STATISTICS IN TRANSITION new series, Spring 2014
Vol. 15, No. 2, pp. 269–282

INCOME INEQUALITY AND INCOME STRATIFICATION IN POLAND

Alina Jędrzejczak[1]

ABSTRACT

Income inequality refers to the degree of income differences among various individuals or segments of a population. When the population has been partitioned into subgroups, according to some criterion, one common application of inequality measures is evaluation of the relationship between inequality in the whole population and inequality in its constituent subgroups in order to work out the within and the between subgroups contributions to the overall inequality. In the paper selected decomposition methods of the well-known Gini concentration ratio were discussed and applied to the analysis of income distribution in Poland. The aim of the analysis was to verify to what extent the inequality in different subpopulations contributes to the overall income inequality in Poland and to what extent their members form distinct segments or strata. To provide the decomposition of the Gini index the population of households was partitioned into several socio-economic groups on the basis of the exclusive or primary source of maintenance. Moreover, the households were divided by economic regions using the Eurostat classification units NUTS 1 as well as by family type defined by the number of children.

Key words: income distribution, income inequality.

1. Introduction

In the analysis of income inequality it may be relevant to assign inequality contributions to various income components (such as labor income or property income) or to various population subgroups associated with socio-economic characteristics of individuals (age, sex, occupation, composition of their household, ethnic groups, etc.). Such an approach can be useful for social policy makers to better understand the influence of various socio-economic determinants on income levels and income inequality. In order to separate the within-groups inequality from the between-groups inequality a decomposable inequality measure has to be used. If the adopted inequality measure is additively

[1] Institute of Statistics and Demography, University of Lodz, Poland. E-mail: jedrzej@uni.lodz.pl.

decomposable, the overall inequality is equal to the sum of the within and between-groups inequality.

The Gini index is a well known and widely used synthetic inequality measure usually expressed in terms of the area under the Lorenz curve. In numerous works on income distribution it is considered the best single measure of income inequality (see e.g.: Morgan, 1962; Gastwirth, 1970), what is mainly due to its statistical properties. In contrast to many other inequality coefficients, measuring only the deviations from the mean and thus interlinking the concept of location with the concept of variability, the Gini index takes into account the income differences between each and every pair of individuals. It has also a clear economic interpretation and thus has been applied in various empirical studies and policy research. On the other hand, being sensitive to both the distribution of income and the distribution of ranks, the Gini index cannot be easily decomposed into two: between-groups and within-groups components. This property can be found a disadvantage of this index which was even claimed decomposable only when the subpopulations do not overlap (see: Shorrocks, 1984). Regardless of these difficulties, for the last 50 years a great effort has been made to specify the conditions under which the decomposition of the Gini coefficient is feasible and many interesting decompositions have been derived. Some of them provide us with the more complex but at the same time more informative tools for income inequality analysis than do many straightforward decompositions of additively decomposable inequality measures.

The first attempts to decompose the Gini index followed the classical Theil's approach and considered only two terms: the within-groups component and the between-groups component, the latter being generally based on the assumption that each individual receives the mean income of his own group. The pioneer Gini index decomposition by subgroups is due to Soltow (1960) who analyzed the effects of changes in education, age and occupation on income distribution. The first Gini index decomposition encompassing comparisons between pairs of subgroups is due to Bhattacharya and Mahalanobis (1967); actually the decomposition proposed by the authors refers first to the Gini mean difference Δ. The decomposition is based on a priori definition of the between-groups component, being the Gini mean difference evaluated among the subgroup means, and leaves the within term to be obtained as a residual.

Both the decompositions mentioned above were rather inadequate as they ignored the existence of overlapping as well as different variances and asymmetries of income distributions in subpopulations. In fact, when the groups ranges overlap the third component called "crossover term" or "interaction" arises, being rather difficult to interpret. The interaction term can be viewed as a measure of income stratification or the degree to which the incomes of different social groups cluster.

An interesting three-term decomposition and interpretation of the Gini coefficient was proposed by Pyatt (1976) in a game theory framework. Following the Pyatt idea, the Gini index can be perceived as an average gain to be expected if an individual had a choice between his own income or any other income selected at random from the population of income receivers. Pyatt split the Gini index into the sum of three non-negative terms: the first depends on the differences in mean incomes between subgroups and remains the only positive term in the special case when there is no variation within subgroups, the other two terms both depend on variation within subgroups. In particular, the second one depends on the Gini indices evaluated within each subgroup and the third term vanishes in the case when subgroups income ranges do not overlap, otherwise it is positive and measures the degree of overlapping. An analogous approach, based on matrix algebra, can be found in Silber (1989); the author decomposes the Gini index into the sum of the within, between and interaction terms giving a clear and intuitive interpretation to the latter in terms of individuals ranking. That “third component” was also discussed by Mehran (1975), Mookherjee and Shorrocks (1982), Yitzhaki and Lerman (1991), Deutsch and Silber (1999), to name only a few, what resulted in numerous interesting decomposition formulas. Some of them are computationally cumbersome and it is not always clear what meaningful interpretation each of the components has. Mehran defined “the third term” as interaction “*interpreted as a measure of income domination of one subgroup over the other apart from the differences between their mean incomes*”. Yitzhaki and Lerman (1991), intended from a sociological point of view, proposed a decomposition of the Gini index into the sum of a within term, a between term, and a third term that accounts for subgroups stratification understood as “*a group’s isolation from members of other groups*”. The within- and between-group terms considered by the authors were based on the covariance formula so they are differently defined with respect to the ones considered above.

The most widespread approach to the decomposition of the Gini index that gives an important contribution to the understanding of the overlapping term was proposed by Dagum (1997). It introduces the concept of economic distance between distributions and relative economic affluence (REA) as an important element in the Gini index decomposition by subpopulation groups.

The objective of the paper is to discuss the most interesting decomposition procedures proposed by Dagum (1997) and Yitzhaki and Lerman (1991) and then apply them to the analysis of income inequality in Poland. The aim of the analysis was to verify to what extent the inequality in different subpopulations contributes to the overall income inequality in Poland and whether their members form distinct segments or strata.

2. The Gini index decomposition by subpopulations

The Gini index of inequality is usually defined by means of a geometric formula since it can be expressed as twice the area between the Lorenz curve and the straight line called the line of equal shares. The Gini index can also be seen as a relative dispersion measure when expressed by means of the mean difference

Δ - a dispersion measure which is defined as the average absolute difference between all possible pairs of observations. This concept can be called a statistical approach and was introduced by Gini (1912). It was subsequently used by many authors to derive various Gini index decompositions but the most widespread decomposition by subpopulations was undoubtedly proposed by Dagum (1997).

The starting point for this decomposition was the Gini index formula based on the Gini mean difference extended to the case of a population divided into *k* subpopulations (groups):

$$G = \frac{\Delta}{2\overline{Y}} = \sum_{r=1}^{n}\sum_{i=1}^{n} | Y_i - Y_r | \Big/ 2n^2\overline{Y} = \sum_{j=1}^{k}\sum_{h=1}^{k}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h} | y_{ji} - y_{hr} | \Big/ 2n^2\overline{y} \tag{1}$$

The Gini index expressed in terms of the Gini mean difference can also be generalized for a two-populations case, measuring the between-populations (or intra-groups) inequality. Thus, the extended Gini index between groups *j* and *h* can be written as follows:

$$G_{jh} = \frac{\Delta_{jh}}{\overline{Y}_j + \overline{Y}_h} = \frac{1}{\overline{Y}_j + \overline{Y}_h}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h} | y_{ji} - y_{hr} | \big/ n_j n_h \tag{2}$$

where: Δ_{jh} - mean difference modified for two income distributions.

Dagum (1997) proved that the Gini ratio *G* for a population of economic units partitioned into *k* subpopulations n_j ($j = 1, \ldots, k$) can be expressed as the weighted sum of the extended Gini ratios weighted by the products of the *j*-th group population share p_j and the *h*-th group income share s_h:

$$G = \sum_j \sum_h G_{jh} p_j s_h \tag{3}$$

Using the symmetry properties of G_{jh} and Δ_{jh} and the equation (3), the Gini index can be decomposed into two elements: the within G_w and gross-between G_{gb} inequality (Dagum, 1997):

$$G = \sum_{j=1}^{k} G_{jj} p_j s_j + \sum_{j=2}^{k}\sum_{h=1}^{j-1} G_{jh}(p_j s_h + p_h s_j) = G = G_w + G_{gb} \tag{4}$$

where: $G_{jj} = \frac{\Delta}{2\overline{y}_j} = \frac{1}{2\overline{y}_j}\sum_{r=1}^{n_h}\sum_{i=1}^{n_j} | y_{ji} - y_{jr} | \big/ n_j^2$ is the Gini index for the subpopulation *j* $\overline{y}_j$ - mean income in group *j*, n_j - frequency in group *j*.

As it can be easily noticed the Gini index provides an unusual “between-group” component. It measures the income inequality between each and every pair of subpopulations, whereas entropy and most of between-groups inequality measures yield only the income inequalities between the subpopulation means. The first component of the decomposition given by the formula (4) (G_w) describes the *contribution of the Gini inequality within subpopulations* to the total inequality of a population described by the Gini ratio G. The second component (G_{gb}) measures the *gross contribution of the extended Gini inequality between subpopulations* to the total Gini G. This component depends on the differences between subpopulations coming from both: *differences in mean* income levels and *differences in shape* (the populations differ in variance and asymmetry which implies that they have different inequality measures).

The income differences between the elements coming from various subgroups can be of the same or of opposite sign as the deviation in their corresponding means.

The interpretation of G_{gb} given above suggests the further decomposition of the Gini index by subgroups. The contribution of gross between-group inequality can be divided into two separate parts: the first one consistent with the differences between the means and the remaining part called transvariation:

$$G_{gb} = \sum_{j=2}^{k}\sum_{h=1}^{j-1} G_{jh}(p_j s_h + p_h s_j) D_{jh} + \sum_{j=2}^{k}\sum_{h=1}^{j-1} G_{jh}(p_j s_h + p_h s_j)(1 - D_{jh}) = G_b + G_t \qquad (5)$$

G_b – the contribution of net between-groups inequality to the Gini index,
G_t – the contribution of "transvariation",
D_{jh} – "economic distance" ratio (Dagum, 1980).

The concept of transvariation (*transvariazione*) was originally introduced by Gini (1916) and it plays a crucial role in the Gini index decomposition by population subgroups. Transvariation between two populations exists when at least one income difference between individuals belonging to different groups has the sign opposite to the sign of the difference between their means. Obviously, the idea of transvariation is similar to the concept of distribution overlapping. The probability of transvariation can be simply defined (Gini, 1916) as the ratio of the actual number of transvarying pairs to its maximum. It takes values in the interval [0,1] and the more the two groups overlap the greater value it takes. Intensity of transvariation accounts not only for the frequency but also for the amount of income differences. The term Djh (eq. 5) called economic distance ratio or REA (relative economic affluence) is related to the normalized intensity of transvariation which is simply 1-Djh , and can be regarded as the measure of relative economic affluence of the j-th subpopulation with respect to the h-th subpopulation. It can be defined as the weighted sum of the income differences yji –yhr for all the members belonging to the population j-th with incomes greater than the income of all the members belonging to the population *h*-th, given that $\overline{Y}_j > \overline{Y}_h$ (for details see: Dagum, 1980).

As pointed out in Monti (2007), it is easy to verify that G_w, G_b and G_t of the Dagum decomposition (eq. 4 and eq. 5) equal, respectively, the within, the between and the interaction term of Mookherjee and Shorrocks decomposition and are also equivalent to Mehran's decomposition. It can be noted that the Dagum between-groups inequality (4) can be obtained without the rigorous assumption about equally distributed income groups. Moreover, it is worth mentioning that only the Dagum decomposition shows clearly how the overlapping term is connected both with between-groups and within-group inequality.

The inequality decomposition proposed by Yitzhaki and Lerman (1991) is based on the covariance formula, presented by the same authors (Lerman, Yitzhaki, 1985), where the Gini index is expressed in terms of twice the covariance between income and its rank divided by the overall mean income. Their decomposition encompasses an index of stratification that highlights the distinction between social stratification and inequality. It captures the extent to which population subgroups occupy distinct strata within an overall distribution. For the *i*-th subpopulation the index of stratification has the following form:

$$Q_i = \frac{\mathrm{cov}_i[F_i(y) - F_{n-i}(y), y]}{\mathrm{cov}_i[F_i(y), y]} \tag{6}$$

where: $cov_i[F_i(y) - F_{n-i}(y), y]$ – covariance over group i between y and the difference between the ranking of a member of group i in his own group and the re-ranking he would have in the rest of the population,

$cov_i[F_i(y), y]$ - covariance over group i between y and its own ranking in group i.

The index of stratification given by (6) measures how members of a group differ from members of other groups. In this context stratification can be understood as "a group's isolation from members of other groups" (Yitzhaki, Lerman 1991). The index (6) has the following properties, making it sensitive to stratification of particular groups over an overall population:

- it measures the level of stratification for each group separately, taking into consideration the relation of its ranking in comparison with the rest of the population;
- Q_i declines when the number of the members of other groups being in the range of i increases;
- Q_i takes values from the interval <–1,1>. If $Q_i = 1$, a group i forms a perfect stratum - no members of other groups fall within its range of income. If $Q_i = 0$, a group i does not form a stratum at all - the ranking of all individuals within this group is identical to their ranking within the overall population (the groups completely overlap). $Q = -1$ in an extreme case when a group i is not well defined as being composed of two perfect strata placed at the tails of the distribution;

– given a number of the members of other groups who fall in the range of a group *i*, Q_i will be lower the closer the members of these groups are to the mean of *i*.

Income stratification is highly related to income inequality and can be a starting point to inequality decomposition by subpopulation groups. In general, high within-group inequality is likely to reduce a group stratification because it often increases overlapping of a group with other groups. On the other hand, high between-group inequality is likely to increase stratification by making the subpopulations more isolated from each other. Complicated connections between within-group inequality, between-group inequality and stratification can be revealed in detail by an unified framework given by a decomposition formula of Lerman and Yitzhaki (1991):

$$G = \sum_i s_i G_i + \sum_i s_i (p_i - 1) G_i Q_i + \sum_i \frac{2\operatorname{cov}[\bar{y}_i, \bar{F}_i(y)]}{\bar{y}} \qquad (7)$$

where: $\bar{F}_i(y)$ – group *i*'s average rank.

The first component represents within-group inequality, the second component reflects the impact of stratification, described as intra-group inequality in overall ranks, while the third component accounts for the between-group inequality. Changes in income distribution may affect only one component of (7) or may have influence on all of them. High stratification implies low variability of ranks so the increases in group stratification exert negative impact on inequality. The between-group inequality is expressed as the between-group Gini index calculated on the basis of covariance between each mean income of a group and the average rank. As the authors point out, it is similar, but not identical to the between-group terms presented in Pyatt (1976), Mookherjee and Shorrocks (1982) and Silber (1989). The substantial difference is in the way the group ranks are established: in Lerman and Yitzhaki (1991) the ranking is obtained by averaging each ranking of observation within each subpopulation, while for the remaining authors it is simply the ranking of mean incomes. It is worth mentioning that when there is no overlapping between groups, all the methods yield the same results.

3. Application

The methods discussed above were applied to the analysis of income inequality in Poland by socio-economic groups, regions and family types. The basis for the calculations was micro data coming from the Household Budget Survey (HBS) conducted by Central Statistical Office in 2009. The data obtained from the HBS allow for the detailed analysis of the living conditions in Poland, being the basic source of information on the revenues and expenditure of the population. In 2009 the randomly selected sample covered 37,302 households, i.e. approximately 0.3% of the total number of households. The adopted sampling scheme was geographically stratified and two-stage one with different selection probability at the first stage. In the estimation of inequality measures and their decomposition the survey weights based on inverse inclusion probabilities were taken into consideration. In order to maintain the relation between the structure of the surveyed population and the socio-demographic structure of the total population, the data obtained from the HBS were weighted with the structure of households by the number of persons and class of locality coming from Population and Housing Census 2002.

The inequality analysis was conducted after separately dividing the overall sample: by region NUTS 1 constructed according to the Eurostat classification, by family type classified according to the number of children, and by socio-economic group established on the basis of the exclusive or primary source of maintenance. The variable of interest was household available income that can be considered the basic characteristic of its economic condition. It is defined as a sum of households' current incomes from various sources reduced by prepayments on personal income tax made on behalf of a tax payer by tax-remitter (this is the case of income derived from hired work and social security benefits and other social benefits); by tax on income from property; taxes paid by self-employed persons (including professionals and individual farmers), and by social security and health insurance premiums. To avoid interpretation problems, rare negative incomes were removed from the original sample.

Table 1 describes in detail the results of income inequality decomposition by socio- economic groups while tables 2 and 3 present the corresponding calculations outcome for the population divided by region and family type, respectively. To allow comparing the conditions of households of different sizes and different demographic structures, the square root scale, popular in recent OECD publications, was applied in the paper (table 3a). All the tables present statistical characteristics of household available income by population groups as well as the final results of inequality decomposition with respect to these groups. In particular, the within-groups, between-groups and "overlapping" components are reported for both Dagum (D) and Yitzhaki-Lerman (Y-L) approach (eq. (4), (5) ,(7)). As it has been mentioned above, these decompositions represent completely different concepts and thus provide us with inequality contributions that can be the basis of income inequality analysis from different perspectives. However, the main interest of this paper is groups overlapping and stratification. The overlapping component in the Dagum decomposition (called transvariation) is based on the relative economic affluence of one subpopulation with respect to another while the "third term" of Y-L method is based on ranking rather than income differences, and can only be regarded as a measure of groups separation. Similarly, the between-group component of the Dagum approach is based on income differences for each and every pair of households in contrast to the Y-L approach where only group means are considered. It results in higher sensitivity of the Dagum decomposition to changes in grouping factors, while the Y-L decomposition is by construction dominated by the within-group component (see. Tables 1-3).

Table 1. Decomposition of income inequality in Poland by socio-economic group

Measure	Socio-economic group					Total
	Employees	Farmers	Self-employed	Pensioners	Unearned sources	
Mean income $\bar{y}_i$ [1000 PLN]	3.781	4.556	4.738	2.108	1.695	3.186
Population proportion p_t	0.491	0.038	0.069	0.361	0.041	1
Income proportion s_i	0.583	0.054	0.103	0.239	0.021	1
Gini index G_i	0.293	0.483	0.319	0.306	0.370	0.352
Stratification index Q_i	0.313	–0.038	0.269	0.189	0.083	
Within-groups term (Y–L)	0.171	0.026	0.033	0.073	0.008	**0.311**
Between-groups term (Y–L)						**0.085**
Stratification term (Y–L)						**–0.044**
Within-groups term (D)	0.084	0.001	0.002	0.026	0.000	**0.114**
Between-groups term (D)						**0.154**
Transvariation (overlapping term) (D)						**0.085**

Source: Author's calculations.

Table 2. Decomposition of income inequality in Poland by region

Measure	Region of Poland						Total
	Central	Southern	Eastern	North-western	South-western	Northern	
Mean income $\bar{y}_i$ [1000 PLN]	3.554	3.093	2.861	3.227	3.159	3.122	3.186
Population proportion p_t	0.218	0.208	0.168	0.154	0.107	0.145	1
Income proportion s_i	0.243	0.202	0.151	0.156	0.106	0.142	1
Gini index G_i	0.381	0.318	0.355	0.342	0.352	0.348	0.352
Stratification index Q_i	–0.025	0.054	–0.023	0.031	–0.001	0.005	
Within-groups term (Y–L)	0.093	0.064	0.054	0.053	0.037	0.049	**0.351**
Between-groups term (Y–L)							**0.006**
Stratification term (Y–L)							**–0.003**
Within-groups term (D)	0.020	0.013	0.009	0.008	0.004	0.007	**0.062**
Between-groups term (D)							**0.042**
Transvariation (overlapping term) (D)							**0.249**

Source: Author's calculations

Table 3. Decomposition of income inequality in Poland by family type

Measure	Family type (number of children)						Total
	0	1	2	3	4	5…	
Mean income $\bar{y}_i$ [1000 PLN]	2.751	3.920	4.013	3.685	3.471	3.667	3.186
Population proportion p_t	0.643	0.183	0.126	0.035	0.009	0.004	1
Income proportion s_i	0.559	0.226	0.160	0.041	0.009	0.005	1
Gini index G_i	0.361	0.313	0.325	0.329	0.294	0.314	0.352
Stratification index Q_i	–0.028	0.169	0.165	0.108	0.104	0.107	
Within-groups term (Y–L)	0.201	0.071	0.052	0.013	0.003	0.002	**0.342**
Between-groups term (Y–L)							**0.027**
Stratification term (Y–L)							**–0.017**
Within groups term (D)	0.129	0.013	0.006	0.000	0.000	0.000	**0.150**
Between-groups term (D)							**0.071**
Transvariation (overlapping term) (D)							**0.131**

Source: Author's calculations.

Table 3a. Decomposition of income inequality in Poland by family type (equivalised income)

Measure	Family type (number of children)						Total
	0	1	2	3	4	5…	
Mean income $\bar{y}_i$ [1000 PLN]	1.947	2.066	1.910	1.563	1.330	1.260	1.942
Population proportion p_t	0.643	0.183	0.126	0.035	0.009	0.004	1
Income proportion s_i	0.645	0.194	0.124	0.028	0.006	0.003	1
Gini index G_i	0.308	0.308	0.322	0.319	0.282	0.293	0.312
Stratification index Q_i	0.034	0.018	-0. 033	-0.058	0.064	0.107	
Within-groups term (Y–L)	0.198	0.060	0.040	0.009	0.002	0.001	**0.310**
Between-groups term (Y–L)							**0.027**
Stratification term (Y–L)							**–0.002**
Within-groups term (D)	0.128	0.011	0.005	0.000	0.000	0.000	**0.144**
Between-groups term (D)							**0.021**
Transvariation (overlapping term) (D)							**0.147**

Source: Author's calculations.

The overall income inequality in Poland in 2009, measured by means of the Gini index and estimated on the basis of the Polish HBS, was 0.352 (for equivalent income G=0.312). These values confirm a high level of income inequality in Poland as compared with other European countries - according to EU-SILC in 2009 the Gini index calculated for equivalent disposable net income was at the level of 0.314 and in 2011 at the level 0.311, what was still above the EU average. It is worth mentioning that one can observe substantial differences in the values of inequality measures while using different data sources. The discrepancies between the values of the Gini index obtained on the basis of HBS, EU-SILC and Social Diagnosis for the same category of income may come from different sample sizes, different sampling designs and what seems the most important from the method of dealing with non-response. For example, the methodology of EU-SILC includes a requirement for the imputation of the missing income, what can lead to the underestimation of inequality measures and their standard errors. Moreover, one can run into difficulties while trying to compare the results over time - EU-SILC and Social Diagnosis are relatively new surveys and their implementation has been disturbed by many methodological changes. On the contrary, the Household Budget Survey is relatively stable and has the largest sample size, but even such a sample can be insufficient to provide reliable estimates in some divisions (see: Jędrzejczak, Kubacki, 2013).

The impact of the number of children on the distribution of household available income is presented in table 3. Applying the Dagum decomposition, the overall Gini index is due to within-group (43%) and overlapping (37%) components, while the contribution of the between-group term was found to be rather small (20%). The families without children form an untypical group ($Q_0<0$), which in fact consists of two smaller ones differing in average income level: a group of individuals (mainly retirees) and a group of couples without children. The significant stratification emerges only for the households with 1 or 2 children ($Q_1=0,169$; $Q_2=0,165$), identifying them as relatively similar within the groups and different from the outside. This result, however, can be misleading for two reasons. Firstly, the stratification indices Q_i proposed by Yitzhaki and Lerman ignore group sizes and can be negligible even for relatively separated groups when they are sufficiently small. Secondly, to compare subpopulations constructed on the basis of the number of children the equivalised income should be considered rather than the nominal one. After the transformation of available incomes with respect to household composition, the stratification indices, except for the first group, were found to be close to 0 (table 3a). Nevertheless, very high economic distance ratios D_{jh} were observed between small but the poorest groups of households (with 4 and 5 or more children) and the wealthiest group of families possessing only one child. They both exceed 60% so the families possessing only one child are 60% more affluent than the families with 4 and more children. The economic distance ratios D_{jh} consider pair comparisons between groups so they better detect income differences between various subpopulations than do Q indices.

The stratification and between-group inequality is much higher when the breakdown by socio-economic group is considered (table 1). The decomposition presented in table 1 takes into account the splitting up into households of self-employed, households of employees (managers, office workers, blue-collar workers, school teachers, etc.), households of not employed (retirees and pensioners) and households of other not employed (mainly unemployed). The households of farmers constitute a separate group.

Using the Dagum decomposition, the total income inequality in Poland by socio-economic group is dominated by between-group term that accounts for 44% of the overall Gini index. This result coincides with serious stratification indices, which were observed for several socio-economic groups and play an important role in Y-L decomposition. The within-group component (32%) reflects the inner polarization of the groups what gives rise to remarkable differentials in average income between managers and blue-collar workers within the group of *employees*, between entrepreneurs and the others within the group of *self-employed* or between retirees and pensioners within the fourth group. The households of *self-employed* are the wealthiest group, the one with the highest average income, but the group representing the highest level of inequality are *farmers* (G=0.48). The households of *employees* constitute a group with the highest share (24%) in the overall Gini index what is mainly due to its size and income share. The contribution of the overlapping component measured by transvariation is rather small (24%), contrary to high stratification indices for socio-economic groups except *farmers* and *unearned sources*. The negative value of the stratification index Q (and high G) observed for *farmers* suggests that this group is nonhomogeneous, being composed of the households that are not of the same kind (small and very large farms).

The impact of regional differences on income inequality in Poland can be observed in table 2. Contrary to family types and socio-economic groups, regional differences contribute slightly to the overall value of the Gini index. The between-group component accounts for only 12% of the overall income inequality. The Gini ratios and means within regions do not differ significantly so the contributions of particular subpopulations to the overall inequality are determined mainly by their sizes. The substantial contribution of transvariation, equal to 71% of the overall Gini index, is an evidence of notable overlapping of income distributions for NUTS 1 regions in Poland (see also: Jędrzejczak 2010).

4. Concluding remarks

Decomposition of the Gini index can be useful for social policy-makers in assessing the contributions of between-groups and within-groups inequalities to the overall inequality of a population. It can also be helpful in stratification and market segmentation by including the concept of overlapping.

The most widespread approach to the group decomposition of the Gini index was given by Dagum and it is based on the concepts of economic distance between distributions and relative economic affluence. It takes into account different variances and asymmetries of income distributions in subpopulations and gives an important contribution to the understanding of the overlapping term.

The Gini index decomposition proposed by Yitzhaki and Lerman encompasses the index of stratification by linking social stratification with inequality. It can be applied to assess isolation of social groups expressed in terms of income.

Estimation results obtained on the basis of Polish HBS revealed high discrepancies between socio-economic groups of households defined on the basis of primary source of maintenance, whereas regional differences were found to be relatively small and to contribute slightly to overall income inequality in Poland. Extremely large income differences were observed between some household groups differentiated by the number of children. One should also be conscious that the estimation results can be biased mainly because of a high non-response rate being an immanent feature of household budgets surveys all over the world.

Acknowledgements

The Author would like to thank the anonymous reviewer for her/his valuable comments and constructive suggestions which have improved the quality of the paper.

REFERENCES

BHATTACHARYA, N., MAHALANOBIS B., (1967). Regional disparities in household consumption in India, Journal of the American Statistical Association 62 (317), 143–161.

DAGUM, C., (1980). Inequality Measures Between Income Distributions with Application. Econometrica 48(7), 1791–1803.

DAGUM, C., (1997). A New Approach to the Decomposition of the Gini Income Inequality Ratio, Empirical Economics 22(4), 515–531.

DEUTSCH, I., SILBER J., (1999). Inequality Decomposition by Population Subgroups and the Analysis of Interdistributional Inequality, in: J. Silber, Handbook of Income Inequality Measurement, 363–397.

GASTWIRTH, J. L., (1972). The Estimation of the Lorenz Curve and the Gini Index, Review of Economics and Statistics 54(3), 306–316.

GINI, C., (1912). Variabilita e Mutabilita, Bologna, Tipografia di Pado Cuppini.

GINI, C., (1916). Il Concetto di Transvariazione e le sue Primi Applicazioni, Giornale degli Economisti e la Rivista Statistica, [in:] Gini (1959), 21–44.

GINI, C., (1959). Memorie di Metodologia Statistica, vol. II. Libreria Goliardica, Roma.

JĘDRZEJCZAK, A., (2010). Decomposition Analysis of Income Inequality in Poland by Subpopulations and Factor Components, Argumenta Oeconomica 1(24), 109–123.

JĘDRZEJCZAK, A., KUBACKI, J., (2013). Estimation of Income Inequality and the Poverty in Poland by Region and Family Type, Statistics in Transition 14(3), 359-378.

MEHRAN, F., (1975). A Statistical Analysis on Income Inequality Based on Decomposition of the Gini Index. Proceedings of the 40th Session of ISI.

MOOKHERJEE, D., SHORROCKS, A., (1982). A Decomposition Analysis of the Trend in UK Income Inequality, The Economic Journal 92(368), 886–902.

MONTI, M., (2007). Note on the Dagum Decomposition of the Gini Inequality Index, Università Degli Studi di Milano Working Papers 2007–16.

MORGAN, J., (1962). The Anatomy of Income Distribution, Review of Economics and Statistics 44(3), 270–282.

PYATT, G., (1976). On the Interpretation and Disaggregation of Gini Coefficient, The Economic Journal 86(342), 243–255.

RADAELLI, P., (2010). On the Decomposition by Subgroups of the Gini and Zenga‘s Uniformity and Inequality Indexes, International Statistical Review, 78(1), 81–101.

SHORROCKS, A., (1984). Inequality Decomposition by Population Subgroups, Econometrica 52(6), 1369–1385.

SOLTOW, L., (1960). The Distribution of Income Related to Changes in the Distributions of Education, Age, and Occupation, The Review of Economics and Statistics 42(4), 450–453.

SILBER, J., (1989). Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality, The Review of Economics and Statistics 71(2), 107–115.

VERNIZZI, A., (2009). Applying the Hadamard Product to Decompose Gini, Concentration , Redistribution and Re-ranking Indices, Statistics in Transition 10 (3), 505–524.

YITZHAKI, S., (1994). Economic Distance and Overlapping of Distributions, Journal of Econometrics 61(1), 147–159.

YITZHAKI, S., LERMAN, R., (1984). A Note on the Calculation and Interpretation of the Gini Index, Economic Letters 15(3-4), 363–369.

YITZHAKI, S., LERMAN, R., (1991). Income Stratification and Income Inequality, Review of Income and Wealth 37(3), 313–329.