STATISTICS IN TRANSITION new series, Autumn 2013
Vol. 14, No. 3, pp. 359–378

ESTIMATION OF INCOME INEQUALITY AND THE POVERTY RATE IN POLAND, BY REGION AND FAMILY TYPE

Alina Jędrzejczak[1], **Jan Kubacki**[2]

ABSTRACT

High income inequality can be a source of serious socio-economic problems, such as increasing poverty, social stratification and polarization. Periods of pronounced economic growth or recession may impact different groups of earners differently. Growth may not be shared equally and economic crises may further widen gaps between the wealthiest and poorest sectors. Poverty affects all ages but children are disproportionately affected by it. The reliable inequality and poverty analysis of both total population of households and subpopulations by various family types can be a helpful piece of information for economists and social policy makers. The main objective of the paper was to present some income inequality and poverty estimates with the application to the Polish data coming from the Household Budget Survey. Besides direct estimation methods, the model based approach was taken into regard. Standard errors of estimates were also considered in the paper.

Key words: income inequality, poverty, variance estimation, small area statistics.

1. Introduction

The range of survey data analysis has expanded enormously over time in response to growing demands of policy makers. Recently, the demand for estimates at a small level of aggregation has increased, in contrast to national estimates that were commonly used in the past. Since income inequality in Poland increased significantly in the period of transformation from the centrally planned to the market economy, reliable inequality and poverty analysis of the total population of households and subpopulations by various family types can provide helpful information for economists and social policy makers.

[1] Institute of Statistics and Demography, University of Łódź; Centre of Mathematical Statistics, Statistical Office in Łódź. E-mail: jedrzej@uni.lodz.pl.
[2] Centre of Mathematical Statistics, Statistical Office in Łódź. E-mail: j.kubacki@stat.gov.pl.

In the paper, some direct and indirect model-based estimation methods for income inequality and poverty parameters are presented and applied. Income inequality is measured by the Gini and Zenga indices. Among the indicators of poverty we consider: at risk of poverty rate, poverty gap and poverty severity with special attention paid to the estimation of their standard errors. The estimates of inequality measures are produced only for large domains (regions or family types considered separately) using direct estimation, while poverty related measures are also calculated for small domains that require small area estimation. For subsets of the Polish population cross-classified by region and family type small area estimators based on linear mixed models are used.

Measures of income inequality and poverty are presented in Section 2. Section 3 provides a brief survey of direct variance estimation methods, while in Section 4 the outline of EBLUP theory is presented. Some empirical applications based on Polish HBS data are included in Section 5.

2. Measures of income inequality and poverty

Income inequality refers to the degree of difference in earnings among various individuals or segments of a population. Measures of inequality, also called concentration coefficients, are widely used to study income, welfare and poverty issues. They can also be helpful in analyzing the efficiency of a tax policy or in measuring the level of social stratification and polarization. They are most frequently applied to dynamic comparisons, i.e. comparing inequality across time. Among numerous inequality measures, the Gini and Zenga coefficients are of the greatest importance. The Gini concentration coefficient is the most widely used measure of income inequality, mainly because of its clear economic interpretation. The Zenga „point concentration" measure based on the Zenga curve has recently received some attention in the literature.

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

$$G = 2\int_0^1 (p - L(p))dp \tag{1}$$

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

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

where μ denotes the expected value of a random variable Y and $F^{-1}(p)$ is the p^{th} quantile.

One can estimate the value of the Gini index from the survey data using the following formula:

$$\hat{G} = \frac{2\sum_{i=1}^{n}(w_i y_{(i)} \sum_{j=1}^{i} w_j) - \sum_{i=1}^{n} w_i y_{(i)}}{(\sum_{i=1}^{n} w_i)\sum_{i=1}^{n} w_i y_{(i)}} - 1 \ , \tag{3}$$

where: $y_{(i)}$ – household incomes in a non-descending order, w_i- survey weight for *i*-th economic unit, $\sum_{j=1}^{i} w_j$ - rank of *i*-th economic unit in *n*-element sample.

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

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

$$Z_p = [y_p^* - y_p] / y_p^*, \tag{4}$$

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

$$y_p^* = Q^{-1}(p). \tag{5}$$

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

Zenga synthetic inequality index can be expressed as the area below the Zenga curve (4), and is defined as simple arithmetic mean of point concentration measures Z_p, $p \in \langle 0.1 \rangle$:

$$Z = \int_0^1 Z_p dp \ . \tag{6}$$

The commonly used nonparametric estimator of the Zenga index (6) was introduced by Aly and Hervas (1999) and can be expressed by the following equation:

$$\hat{Z} = 1 - \frac{1}{n\bar{y}} \left\{ y_{1:n} + \sum_{j=1}^{n-1} y_{\left\langle \frac{\sum_{i=1}^{j} y_{i:n}}{\bar{y}} \right\rangle :n} \right\} , \tag{7}$$

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

The poverty measures are statistical functions which translate the comparison of the indicator of well-being and the poverty line, made for each household, into one aggregate number for the population as a whole or a population sub-group. Since the publication of Sen (1976) article on the axiomatic approach to the measurement of poverty, several indices of poverty have been developed that make use of the three basic poverty indicators (Panek, 2008). The most popular poverty measure is *headcount ratio* also called *at-risk-of-poverty rate ARPR*. It represents the share of the population whose equivalent income or consumption is below the poverty line:

$$H = \frac{n_p}{n} 100, \tag{8}$$

where: n_p- number of the poor, n- total number of households.

Poverty gap index provides information regarding the distance of households from the poverty line. This measure captures the mean aggregate income or consumption shortfall relative to the poverty line across the whole population. It is obtained by adding up all the shortfalls of the poor and dividing the total by the population size:

$$PG = \frac{1}{n} \sum_{i=1}^{n_p} \left(\frac{y^* - y_i}{y^*} \right), \tag{9}$$

where: y^* denotes the poverty line (poverty threshold). The poverty gap (9) can be used as a measure of the minimum amount that one would have to transfer to the poor under perfect targeting, i.e. each poor person getting exactly the amount he/she needs to be lifted out of poverty so that they are all brought out of poverty. By replacing the number of households n by the number of the poor n_p in the formula (9), we obtain the alternative poverty gap index:

$$PG_p = \frac{1}{n_p} \sum_{i=1}^{n_p} \left(\frac{y^* - y_i}{y^*} \right). \tag{10}$$

Poverty severity index (squared poverty gap) takes into account not only the distance separating the poor from the poverty line (the poverty gap), but also the inequality among the poor. That is, higher weights are placed on those households which are further away from the poverty line.

$$PS_p = \frac{1}{n_p} \sum_{i=1}^{n_p} \left(\frac{y^* - y_i}{y^*} \right)^2. \tag{11}$$

According to their definitions, the headcount index (ARPR), the poverty gap index and the poverty severity index (Panek, 2008) can be expressed as ratio

estimators. Thus, the precision estimation algorithms can be similar to the algorithm for a ratio estimate. The headcount index estimator can be expressed as follows:

$$\hat{H} = \frac{\sum_{i \in U} I_i w_i}{\sum_{i \in U} w_i}, \tag{12}$$

while the estimator of poverty gap index given by (10) takes the form:

$$\hat{PG}_p = \frac{\sum_{i \in U_p} ((y^* - y_i)/y^*) w_i}{\sum_{i \in U_p} w_i}, \tag{13}$$

where: I_i – indicator function taking value 1 when *i*-th household equivalent income is below a poverty line, and taking value 0 in the opposite situation, w_i – survey weight for *i*-th economic unit, U_p denotes the poor families population (or subpopulation).

3. Methods of variance estimation

The precision of an estimator is usually discussed in terms of its variance or standard error. When the standard errors of inequality and poverty measures are large, many conclusions about the comparisons over time and between groups may not be warranted. For most income concentration measures, the Gini and Zenga indices included, explicit variance estimators are theoretically complicated, i.e. it is hard to derive general mathematical formulas for nonlinear statistics, especially when the sampling design is complex. Also, most widely used poverty statistics are nonlinear functions of sampling observations so their standard errors are rather difficult to obtain and have been rarely reported in practice. To solve this problem, some special approximate techniques for variance estimation can be used. They include: Taylor linearization technique, random groups method, jackknife, bootstrap, balanced half samples, also called balanced repeated replication BRR. (Wolter, 2003; Särndal et al., 1997).

In the context of inequality measures, the Taylor linearization, the bootstrap and the parametric approach based on a theoretical income distribution model are the methods of variance estimation most often used (Jędrzejczak, 2011); while standard errors of poverty statistics are usually estimated by means of the bootstrap and balance repeated replication. An interesting outline of the variance estimations methods for the Gini index was offered by Langel and Tillé (2013).

The parametric approach uses a model-based variance with respect to hypothesized data generating process, provided that an empirical income distribution can be approximated by a theoretical model described by a probability density function $f(y,\boldsymbol{\theta})$. Applying the maximum likelihood (ML)

theory, the estimators obtained are asymptotically unbiased and normally distributed with variances given by the Cramer-Rao bound. Let us assume that an inequality measure of interest can be expressed as a function $g(\boldsymbol{\theta})$ of the model parameters $\boldsymbol{\theta}$. The variance of the ML estimator of an inequality measure $g(\boldsymbol{\theta})$ takes the form:

$$\mathrm{D}^2[\mathrm{g}(\hat{\boldsymbol{\theta}})] = \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]^T \mathbf{I}_\theta^{-1} \left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right], \tag{14}$$

where: $\mathbf{I}_{\boldsymbol{\theta}}$ denotes the Fisher information matrix.

The estimator of the variance (14) can be obtained by replacing the unknown parameter values $\boldsymbol{\theta}$ by their large sample ML estimates $\hat{\boldsymbol{\theta}}$. It preserves the asymptotic properties of maximum likelihood estimators (Zehna, 1966). The parametric approach can be very effective for large samples, assuming that the income distribution model as well as the parametric formula for an inequality statistic are both known.

Another method of variance estimation that can be used for poverty and inequality measures is balanced repeated replication. It proves especially useful when data come from a complex survey design with a large number of strata.

Let a stratified sampling frame with H strata be considered, with two subsets of primary sampling units (PSU) obtained for each stratum. They should be constructed in such a way that every stratum consists of two subsamples, each of them having similar number of units. A half-sample is a set consisting of one of the two subsets for each stratum. The number of all possible half-samples is 2^H, what may cause complication when the number of strata is large. To avoid such difficulties, we can choose the balanced set of R half-samples, so that the number of variants is significantly smaller than 2^H. The subset of balanced half-samples can be defined as a matrix of dimension $R \times H$ with the elements (r,h) equal δ_{rh} = +1 or –1, indicating whether the PSU from the h-th stratum selected for the r-th half sample is the first or the second PSU. The set of R half-samples is considered as balanced, if

$$\sum_{r=1}^{H} \delta_{rh}\delta_{rh'} = 0 \qquad \forall h = h'. \tag{15}$$

Balanced matrix RH can be obtained from Hadamard matrix, that has dimensions $R \times R$. The rows of Hadamard matrix denote half-samples while columns denote the strata, and the following condition is satisfied $H+1 \le R \le H+4$. Because the lines and columns in such a matrix are mutually orthogonal the half-samples selected are mutually independent (examples of Hadamard matrices to be found in: Bruch, Münnich and Zins, 2011).

The weights for selected elements may be equalized and are usually multiplied by 2 (Shao et al., 1998). Next, the estimated values for parameter of

interest $\hat{\theta}_r$ are determined by balanced repeated replication for each half-sample. The standard variance estimator can be expressed by the following formula:

$$\hat{V}_{BRR}(\hat{\theta}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \hat{\theta}_{StrRS})^2 \ , \tag{16}$$

where $\hat{\theta}_{StrRS}$ is parameter estimate for the whole sample in the case of stratified random sampling (StrRS).

4. Model-based approach and EBLUP estimation

Sample survey data can be used to derive reliable direct estimates for large domains (discussed in Section 3), but sample sizes in small domains are seldom large enough for direct estimators to provide adequate precision for these domains. Thus, it is necessary to employ indirect estimation methods that borrow strength from related areas. Many subpopulation parameters, including means and totals, can be expressed as linear combinations of fixed and random effects of small area models. Best linear unbiased prediction (BLUP) estimators of such parameters can be obtained in a classical way using BLUP estimation procedure. BLUP estimators minimize Mean Square Error (MSE) within the class of linear unbiased estimators and do not depend on the normality of random effects. Maximum likelihood (ML) or restricted maximum likelihood (REML) methods can be used to estimate the variance and covariance components, assuming normality.

The EBLUP procedure has been applied in many important statistical surveys conducted all over the world. The pioneer work in this area was that of Fay and Herriot (Fay, Herriot, 1979), where the EBLUP technique was used for evaluating per capita income and some other statistics obtained for counties. The model-based approach to small area estimation of inequality indices for regions in Poland was applied in the paper of Jędrzejczak, Kubacki (2010). The authors provided empirical Bayes (EB) and empirical best linear unbiased prediction (EBLUP) estimators under area level models. Molina and Rao (2010) estimated poverty indicators as examples of nonlinear small area population parameters by using the empirical Bayes (best) method, based on a nested error model. Hierarchical Bayes multivariate estimation of poverty rates for small domains was lately discussed by Fabrizi et al. (2011).

Many applications of EBLUP in the context of small area estimation are based on a special kind of the general linear mixed model, widely known as basic area level model (Rao, 2003):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} + \mathbf{e} \tag{17}$$

where: **y** is $n \times 1$ vector of sample observations, **X** – known matrix of explanatory variables, **β** is a vector of linear regression coefficients, **v** denotes area-specific random effect vector, **e** is sampling error vector.

It is usually assumed that **v** and **e** are independently distributed with mean **0** and covariance matrices **G** and **R,** respectively. EBLUP estimator for the small area model given by (17) has the following form:

$$\boldsymbol{\theta}_{EBLUP} = \mathbf{X}\boldsymbol{\beta} + \mathbf{MGV}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{18}$$

where: $\boldsymbol{\beta} = (\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{y}$, **M** is the identity matrix, **G** is the matrix with non-zero diagonal and its values are equal to σ_v^2, which is the model variance. It is usually computed using special iterative procedure that applies Fisher algorithm.

Mean square error estimate (MSE) of EBLUP can be obtained from the following formula:

$$MSE(\theta_{EBLUP}) = g_1(\hat{\delta}) - b_\delta^T(\hat{\delta})\nabla g_1(\hat{\delta}) + g_2(\hat{\delta}) + 2g_3(\hat{\delta}) \tag{19}$$

where δ is a variance dependent parameter. Using this formula we usually assume that the mean square error of EBLUP is the sum of three main elements g_1, g_2 and g_3 which are described by the following equations (Rao, 2003):

$$g_1(\hat{\delta}) = diag(\mathbf{G} - \mathbf{GV}^{-1}\mathbf{G}) \tag{20}$$

$$g_{2i}(\hat{\delta}) = (\mathbf{X_i} - \mathbf{m}_i^T\mathbf{GV}^{-1}\mathbf{X})(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X_i} - \mathbf{m}_i^T\mathbf{GV}^{-1}\mathbf{X}) \tag{21}$$

$$g_{3i}(\hat{\delta}) = (\mathbf{m}_i^T(\mathbf{V}^{-1} - \mathbf{G}(\mathbf{V}^{-1}\mathbf{V}^{-1}))\mathbf{V}(\mathbf{m}_i^T(\mathbf{V}^{-1} - \mathbf{G}(\mathbf{V}^{-1}\mathbf{V}^{-1}))^{\mathbf{T}})\mathbf{I} \tag{22}$$

where $\mathbf{m_i}$ is a vector with zeros for all elements with exception for the element having an index *i* while **I** is the inversed Fisher information matrix.

5. Application

The methods given above were applied to the estimation of inequality and poverty measures in Poland by region and family type. The basis for the calculations was micro data coming from the Polish Household Budget Survey (HBS) conducted in 2009. The data obtained from the household budget survey allow for the analysis of the living conditions of the population, 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 sample was selected by two-stage stratified sampling with unequal inclusion probabilities for primary sampling units. 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

number of persons and class of locality coming from the Population and Housing Census 2002.

The analysis has been conducted after dividing the overall sample by region NUTS1, constructed according to the EUROSTAT classification, and by 6 family types, classified according to the number of children. The variable of interest was household's 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 obtain the estimates of income inequality coefficients for selected subpopulations, the formulas (3) and (7) were applied, while poverty indicators were estimated by means of (12) and (13). Standard errors of the Gini and Zenga inequality measures were estimated using the parametric approach based on the three-parameter Burr type-III distribution, also called the Dagum model. The model is known to be well fitted to empirical income and wage distributions in different divisions. First, the maximum likelihood estimates of the Dagum model were calculated and then the formula (14) was applied to obtain the variances of the Gini and Zenga indices. The precision of poverty indicators was estimated by means of the balanced repeated replication technique BRR (See: formula (16)). It is worth mentioning that BRR estimators are typically estimators of design-based variances while methods based on ML, and specifically the one based on the Dagum model, are model-based. They generalise sample observations using predefined income distribution model instead of survey weights.

The results of the calculations are presented in Tables 1-3 and in Figures 1-6. To allow comparing the conditions of households of different sizes and different demographic structures, various income equivalence scales are used. The square root scale, popular in recent OECD publications, was applied in the paper. It is based on division of individual household income by the square root of the household size. As a poverty threshold we used 60% of median equivalised income value.

Table 1. Estimates of inequality and poverty indices with their standard errors by family type

No.	Number of children	Inequality coefficient		Poverty index		
		Gini	Zenga	Headcount ratio	Poverty gap	Poverty severity
1	0	0.36 (0.004)	0.37 (0.007)	15.4 (0.230)	26.2 (0.442)	12.2 (0.402)
2	1	0.32 (0.009)	0.30 (0.016)	13.2 (0.405)	29.7 (0.952)	16.0 (0.930)
3	2	0.33 (0.014)	0.31 (0.021)	16.9 (0.597)	28.4 (0.912)	14.2 (0.911)

Table 1. Estimates of inequality and poverty indices with their standard errors by family type (cont.)

No.	Number of children	Inequality coefficient		Poverty index		
		Gini	Zenga	Headcount ratio	Poverty gap	Poverty severity
4	3	0.32 (0.031)	0.30 (0.059)	27.1 (1.525)	27.8 (1.278)	13.3 (1.234)
5	4	0.30 (0.057)	0.28 (0.125)	33.5 (2.763)	31.1 (2.515)	15.7 (2.521)
6	5 …	0.29 (0.072)	0.26 (0.137)	39.6 (4.188)	29.0 (2.593)	15.8 (2.435)
7	Total	0.35 (0.003)	0.36 (0.005)	15.9 (0.253)	27.2 (0.396)	13.2 (0.381)

Source: Authors' calculations on the basis on HBS 2009.

Table 2. Estimates of inequality and poverty indices with their standard errors by region

No.	Region	Inequality coefficient		Poverty index		
		Gini	Zenga	Headcount ratio	Poverty gap	Poverty severity
1	Central	0.39 (0.006)	0.43 (0.011)	13.9 (0.419)	30.0 (1.137)	16.3 (1.228)
2	Southern	0.32 (0.008)	0.30 (0.015)	13.2 (0.405)	24.5 (0.755)	10.4 (0.673)
3	Eastern	0.35 (0.008)	0.36 (0.014)	23.5 (0.630)	28.6 (0.737)	14.3 (0.676)
4	North-western	0.33 (0.009)	0.32 (0.016)	14.5 (0.859)	23.7 (1.046)	10.1 (0.852)
5	South-western	0.35 (0.010)	0.36 (0.018)	15.3 (0.880)	28.5 (0.974)	14.3 (0.913)
6	Northern	0.34 (0.009)	0.35 (0.016)	15.7 (0.689)	26.7 (1.166)	13.0 (0.913)
7	Total	0.35 (0.003)	0.36 (0.005)	15.9 (0.253)	27.2 (0.396)	13.2 (0.381)

Source: Authors' calculations on the basis on HBS 2009.

Table 3. Estimates of poverty indices and their standard errors by region and family type

	Region	Number of children	Sample size	Poverty index					
				Headcount ratio		Poverty gap		Poverty severity	
				Estimate	Standard Error	Estimate	Standard Terror	Estimate	Standard error
1	Central	0	5469	18.047	0.629	27.482	0.676	13.173	0.655
		1	1393	13.516	0.892	31.206	2.502	17.854	2.784
		2	962	19.465	1.238	31.002	2.518	17.927	2.564
		3	216	33.325	4.478	34.347	3.226	19.130	3.276
		4	39	48.989	9.800	33.177	5.712	17.840	6.112
		5…	22	38.354	12.793	30.144	12.016	17.045	13.231
2	Southern	0	4871	13.510	0.633	24.624	0.958	10.033	0.794
		1	1374	13.407	1.095	26.455	1.358	12.261	1.267
		2	924	16.706	1.292	22.026	1.011	8.614	1.063
		3	235	29.368	3.447	21.598	2.850	7.384	1.968
		4	55	29.153	4.683	26.079	4.139	12.678	4.284
		5…	26	51.757	8.413	22.207	7.121	10.841	6.507
3	Eastern	0	4009	16.270	0.731	27.201	1.286	14.540	1.169
		1	1276	12.619	0.889	35.823	2.077	21.860	1.826
		2	914	14.765	1.050	33.173	2.165	18.960	2.130
		3	293	22.069	1.554	28.704	3.654	15.467	3.201
		4	76	23.275	5.846	20.817	4.501	6.446	2.439
		5…	35	37.941	8.772	32.748	6.853	17.063	7.159
4	North-western	0	3618	14.759	1.136	23.029	1.301	9.202	1.039
		1	1155	13.004	1.063	24.543	2.004	11.538	1.834
		2	719	17.003	1.934	23.129	1.654	9.522	1.390
		3	197	24.825	4.309	23.835	3.164	11.100	3.352
		4	47	32.767	6.989	32.015	9.682	20.310	9.774
		5…	23	38.979	7.540	13.452	3.689	3.154	2.089
5	South-western	0	2640	16.196	0.926	28.338	1.225	13.440	1.174
		1	752	12.962	1.062	33.131	3.589	20.641	3.696
		2	418	19.401	1.518	25.956	2.719	12.101	2.470
		3	109	30.216	6.194	24.314	2.995	9.256	2.364
		4	32	44.162	7.695	29.473	4.901	13.911	5.686
		5…	5	29.683	–	27.222	–	12.639	–
6	Northern	0	3378	13.037	0.648	25.336	1.286	12.460	1.424
		1	996	12.824	0.925	28.000	2.734	13.766	2.185
		2	720	18.372	2.048	27.459	2.103	13.194	1.841
		3	223	22.826	2.457	32.090	2.636	16.275	3.107
		4	48	42.105	6.384	35.491	5.465	18.427	5.683
		5…	33	35.697	12.660	36.954	8.481	25.483	9.988

Source: Authors' calculations on the basis on HBS 2009.

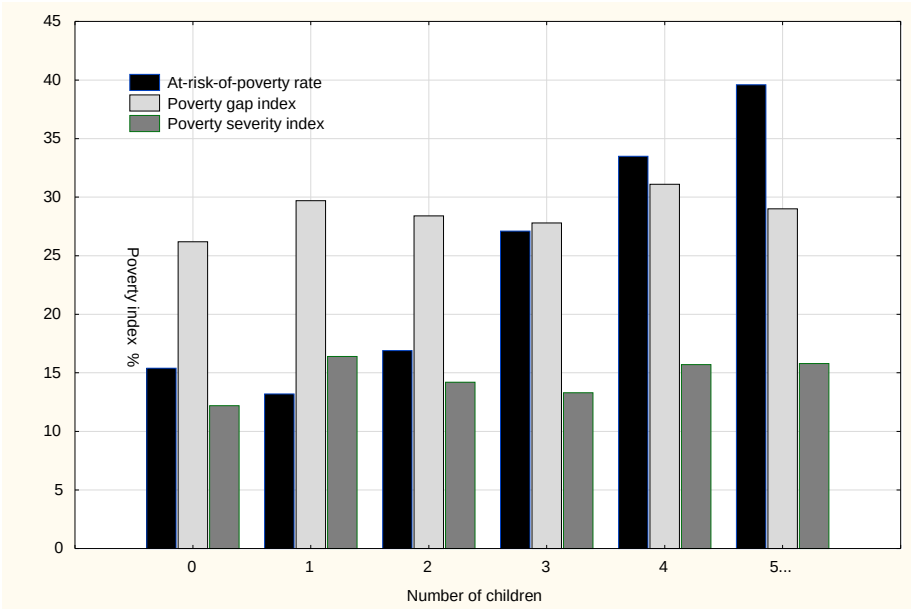

Figure 1. Poverty characteristics of families with different number of children

Source: Authors' calculations.

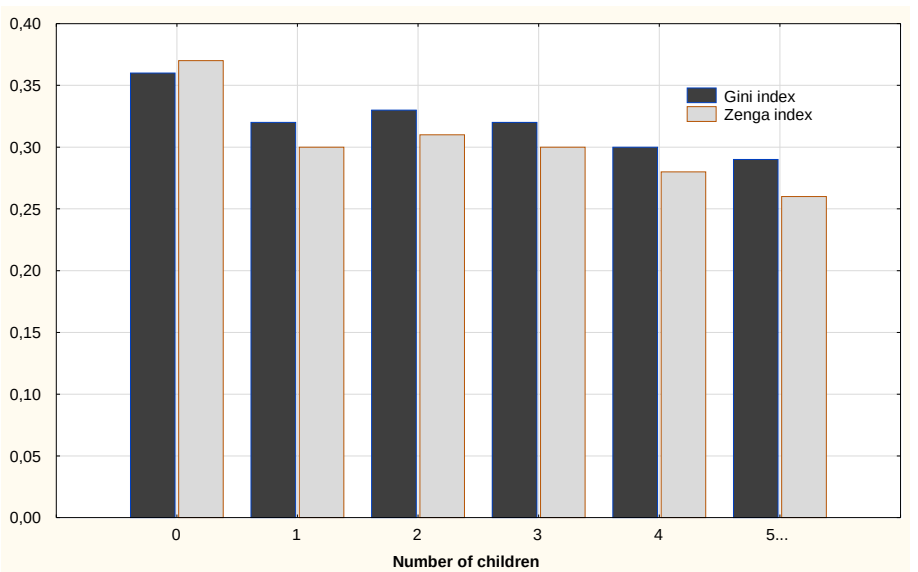

Figure 2. Inequality characteristics of families with different number of children

Source: Authors' calculations.

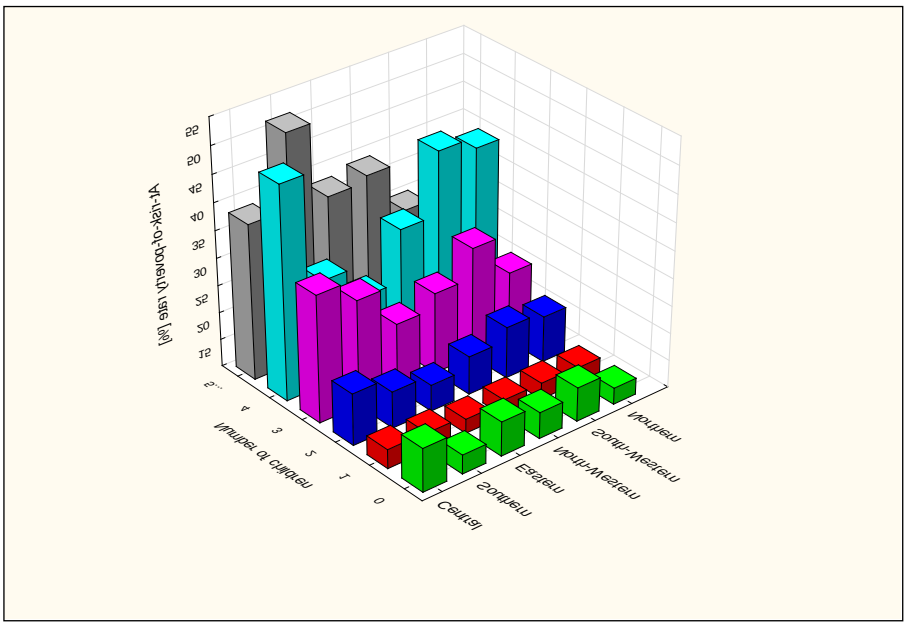

Figure 3. At- Risk-of-Poverty Rates (ARPR) by family type and region
Source: Authors' calculations.

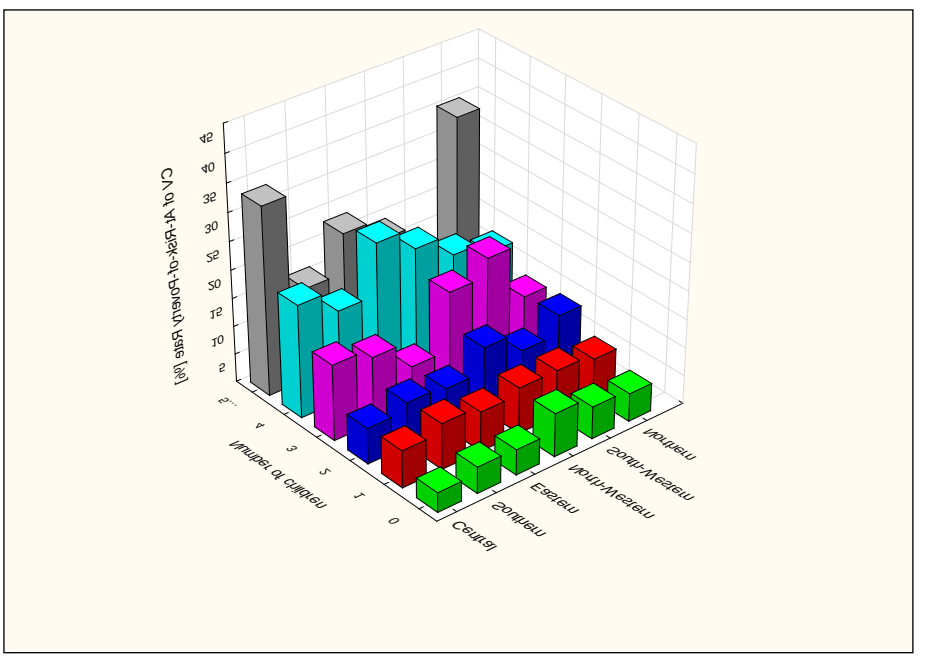

Figure 4. Coefficients of variation for ARPRs by family type and region
Source: Authors' calculations.

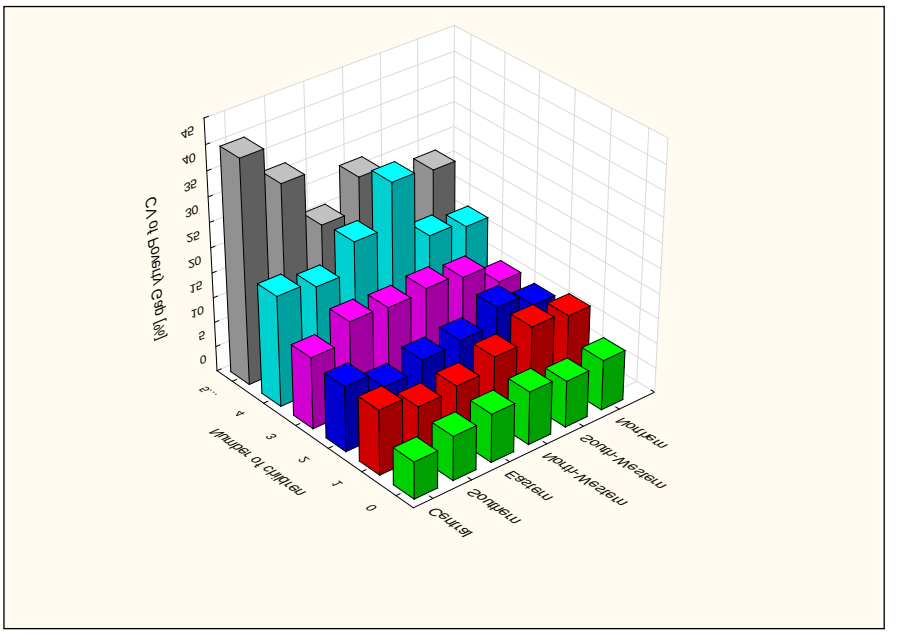

Figure 5. Coefficients of variation for Poverty Gaps by family type and region
Source: Authors' calculations.

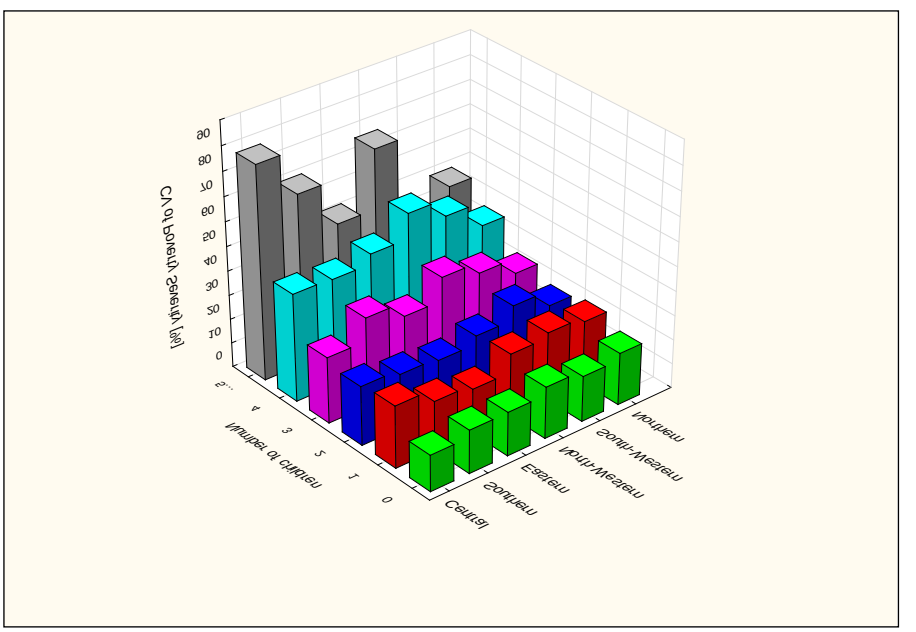

Figure 6. Coefficients of variation for Poverty Severity indices by family type and region

Source: Authors' calculations.

Tables 1 and 2 comprise estimates of inequality and poverty indices by family type and by region, separately. Diversification of household income, expressed by the Gini and Zenga coefficients (relatively stable since 2003), can be considered high in comparison with other European countries. The value of the Gini index for the total household was 0.35 while the Zenga index estimate was even higher (0.36). It is interesting to observe that the number of children is a factor clearly differentiating the level of income inequality and poverty of households. The higher the number of children, the higher the level of poverty indices (especially ARPR), and generally the lower is the level of income inequality. The latter is expressed, among others, by the fact that the poorest families with many children live mainly on social benefits, so their equivalised incomes are relatively similar. Families with five or more children present at-risk-of- poverty rate close to 40%, with simultaneously low level of the Gini index (0.29). On the other hand, discrepancies between regions are much smaller, with the *Eastern* region still having the worst position in terms of income (*ARPR*=23.5%, *G*=0.35, *Z*=0.36).

The detailed estimation results for the division of the entire population by family type (number of children), and at the same time by region, are presented in Table 3. The estimated values of three basic poverty measures for subpopulations, headcount ratio (ARPR), poverty gap index and poverty severity index, are accompanied by their standard errors estimated by means of balanced repeated estimation technique BRR. The calculations were done using WesVar package.

All the results presented in Tables 1-3 are supported by their estimated standard errors. Moreover, Figures 4-6 show the coefficients of variation (CV) for poverty coefficients outlined in Table 3. Analyzing the results of the calculations given above, one can easily notice that the precision of poverty indicators is unsatisfactory, especially when the division of households by family type and simultaneously by region is considered (See: Table 3 and Figures 4-6). The standard errors are relatively small only for the first household type, i.e. families without children, and usually account for about 5% of the corresponding estimates. For the remaining family types the efficiency of estimation is poor as coefficients of variation exceed 30% in many cases.

Thus, we have come to the conclusion that sample sizes of the subpopulations are apparently too small to provide reliable direct estimates. When a subpopulation is too small to yield direct estimates with adequate precision, it is regarded as a small area. Some indirect estimation methods based on borrowing strength in time and in space have been developed to overcome small area estimation problems (Rao, 2003). It is now generally accepted that when indirect

estimation is to be used, it should be based on explicit small areas models (See: §4). To improve the precision of head-count-ratio estimation for family types in regions, a model-based approach using the basic area level model (17) was applied. First, a standard linear regression model was constructed using some auxiliary sources of data that come from Polish Public Statistics and administrative registers. Regional per capita income GDP and the number of children NC played the role of explanatory variables in the model:

Model parameter	Coefficient value	Standard error	t-statistic	*p*-value
Intercept (α_0)	0.000865	0.050108	0.017259	0.986334
GDP (α_1)	0.001067	0.000477	2.239633	0.031971
NC (α_2)	0.056487	0.005636	10.02267	1.53E-11

For the specified model, the value of determination coefficient ($R^{2)}$ is 0.762, the value of corrected R^2 is equal to 0.747, the F statistics is $F(2.33)=52.735$ $p<0.00000$, and standard estimation error is equal to 0.05775.

The results of EBLUP estimation of ARPRs are summarized in Table 4. As it can be easily noticed, the model-based approach induced significant refinement for almost all subpopulations. Mean squared errors of EBLUP estimates are smaller than the corresponding standard deviations of direct estimates. In the last column we show the reduction of CV which exceeds 60% in some cases, yet on average it is equal to 21.54% .

Because the correlations between poverty gap (or poverty severity) and the same auxiliary variables as for the model of headcount ratio are relatively weak, we do not include the models for these cases. However, a preliminary correlation analysis for other explanatory variables, including the Gini and Zenga coefficients, has been carried out. It reveals a relatively strong correlation between the poverty and inequality measures that was found to be 0.46 for Gini and poverty gap and 0.43 for Gini and poverty severity (for the Zenga measure the correlation coefficients were 0.52 for poverty gap and 0.48 for poverty severity). It can be assumed that also other poverty related variables, e.g. unemployment rate, may be included in such computations. A more detailed analysis of such cases goes beyond the scope of the paper yet it may be performed in the future.

Table. 4. ARPR estimation results using direct and model-based approach

	Region	Number of children	Sample size	Direct estimation		EBLUP estimation		CV Red. [%]
				Estimate	Standard error	Estimate	Root mean squared error	
1	Central	0	5469	18.047	0.629	17.926	0.626	**0.1**
		1	1393	13.516	0.892	13.834	0.877	**3.9**
		2	962	19.465	1.238	19.901	1.197	**5.4**
		3	216	33.325	4.478	30.071	3.085	**23.7**
		4	39	48.989	9.800	34.564	3.895	**43.7**
		5…	22	38.354	12.793	37.416	4.210	**66.3**
2	Southern	0	4871	13.510	0.633	13.447	0.627	**0.6**
		1	1374	13.407	1.095	13.649	1.059	**5.0**
		2	924	16.706	1.292	17.190	1.233	**7.3**
		3	235	29.368	3.447	27.438	2.598	**19.3**
		4	55	29.153	4.683	29.739	3.063	**35.9**
		5…	26	51.757	8.413	37.250	3.740	**38.2**
3	Eastern	0	4009	16.270	0.731	15.996	0.722	**-0.5**
		1	1276	12.619	0.889	12.719	0.872	**2.7**
		2	914	14.765	1.050	15.105	1.021	**4.9**
		3	293	22.069	1.554	22.313	1.466	**6.7**
		4	76	23.275	5.846	26.895	3.362	**50.2**
		5…	35	37.941	8.772	33.555	3.837	**50.5**
4	North-western	0	3618	14.759	1.136	14.447	1.098	**1.2**
		1	1155	13.004	1.063	13.263	1.030	**5.0**
		2	719	17.003	1.934	17.870	1.743	**14.2**
		3	197	24.825	4.309	25.158	2.886	**33.9**
		4	47	32.767	6.989	30.577	3.441	**47.2**
		5…	23	38.979	7.540	35.436	3.666	**46.5**
5	South-western	0	2640	16.196	0.926	15.906	0.906	**0.4**
		1	752	12.962	1.062	13.246	1.029	**5.2**
		2	418	19.401	1.518	19.654	1.423	**7.5**
		3	109	30.216	6.194	26.696	3.245	**40.9**
		4	32	44.162	7.695	32.696	3.510	**38.4**
		5…	5	29.683	–	34.364	3.841	**-**
6	Northern	0	3378	13.037	0.648	12.962	0.641	**0.5**
		1	996	12.824	0.925	12.988	0.903	**3.6**
		2	720	18.372	2.048	18.786	1.826	**12.8**
		3	223	22.826	2.457	23.425	2.105	**16.5**
		4	48	42.105	6.384	32.281	3.379	**31.0**
		5…	33	35.697	12.660	34.064	3.927	**67.5**

Source: Authors' calculations on the basis of HBS 2009 and CSO Local Data Bank.

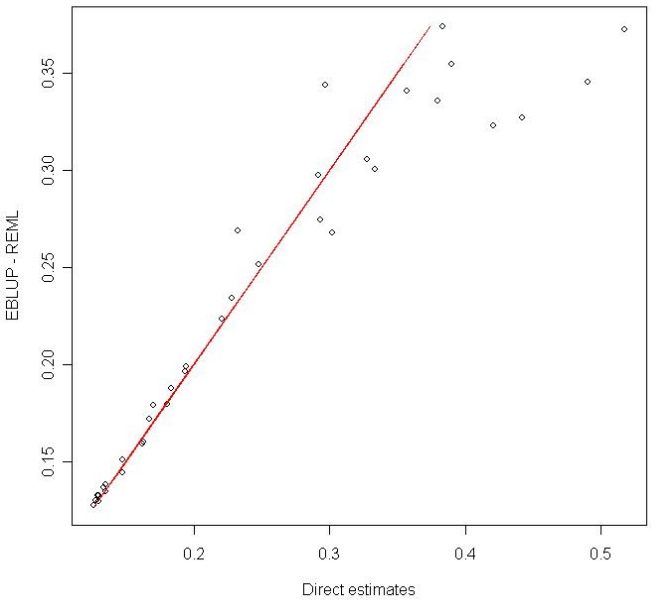

Figure 7. EBLUP (REML) estimates versus direct estimates

6. Conclusions

Efficient estimation of income distribution parameters, especially inequality and poverty characteristics and their standard errors, can be a serious problem for small domains and should be analyzed in detail. The EBLUP estimators for headcount ratios applied in the paper have proved to be more efficient than the corresponding direct estimators, as a result of "borrowing strength" from other subpopulations. The EBLUP estimation procedure, based on a general linear mixed model, has an additional advantage of taking into account the between-area variation beyond that explained by the auxiliary variables included in a classical regression model. The estimates of inequality and poverty measures by subpopulations presented in the paper can provide economists and social policy makers with valuable information that can help them to improve the decision-making process and bring them to adequate economic allocations. Understanding the domain-specific profiles may prove crucial in developing appropriate polices on most efficient reduction of the global poverty. In the future, it would also be interesting to consider more advanced cases of small area models, including poverty gap and poverty severity ratio models, that can account for differences between domains of interest more precisely.

REFERENCES

ALY, E., HERVAS, M., (1999). Nonparametric Inference for Zenga's Measure of Income Inequality, Metron, LVII (1-2), pp. 69–84.

BRUCH, CH., MÜNNICH, R., ZINS, S., (2011). Variance Estimation for Complex Surveys, AMELI project, Deliverable 3.1.

FABRIZI, E., FERRANTE, M. R., PACEI, S., TRIVISANO, C., (2011). Hierarchical Bayes Multivariate Estimation of Poverty Rates Based on Increasing Thresholds for Small Domains, Computational Statistics & Data Analysis, 55 (4), pp. 1736–1747.

FAY, R. E., HERRIOT, R. A., (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data, Journal of the American Statistical Association, 74, pp. 269–277.

JĘDRZEJCZAK, A., (2011). Metody analizy rozkładów dochodów i ich koncentracji, Łódź University Press, Łódź.

JĘDRZEJCZAK, A., KUBACKI, J., (2010). Estimation of Gini Coefficient for Regions from Polish Household Budget Survey using Small Area Estimation Methods, [in:] *Survey Sampling Methods in: Economic and Social Research*, Edited by Wywiał J., Gamrot W., Akademia Ekonomiczna w Katowicach, Katowice, pp. 109–124.

MOLINA, I, RAO, J. N. K., (2010). Small Area Estimation of Poverty Indicators, Canadian Journal of Statistics 38, pp. 369–385.

LANGEL, A., TILLÉ, Y., (2013). Variance Estimation of the Gini Index: Revisiting a Result Several Times Published, Journal of the Royal Statistical Society: Series A, 176 (2), pp. 521–540.

PANEK, T., (2008). Ubóstwo i nierówności: dylematy pomiaru, in: Statystyka społeczna – dokonania, szanse, perspektywy, BWS, t. 57, GUS, Warszawa, 2008, pp. 96–108.

RAO, J. N. K., (2003). Small Area Estimation, Wiley, London.

SÄRNDAL, C. E., SWENSSON, B., WRETMAN, J., (1997). Model Assisted Survey Sampling, Springer, New York.

SEN, A., (1976). Poverty - an Ordinal Approach to Measurement, Econometrica 44, pp. 219–231.

SHAO, J., CHEN, Y., CHEN, Y., (1998). Balanced Repeated Replication for Stratified Multistage Survey Data under Imputation. Journal of the American Statistical Association, 93 (442), pp. 819–831.

WOLTER, K., (2003). Introduction to Variance Estimation, Springer-Verlag, New York.

ZEHNA, P. W., (1966). Invariance of Maximum Likelihood Estimation, Annals of Mathematical Statistics 37, pp. 744–745.

ZENGA, M., (1984). Proposta per un Indice di Concentrazione Basato sui Rapporti tra Quantili di Popolazione e Quantili di Reddito. Giornale degli Economisti e Annali di Economia, 48, pp. 301–326.

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