2015). When the negative taxes are possible, progression of tax scale could mean that the lower is income, the more negative is calculated tax.

In case of negative data Lorenz and concentration curves can lie partly below X-axis for positive mean, or can be flipped vertically for negative mean. Both situations are shown in Figure 3 for artificial datasets containing a few units. Even if the value of Gini index falls between 0 and 1, inequality can be overestimated in case of negative data. All four compared approaches use Kakwani progressivity index adapted for negative values, two first methods are taken from literature, two next are own ideas:

Figure 3: Lorenz curves for negative data with positive mean (left) and negative mean (right)




Source: the author.

1) Method (Chen et al., 1982) with correction in formula by Berebbi and Silber (1985), called later CTR-BS method, employs normalization term including area below X-axis. Gini index by CTR-BS was calculated with *GiniWegNeg* package (Rafinetti and Aimar, 2016) in *R* project enviroment (R Core Team, 2015).
2) Rafinetti et al. (2015) use different normalization term, modifying standard method with polarized mean value. They also link their method to concentration areas of generalized Lorenz curve, which has in Y-axis mean of cumulated quantity, rather than the cumulated share in total quantity. The authors of this method show undesirable properties of CTR-BS method, when mean is positive close to zero, and also show good properties of their method in this case. Their method, called later RSV method, can also be used for negative or null mean. But not every axiom of inequality measurement is satisfied by RSV method: RSV show calculated polarized Gini index equal to 0.56 for the vector (-5,-5,-5,-5,-5,-5,-5,-5,-5,45.01), but its value undesirably increases to 0.9 after adding to every unit a constant value equal to 5 or after transfer of 45 from the last unit equally divided into the rest of units. Gini index by RSV was calculated also with *GiniWegNeg* package (Rafinetti and Aimar, 2016), where calculation formula presented in Rafinetti et al. (2015) paper is corrected by multiplying by $(n–1)/n$, because the maximum of Gini by RSV equals 1 in their paper. The maximum of Gini index by RSV after corrections in the package is equal to $1–1/n$ as in standard Gini formula for finite population.
3) The first own method of avoiding negative values in progressivity index calculation is quite simple: zero is treated as not distinguished value, but only conventional. Therefore it is assumed that the analyzed variable is measured not on the ratio scale, but rather on the interval scale. In this situation one can obtain transformed non-negative quantities

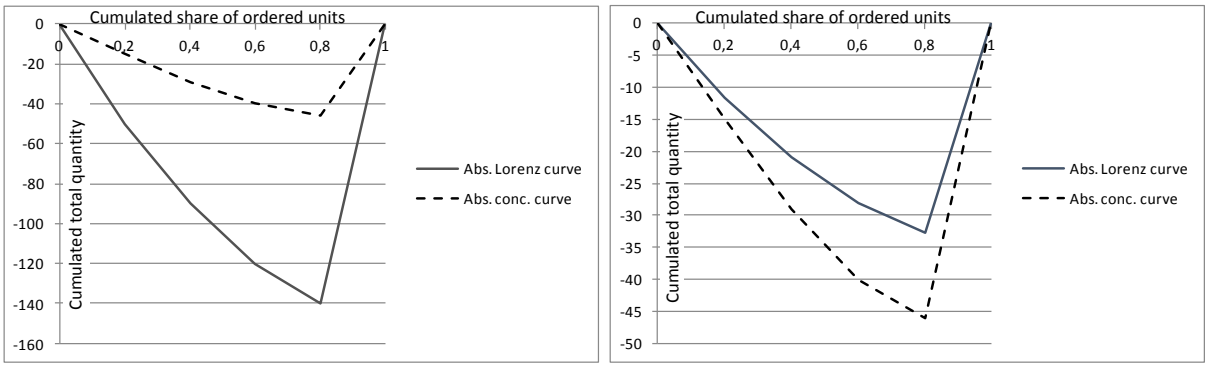
$T$'=$T$-min($T$) and calculate Gini index for $T'$ in the standard way. Such an approach is rather *ad hoc* in its nature, because the total quantity is not kept fixed. The same transformation could be done with data containing only positive values. Such inequality measurement can be called as relative to minimum, because zero conventionally equals the minimum of values. However, in case of equality it is equal to 0/0, which should be interpreted as inequality index equal to zero. Because for negative taxes inequality is reduced through such transformation (minimum is negative), also income before taxation, which is non-negative, should be transformed as $X$'=$X$-min($T$) for calculation of Kakwani progressivity index from equation (2).

4) The last method is applying absolute Gini index (Moyes, 2007; Cowell and van Kerm, 2015), which is derived from absolute Lorenz curve defined by Moyes (1987, 1988) as generalized Lorenz curve of quantities subtracted by mean $Y' = Y - \bar{Y}$. Therefore, it is assumed that quantities are measured on the interval scale, as in the previous method. Classic Lorenz curve, called relative contrary to absolute, can be defined similarly as generalized Lorenz curve of quantities divided by mean $Y' = Y / \bar{Y}$. Hence, measurement on ratio scale is reasonable, which excludes non-negative values from the analysis based on classic Lorenz curve. When data are ordered according to another variable, absolute concentration curve is defined similarly to classic (relative) concentration curve. The example of absolute Lorenz, which always lies below X-axis or as a straight line on X-axis only in case of equality, is presented in Figure 4 (left) with the example of absolute concentration curve. Absolute Gini (concentration) index is defined as twice the area between Lorenz (concentration) curve and line of equality (between points 0 and 1 on X-axis). Concentration curve may lie above X-axis and in this case the area is measured with a negative sign. Absolute Gini index has different properties from the classic (relative) Gini index: according to review of inequality measurement axioms by Ostasiewicz and Ostasiewicz (2015), there is a common subset of axioms satisfied by both measures, but there are also different axioms satisfied separately by relative and absolute measures of inequality. According to Kolm (1976), absolute Gini index could be also classified as "leftist", because it remains constant, when a positive constant is added to all quantities, and increases when all quantities are multiplied by a constant greater than 1. On the contrary relative Gini index could be classified as "rightist", because it decreases, when a positive constant is added to all quantities, and remains constant, when all quantities are multiplied by a constant greater than 1. Absolute Gini coefficient can also be calculated as the half of mean absolute difference $\Delta/2$ (Cowell and van Kerm, 2015). This measure was used in the analysis of quantities with negative data several years ago (e.g. Amiel et al., 1996), but own idea verified in this paper is use of the absolute concentration index. Because tax is obviously far less than income before taxation, absolute concentration curve for tax and absolute Lorenz curve for income are not directly comparable. Hence, concentration curve for tax was adjusted to Lorenz curve for income through dividing Y-axis values of concentration curve by the average rate of the tax $\tau = \sum T \big/ \sum X$. Adjusted concentration curve is presented in Figure 4 (right) with the example of Lorenz curve. For the sake of comparability of absolute curves, the progressivity is measured mainly as the relative difference between adjusted concentration index and absolute Gini index $\left(\Pi^K / G_X\right) \cdot 100\%$ with the same interpretation as for relative curves in formula (2).

Relative Gini coefficients by above described methods were calculated in *GiniWegNeg* package with values of equivalence scale (described in section 4) as weights. Relative concentration indices for tax were calculated in *R* program similarly to Gini index, but with data ordered according to income before taxation. Normalization terms in CTR-BS and RSV

methods were calculated with tax sorted non-decreasingly to get the same denominator as in formula for Gini index by these methods. Absolute Gini and concentration indices were calculated in Microsoft Excel. The copy of employed Microsoft Excel and *R* calculation formulas can be obtained by email on request from the author of this paper.

Figure 4: Example of absolute Lorenz and absolute concentration curves in case of progressive tax scale: original (left) and adjusted (right)




Source: the author.

## 4. Data and Results of Analysis

The empirical comparison of described above approaches is shown by the analysis of real data collected in one of the tax offices in Wrocław in the year 2007. The size of the dataset is $n$=17165, and data concern only the gross income and number or children of each marriage employing tax splitting. Although average wages increased by ca. 40% in Poland in the years 2007-2014, the empirical distribution of gross income was remained the same for better representation of Polish earnings: wages earned in rural areas are lower, but the number of children is greater than in Wrocław, which is one of biggest cities in Poland. Negative tax calculated for the data with child tax credit rules of the year 2014 is equal to 1.9% of the total PIT due calculated for the data. It is less than reported by Ministry of Finance (2.4% as stated in section 2), but our calculations are based on some simplifying assumptions: both spouses have got one employment contract, both pay all social and health contribution and no other tax allowances are used. Because the analyzed households concern a different number of children, modified OECD equivalence scale was applied to data in order to compare equivalized units, but slightly corrected: weight equal to 1 was as usually applied to the first spouse, weight equal to 0.5 was applied as usually to the second spouse, but the weight for a child was calculated as 0.364 proportionally to the structure of Polish population by age groups and net enrollment rate to education in age of 19-24 years, because parents are allowed to child tax credit only for studying children, who are less than 25 years old. This calculation was based on data provided by Central Statistical Office of Poland.[3]

Kakwani progressivity index was calculated for tax before applying child tax credit and tax after applying child tax credit. Polish PIT calculated with progressive scale is reduced first by the health insurance contribution and then children tax credit may be applied. Health insurance contribution is equal to 9% of gross income reduced by social insurance contributions, but 7.75% of health insurance base is subtracted from the calculated PIT. It is income before applying child tax credit in our analysis, and this net income cannot be greater

[3] Available at http://stat.gov.pl/en/topics/population/population/structure-of-the-population-by-2014,7,1.html and then also at http://stat.gov.pl/download/gfx/portalinformacyjny/en/defaultaktualnosci/3306/2/8/1/higher_education_institutions.pdf (accessed in April 20, 2016).

than taxation base, because tax cannot be negative in this stage of calculation. Income after applying child can be greater than the taxation base, and the final tax can be negative as stated in section 2.

Results of the analysis are shown in Table 2 for non-negative tax before applying child tax credit, and in Table 3 for negative tax after applying child tax credit.

Table 2: Calculation of Gini, concentration and progressivity indices for the tax before applying child tax credit

| Method | $G_X$ | $C_{T\|X}$ | $\Pi^K$ | $(\Pi^K / G_X)\cdot 100\%$ |
|---|---|---|---|---|
| Classic indices | 0.36359 | 0.41657 | 0.05297 | 14.5689 |
| CTR-BS method | 0.36359 | 0.41657 | 0.05297 | 14.5689 |
| RSV method | 0.36359 | 0.41657 | 0.05297 | 14.5689 |
| Indices of values relative to tax minimum | 0.36359 | 0.41657 | 0.05297 | 14.5689 |
| Absolute indices | 12444.96 | 14258.06 | 3077. 46 | 14.5689 |

Source: the author.

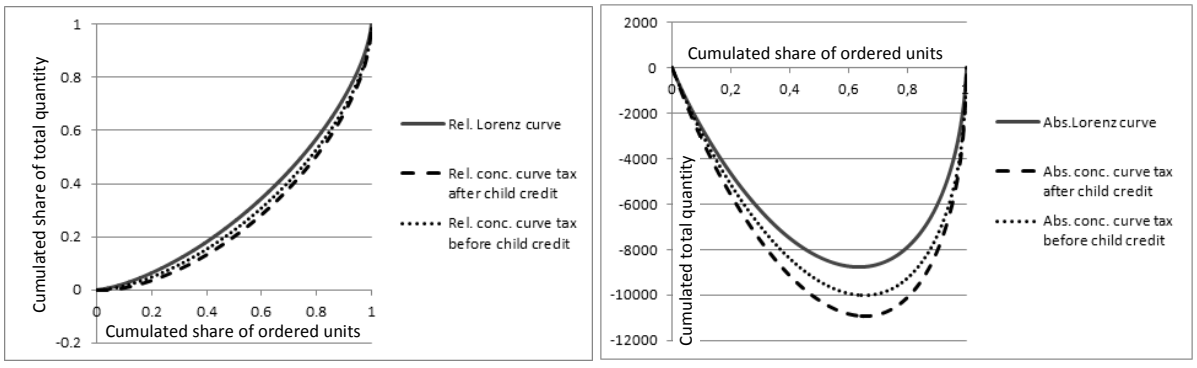
Table 3: Calculation of Gini, concentration and progressivity indices for the tax after applying child tax credit

| Method | $G_X$ | $C_{T\|X}$ | $\Pi^K$ | $(\Pi^K / G_X)\cdot 100\%$ |
|---|---|---|---|---|
| Classic indices | 0.36359 | 0.45351 | 0.08991 | 24.7286 |
| CTR-BS method | 0.36359 | 0.45351 | 0.08991 | 24.7288 |
| RSV method | 0.36359 | 0.45237 | 0.08878 | 24.4168 |
| Indices of values relative to tax minimum | 0.35084 | 0.37285 | 0.02201 | 6.2726 |
| Absolute indices | 12444.96 | 15522.42 | 1813.10 | 24.7286 |

Source: the author.

In aim to illustrate results shown in the Tables 2 and 3, the relative and adjusted absolute concentration curves, which were used to measure the tax progressivity, are presented in Figure 5 together with Lorenz curves.

Figure 5: Relative and absolute Lorenz and concentration curves for the analyzed data




Source: the author.

## 5. Conclusions

The results of the analysis presented in the paper, especially in the Tables 2 and 3, show that different methods of analysis inequality concentration measures adapted to existence of

negative data give similar results, when the amount of negative quantities is not very high as in case of Polish PIT. Only the third method should be abandoned, because it is not useful when progressivity index or other measures are calculated as differences: the influence of subtracting minimum is different for quantities with different means and increased progression is not discovered in case of tax reduced by child tax credit. All other methods give similar results to classic analysis designed for non-negative data. The most perspective method in the opinion of the author is the last one, based on absolute concentration curves, because it can be applied in the same way to data having negative or null mean. Additionally in the author's subjective view, the absolute concentration curves show the most clearly growing progression of tax in case of applying child credits. Further steps of research need investigating properties of described measures in the analysis of artificial data with different amount of negative values and different sign of mean value.

**Acknowledgements**

Special thanks to Walenty Ostasiewicz, Achille Vernizzi, Edyta Mazurek and two anonymous reviewers for their helpful comments. Any errors are solely those of the author.

**References**

[1] AMIEL, Y., COWELL, F., POLOVIN, A. 1996. Inequality among the kibbutzim. In Economica, 1996, vol. 63, iss. 250, supplement S, pp. S63-S85.

[2] BERREBI, Z. M., SILBER, J. 1985. The Gini coefficient and negative income : A comment. In Oxford Economic Papers, 1985, vol. 37, iss. 3, pp. 525-526.

[3] CHEN, C.-N., TSAUR, T.-W., RHAI, T.-S. 1982. The Gini coefficient and negative income. In Oxford Economic Papers, 1982, vol. 34, iss. 3, pp. 473-478.

[4] COWELL, F. A., VAN KERM, P. 2015. Wealth inequality : A survey. In Journal of Economic Surveys, 2015, vol. 29, iss. 4, pp. 671-710.

[5] KAKWANI, N. C. 1977. Measurement of tax progressivity : International comparison. In Economic Journal, 1977, vol. 87, iss. 345, pp. 71-80.

[6] KOLM, S. C. 1976. Unequal inequalities I. In Journal of Economic Theory, 1976, vol. 12, iss. 3, pp. 416-442.

[7] MAZUREK, E. 2015. Losses and profits on the child tax credit in the Polish tax system. In HRONOVÁ, S., VLTAVSKÁ, K. (eds.) Conference Proceedings: 18th International Conference on Applications of Mathematics and Statistics in Economics (AMSE), Jindřichův Hradec, Czech Republic, date: September 2-6, 2015, http://amse2015.cz/doc/Mazurek.pdf.

[8] MOYES, P. 1987. A new concept of Lorenz domination. In Economic Letters, 1987, vol. 23, iss. 2, pp. 203-207.

[9] MOYES, P. 1988. A note on minimally progressive taxation and absolute income inequality. In Social Choice and Welfare, 1988, vol. 5, iss. 2-3, pp. 227-234.

[10] MOYES, P. 2007. An extended Gini approach to inequality measurement. In Journal of Economic Inequality, 2007, vol. 5, iss. 3, pp 279-303.

[11] OSTASIEWICZ, W., OSTASIEWICZ, K. 2015. Towards a science of equity and sustainability. In HRONOVÁ, S., VLTAVSKÁ, K. (eds.) Conference Proceedings: 18th International Conference on Applications of Mathematics and Statistics in Economics (AMSE), Jindřichův Hradec, Czech Republic, date: September 2-6, 2015, http://amse2015.cz/doc/Ostasiewicz_Ostasiewicz.pdf.

[12] R CORE TEAM 2015. R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

[13] RAFFINETTI, E., AIMAR, F. 2016. GiniWegNeg : Computing the Gini Coefficient for Weighted and Negative Attributes. R package version 1.0. http://CRAN.R-project.org/package=GiniWegNeg.

[14] RAFFINETTI, E., SILETTI, E., VERNIZZI, A. 2015. On the Gini coefficient normalization when attributes with negative values are considered. In Statistical Methods and Applications, 2015, vol. 24, iss. 3, pp. 507-521.

[15] SCHUTZ, R. R. 1951. On the measurement of income inequality. In American Economic Review, 1951, vol. 41, iss. 1, pp. 107-122.