STATISTICS IN TRANSITION new series, September 2025
Vol. 26, No. 3, pp. 135–154, https://doi.org/10.59139/stattrans-2025-031
Received – 17.12.2024; accepted – 11.03.2025

Regional differentiation of income distributions in Poland

Alina Jędrzejczak[1], Małgorzata Misztal[2], Dorota Pekasiewicz[3]

Abstract

Income inequality is observed to have recently increased both at the country and regional level. Consequently, inequality, poverty and social stratification have become important issues of debate among economists, sociologists and social policy-makers. Thus, an in-depth statistical analysis of the income situation of households is particularly important when counteracting the social effects of the discussed phenomenon. In the literature, the situation of regions in Poland is compared primarily in terms of the differences in GDP or average wages observed for households. The aim of the article is to analyze the regional differences in the entire income distribution in Poland, taking into account not only average income levels, but also income inequality and poverty parameters. The study, based on individual data from the Household Budget Survey, used parametric and non-parametric methods for estimating inequality and poverty measures, as well as cluster analysis methods. In the parametric approach, the empirical income distributions in Poland were approximated using the theoretical Dagum distribution. This enabled the segmentation of voivodships in terms of the estimated characteristics of the equivalent household income distribution. The results of the calculations confirmed the assumption that income distributions in Poland differ significantly across regions, and the obtained clusters allowed the detection of groups of regions that may require a separate social policy aimed at improving the material situation of households.

Key words: income distribution; inequality; poverty; Dagum model; cluster analysis.

1. Introduction

Reducing regional inequality was one of the key means of promoting the "harmonious development" within Europe envisioned in the EEC Treaty of 1957.

[1] Department of Statistical Methods, University of Lodz; Lodz, Poland & Statistical Office in Lodz, Lodz, Poland. E-mail: alina.jedrzejczak@uni.lodz.pl, ORCID: https://orcid.org/0000-0002-5478-9284

[2] Department of Statistical Methods, University of Lodz, Lodz, Poland. E-mail: malgorzata.misztal@uni.lodz.pl, ORCID: https://orcid.org/0000-0002-8719-2097

[3] Department of Statistical Methods, University of Lodz, Lodz, Poland. E-mail: dorota.pekasiewicz@uni.lodz.pl, ORCID: https://orcid.org/0000-0001-8275-3345

The pursuit of "economic, social and territorial cohesion" through ever closer regional and national harmonization was also proclaimed in the 2007 Lisbon Treaty, but deepening European integration has not always been matched with convergence in living standards between sub-national regions. The gap between poorer and richer areas increased during the last economic crisis even in some developed economies, and the income discrepancy between richer and poorer regions is likely to widen further as government-spending cuts disproportionately hurt less prosperous regions. Over the last few decades, income inequality measured by the Gini index has been growing not only in Europe (see: *Growing Unequal*, OECD 2008; *Divided We Stand. Why Inequality Keeps Rising*, OECD 2011; *In It Together: Why Less Inequality Benefits All*, OECD 2015), and in 2014 most OECD countries recorded the highest level of inequality in 30 years. An even more pessimistic picture of income and wealth inequalities can be seen in a recent publication of OECD (2024), where it has been revealed that more than one-third of people in the OECD are at risk of poverty. The total income inequality can be decomposed into between- and within- regions inequality. Recent trends on income inequality in the EU confirm that, by decomposing overall inequality, as much as 85% is explained by within countries inequality (Bonesmo Fredriksen, 2012), while the findings of Barca (2009) reveal how prosperous regions in EU countries show at the same time strong internal, i.e. subnational, disparities. It has also been found that nearly 40% of total income goes on average to people in the highest income quintile, and less than 10% to people in the first quintile (Eurostat, 2014). In a working paper of Savoia (2020), devoted to the question whether the EU Cohesion Policy may have contributed to reduce the inequalities within countries or regions, it has been found that income inequality observed within NUTS 2 is converging, but to a higher level, so the regions "have tended to become equally more unequal". As a result of the increasing inequalities within and across countries and regions, the reduction in the number of persons at risk of poverty or social exclusion was one of the five key targets of the Europe 2020 strategy.

Income inequality in Poland increased significantly during the process of transformation from a centrally-planned to a market economy - the Gini index went up by approximately 10 percentage points. After the period of rapid economic changes the rate of growth of the Gini index slowed down and for a long time we could observe only slight fluctuations, at about the level of 0.34-0.35, according to the Household Budget Survey (HBS) data, and 0.31 according to EU-SILC. After the decline between 2016 and 2019 the Gini ratio in Poland began increasing again, and in 2023 it reached a very high level of 0.414 (GUS, 2024, p. 212), with high discrepancy between income distributions for socio-economic groups.

Regional inequalities can be measured and analyzed in many different ways - the extent of inequality may be mapped in terms of demography, income and wealth, labor markets, and education and skills. The main focus of this presentation is on the analysis of regional inequalities in terms of household income distribution. Income distributions can differ in many aspects and it seems definitely not enough to compare them only from the point of view of average income levels. Economic inequalities are particularly important today and can be measured by income inequality between and within regions, among other things. Regional income inequality can be strongly linked to the poverty and social exclusion observed across regions and socioeconomic groups. Having said that, though, it is worth noting that the dependencies between average income, inequality and poverty are not straightforward and they depend on many other socio-economic factors.

It is well known that high income inequality can have several undesirable political and social consequences, such as poverty and the polarization of particular economic groups. Although they are usually perceived as similar and are in fact highly related concepts, inequality and poverty may not always come together. One can imagine a strictly egalitarian distribution of incomes, where all the income receivers are poor, or a highly dispersed population without poverty. Setting aside these not very likely situations, there is a strong empirical evidence based on income data from many countries that confirms a strong positive correlation between inequality and poverty. As a consequence, the countries with a more dispersed income distribution tend to have a higher relative level of income poverty, with only a few exceptions. The situation can be different, however, when analyzing these variables across regions. The relationships between economic development (measured by GDP per capita or average household income) and the corresponding income inequality (measured by, e.g. the Gini index) can also be different across countries and regions, what have been discussed in Jędrzejczak (2015) for the Polish and Italian regions. The observed regularities seem to partially confirm the well-known Kuznets' "inverted-U" relationship between the level of economic development and income inequality.

It is worth noting that the discussion on the possible relationship between GDP and the inequality level, which has been present in the economic literature since the mid-1950s, has produced very inconclusive results. We can find many countries (e.g. the Czech Republic) where the process of transformation was connected with no substantial inequality growth, while for many developed countries the inequality first declined, then increased again after a tipping point had been reached (e.g.: Italy in the 1990s). Deininger and Squire (1998, pp. 259-287), using their famous panel data on income inequality, did not find any significant relationship between income inequality and the level of development, even when country-effects were included into the

analysis. Li, Squire and Zou (1998, pp. 26-43) found out that the Kuznets relationship seems to work better in cross-sectional than time-series analyses.

Polish voivodeships differ in the level of economic development. There are many reasons for this differentiation and they concern various areas of socio-economic policy. The most important determinants of economic development include: the levels of regional GDP per capita, average incomes of households and individuals, characteristics of the labor market, demographic indicators, infrastructure and many others. Various works have recently been published on this issue, including: Malina (2020), Dańska-Borsiak (2024), where the authors take into account many indicators, the basic ones being GDP and household income, and then use multivariate statistical methods, which leads to interesting results. The data on GDP per capita and the estimates of household income suggest that there are substantial differences in regional income levels across the country, but little can be deduced from this about differences within regions and the relative number of people in different regions with income below the poverty line, as defined at the national level. In our study, we focus on income distribution characteristics which constitute an important dimension of regional disparities.

In the paper we would like to compare and contrast the regions of Poland (voivodeships) from the point of view of differences in their income distributions. The main aim is to analyze regional differences in the entire income distribution in Poland, taking into account not only average income levels, but also the parameters of income inequality and poverty. The study used non-parametric and parametric methods for estimating inequality and poverty measures, as well as cluster analysis methods for grouping the regions. This enabled the segmentation of voivodeships in terms of the estimated characteristics of the equivalent household income distribution. The applied parametric approach fills the gap in the existing grouping of regions, allowing for the estimation of income distribution parameters based on the Dagum distribution model, which has a strong stochastic basis.

The empirical evidence was micro-data coming from the Household Budget Survey conducted by Statistics Poland (GUS) for the year 2021. The detailed objectives of the paper are the following:

- analysis of income distributions in Polish voivodeships and their approximation based on the Dagum distribution,
- segmentation of Polish voivodeships according to estimated distribution characteristics, including the level of poverty and income inequality, using hierarchical cluster analysis methods.

The novelty of the paper is the segmentation of Polish voivodeships combined with the parametric approach based on the density curves of their income distributions.

2. Methodological remarks

2.1. Income distribution characteristics

Let $Y > 0$ be a random variable representing gross or net incomes as well as taxes. Let $F(y)$ be its cumulative distribution function (cdf), $f(y)$ the corresponding density function (pdf) and $F^{-1}(p) = \inf\{y: F(y) \geq p\}$ its quantile function for $0 < p < 1$.

Income distribution characteristics comprise many popular descriptive statistical measures, frequently applied in numerous statistical analyses, including measures of central tendency, dispersion and shape, but a special attention is paid to the measurement of distribution inequality. 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. Amongst numerous inequality measures the most popular is the Gini index, defined in terms of the Lorenz curve (Gini, 1912;1914). In its over 100-year history, the index has had numerous formulas and interpretations (see, e.g.: Jędrzejczak, 2011).

In practice, inequality coefficients, including the **Gini index**, are usually estimated from empirical data coming from representative samples. One can estimate the value of the Gini index from the survey data using the following formula (1), proposed by Fei et al. (1978):

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

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.

Formula (1) can be applied for individual data coming from random samples, including the samples obtained within the Household Budget Survey, as it incorporates survey weights available for each statistical unit and is constructed on the basis of their inclusion probabilities. These weights make it possible to adjust for unequal selection probabilities and non-response bias.

Another type of inequality indices are the measures based on distribution quantiles or the estimators of these quantiles – these measures are very popular due to their simplicity and straightforward economic interpretation. In this class the most useful measures are: quintile and decile dispersion ratios (Panek, 2011) and the reciprocal of

the decile dispersion ratio called **the dispersion index for the end portions of the distribution**. The estimator of this index is:

$$\widehat{K}_{1:10} = \frac{\sum_{i \in GD_1} y_i}{\sum_{i \in GD_{10}} y_i} \tag{2}$$

where GD_j is *j*-th decile group.

This index (2) takes values from the interval [0,1] and if it is closer to 1, the inequality is lower (mean incomes in the extremal decile groups are the same).

A natural consequence of income inequality can be material poverty and social exclusion. Starting with the seminal publication of Sen (1976), numerous poverty measures have been proposed and analyzed but in practice the most popular are poverty indexes belonging to the class FGT (Foster–Greer–Thorbecke**),** which addresses three basic aspects of this phenomenon: poverty incidence, depth and severity (see: Panek, 2011). In this class, the most popular index is undoubtfully at-risk-of-poverty rate also called **poverty head-count ratio**, which can be estimated from the *n*-element sample by means of the following formula:

$$\widehat{W}_p = \frac{\sum_{i=1}^{n} I_i w_i}{\sum_{i=1}^{n} w_i}, \tag{3}$$

where: I_i – indicator function taking value 1 when *i*-th household's equivalent income is below a poverty line, and taking value 0 in the opposite situation,

w_i – survey weight for *i*-th economic unit.

The poverty threshold is usually assumed as 60% of the national median equivalent income - this approach is used in EU-SILC, the European Union Statistics on Income and Living Conditions, anchored by Eurostat. In some publications of Statistics Poland one can find another relative poverty line - 50% of mean equivalent income. In this study we utilize the Eurostat poverty threshold equal to $y^* = 0.6Me$, where *Me* is established as the estimate of median equivalent household income for the whole country.

Poverty gap index provides additional 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 number of the poor.

The estimator of the poverty gap index, which incorporates sampling weights w_i, is the following:

$$\widehat{PG}_p = \frac{\sum_{i=1}^{n_p} ((y^* - y_i)/y^*) w_i}{\sum_{i=1}^{n_p} w_i}, \tag{4}$$

where: y_i is household equivalent income,

y^*denotes poverty line (poverty threshold),

n_p is the number of poor households,
w_i accounts for the survey weight for i-th household.

The index (4) can be interpreted as minimum income that should have been transferred from the rich to the poor households to totally cancel poverty.

For many practical applications it seems reasonable to approximate empirical income data by means of a well-fitted theoretical model.

Since Pareto proposed his first income distribution model in 1896, many economists and mathematicians have tried to describe empirical distributions by simple mathematical formulas with a small number of parameters. These formulas can be useful for many reasons. Firstly, applying a theoretical model simplifies the analysis, because different distribution characteristics can be performed by the same parameters. Secondly, a theoretical model well fitted to the data can be used to the prediction of wage and income distributions in different divisions. Moreover, approximating empirical wage and income distributions using theoretical curves can smooth out irregularities resulting from imperfect data collection methods, which is often the case with income data. After decades of applications to real data in different divisions, we can propose the Dagum model as an excellent candidate to model income distributions of males and females (see: Jędrzejczak, Pekasiewicz, 2020).

The Dagum distribution is based on both empirical and stochastic foundations, similarly to the Pareto model (Dagum, 1977), but in contrast to the Pareto it fits population income quite well throughout the entire domain and has been successfully applied for many countries all over the world. Therefore, in the case of different subpopulations, e.g. regions or social groups, fitting the income distribution by means of the Dagum distribution can also be applied.

The probability density function of the Dagum distribution is given by (Kleiber and Kotz, 2003, Jędrzejczak and Pekasiewicz, 2020):

$$f(x) = \frac{ap}{b^{ap}} x^{ap-1} \left(1 + \left(\frac{x}{b}\right)^{a}\right)^{-1-p} \quad \text{for } x > 0, \tag{5}$$

and the corresponding cdf takes the form:

$$F(x) = \left(1 + \left(\frac{x}{b}\right)^{-a}\right)^{-p} \quad \text{for } x > 0. \tag{6}$$

where $a > 0$ is scale parameter, $b > 0$ and $p > 0$ are shape parameters.

Using a fit to the Dagum distribution, the Gini index can be determined from the formula:

$$G = \frac{\Gamma(\hat{p})\Gamma\left(2\hat{p}+\frac{1}{\hat{a}}\right)}{\Gamma(2\hat{p})\Gamma\left(\hat{p}+\frac{1}{\hat{a}}\right)} - 1 \quad \text{for } \hat{a} > 1, \tag{7}$$

and poverty head-count ratio from the formula:

$$\widehat{W} = \hat{F}(y^*) \tag{8}$$

Deciles necessary to determine the decile groups in the definition of the dispersion index for the end portions of the distribution (see: formula (2)) are determined based on the quantile function of the Dagum distribution.

Any measure of deviation, based on a synthesis of the absolute or squared differences between observed and estimated frequencies obtained for class intervals, can be a candidate for assessing the goodness of fit. We will evaluate the following goodness-of-fit measures based on this approach: the Mortara index (A_1) and the Pearson index (A_2), together with the coefficient of distribution similarity *Wp,* also called an overlap measure (Gastwirth, 1976).

The coefficient of distribution similarity can be obtained from grouped data by means of the following formula:

$$Wps = \sum_{j=1}^{s} \min(w_j, \widehat{w}_j), \tag{9}$$

where: *s* is the number of class intervals, w_j and $\widehat{w}_j$ denote empirical and theoretical frequencies, respectively.

The other consistency measures mentioned above take the form (Zenga et al., 2010):

$$A_1 = \frac{1}{n}\sum_{j=1}^{s} |n_j - \hat{n}_j|, \tag{10}$$

$$A_2 = \sqrt{\frac{1}{n}\sum_{j=1}^{s} \frac{(n_j - \hat{n}_j)^2}{\hat{n}_j}}, \tag{11}$$

where n_j and $\hat{n}_j$denote the empirical and theoretical frequencies for class intervals.

2.2. Methods of cluster analysis adopted in the study

The segmentation of Polish voivodeships from the point of view of theoretical income distribution parameters estimated on the basis of the HBS 2003-2011 was proposed in Jędrzejczak, Kubacki (2017). Brzezińska (2018) applied correspondence analysis and the Ward hierarchical method to a multivariate analysis of poverty in Poland, based on the reports on economic poverty published by Statistics Poland in 2015.

In the paper, the hierarchical cluster analysis methods were used to perform the segmentation of Polish voivodeships in terms of selected income distribution characteristics, concerning income inequality and poverty indices.

At the first stage, the clusterSim package (Walesiak, Dudek, 2020) was applied to determine the optimal clustering procedure. In our study, complete-linkage clustering with the general distance measure $GDM1$ (Walesiak, 2016) turned out to be optimal, as it maximizes the silhouette index considered the best measure of classification quality. The procedure was finally implemented by adopting interval or ratio scales for measuring variables and taking into account the number of clusters no greater than 7, after a prior normalization of variables using classical standardization.

The general distance measure $GDM1$ for metric data is given by the following equation:

$$GDM1 = d_{ik} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(z_{ij}-z_{kj})(z_{kj}-z_{ij}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l\neq i,k}}^{n}(z_{ij}-z_{lj})(z_{kj}-z_{lj})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(z_{ij}-z_{lj})^2 * \sum_{j=1}^{m}\sum_{l=1}^{n}(z_{kj}-z_{lj})^2\right]^{\frac{1}{2}}} \quad (12)$$

where:
z_{ij} (z_{kj}, z_{lj}) – i-th (k-th, l-th) normalized observation on j-th variable.
i, k, l – the numbers of the objects; $i, k, l \in \{1, \ldots, n\}$
j – the number of the variable; $j \in \{1, \ldots, m\}$

The silhouette index (SI), applied to assess the general quality of clustering, is defined as (Kaufman, Rousseeuw, 1990; Dudek, 2020):

$$SI = \frac{1}{n}\sum_{i=1}^{n}\frac{b(i)-a(i)}{\max\{a(i);\ b(i)\}} \quad (13)$$

where:
n – number of objects in dataset,
$a(i)$ - average distance from object i to other objects belonging to cluster P_r, (object i belongs to cluster P_r),
$b(i)$ - minimum of average distance from object i to other objects belonging to cluster P_s (object i does not belong to cluster P_s).

Therefore,

$$a(i) = \sum_{k\in\{P_r\setminus i\}} d_{ik}/(n_r - 1) \quad (14)$$

$$b(i) = \min_{s\neq r}\left\{\sum_{k\in P_s} d_{ik}/n_s\right\} \quad (15)$$

where:
$\{d_{ik}\}$ – distance matrix,
n_r – number of objects in cluster P_r,
n_s – number of objects in cluster P_s.

The silhouette index values, defined by formula (13), range from -1 to +1 (Kaufman, Rousseeuw, 1990). The SI score from 0.70 to 1.00 indicates very good clustering results with well-separated clusters. Values between 0.50 to 0.70 identify moderate clustering quality with some clusters well separated and the others overlapping. If the SI index is between 0.25 and 0.50, one can recognize poor clustering performance, where many objects are misclassified or clusters overlap. The SI score below 0.25 indicates a very poor clustering quality, suggesting that no substantial structure has been found.

Calculations were performed using the R environment (clusterSim package) and the STATISTICA PL v. 13.3 statistical package.

3. Results

All the calculations were based on the random sample coming from the Household Budget Survey (HBS) 2021 provided by Statistics Poland and being the main source of information on income of the population of households. The variable of interest was household equivalent income, with an equivalence scale established as the LIS (Luxembourg Income Study) scale (square root of the number of household members). The calculations were carried out taking into account the sampling weights (for details see: GUS, 2024).

In the empirical analysis, which adopted a complete-linkage clustering of the Polish regions, we included the following variables, which in our opinion comprise basic statistical characteristics of income distributions:

- Average equivalent income,
- Poverty head-count ratio,
- Poverty gap index,
- Gini index,
- Dispersion index for the end portions of the distribution.

The above-mentioned variables reflect various aspects of income distribution, i.e. its central tendency, dispersion and shape, with particular emphasis on how total income in a voivodeship is divided among households. The choice of the variables was supported by the considerations presented in the Introduction regarding possible (and sometimes ambiguous) interactions between the level of average income, inequality and poverty.

Based on complete-linkage clustering with the general distance measure (12), the optimal number of clusters obtained was 5. The value of the silhouette index equal to 0.64 confirms that a meaningful group structure was found.

Figure 1 presents the dendrogram obtained for the Polish voivodeships. The layout of the dendrogram tells us which voivodeships are most similar to each other. "Branch" height indicates how similar or different they are: the greater the height, the greater difference. The red line illustrates the division of the dendrogram into 5 clusters.

The clusters created as a result of multivariate statistical analysis reveal hidden similarities and differences between voivodships that may be the basis for interesting statistical and economic conclusions. The regions constituting the clusters are not always adjacent to each other, which is a natural consequence of the procedure based on economic rather than geographical characteristics. Cluster no. 1 comprises the Dolnośląskie and Mazowieckie voivodeships, cluster no. 2 the Opolskie, Pomorskie and Wielkopolskie voivodeships, cluster no. 3 the Kujawsko-Pomorskie, Lubelskie, Łódzkie, Podlaskie and Warmińsko-Mazurskie voivodeships, cluster no. 4 the Lubuskie, Małopolskie and Śląskie voivodeships, and cluster no. 5 the Podkarpackie, Świętokrzyskie and Zachodniopomorskie voivodeships.

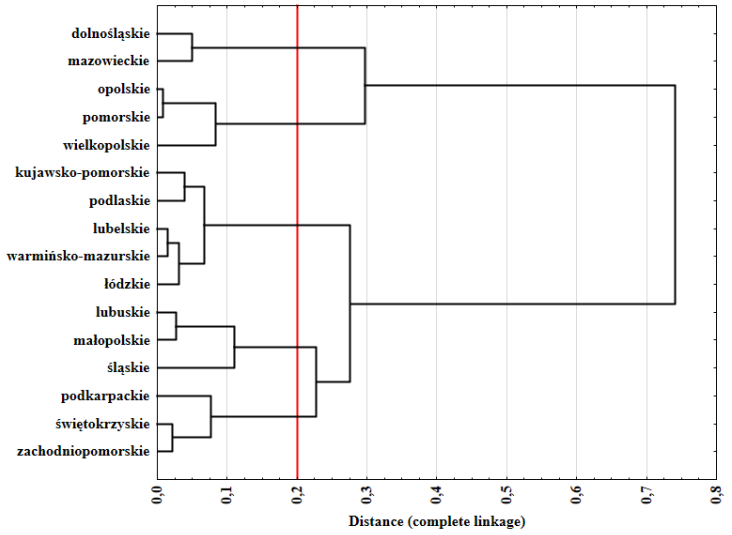


Figure 1: Dendrogram obtained for voivodeships

Source: authors' calculations

The identified clusters are visualized in Figure 2.

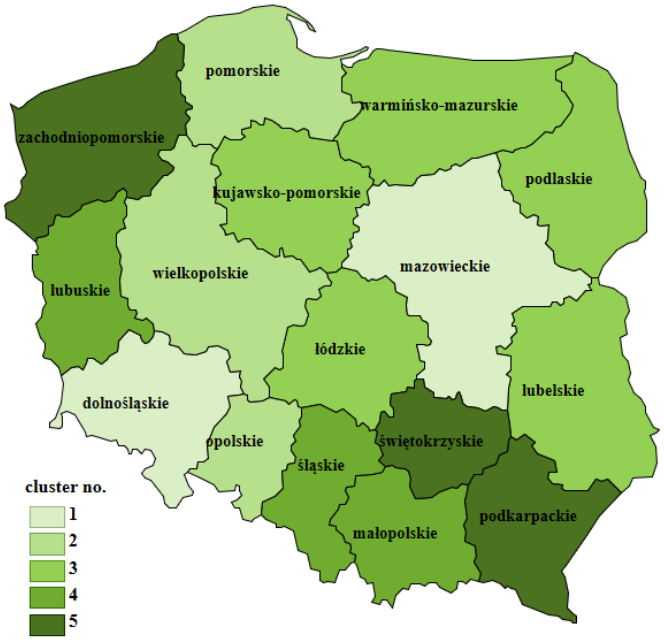


Figure 2: Visualization of the resulting clusters

Source: authors' calculations

The preliminary description of the clusters with regard to the analyzed variables is illustrated in a heatmap (Figure 3). In addition, descriptive statistics for the obtained clusters are presented in Table 1 and Figure 4.

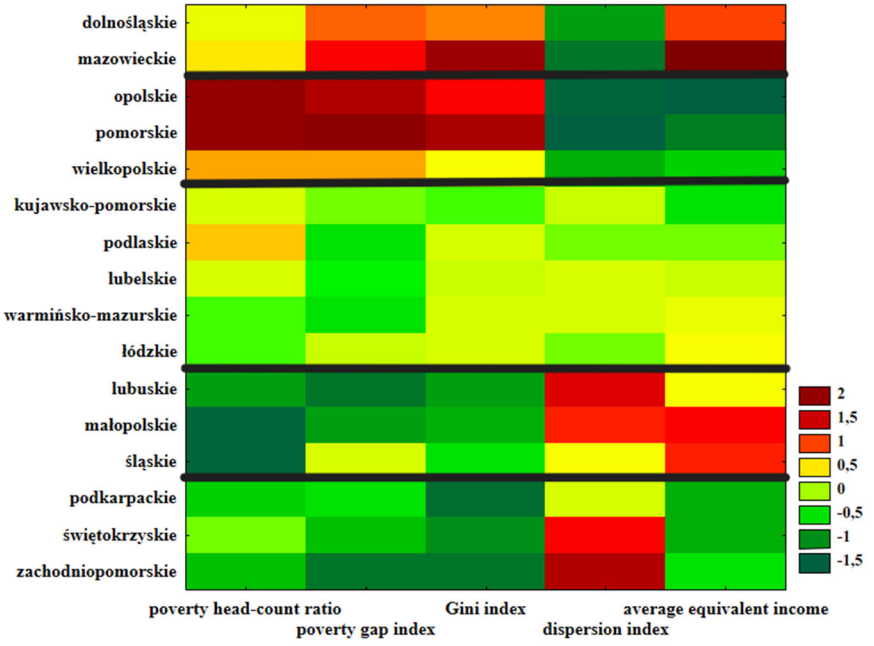


Figure 3: Heatmap – cluster characteristics

Source: authors' calculations

The legend of the heatmap (Figure 3) shows the boundaries of the intervals based on the standardized values of the variables taken into account (average equivalent income, poverty head-count ratio, poverty gap index, Gini index and dispersion index). The light colors indicate close-to-average values while dark red or dark green correspond to very high or very low levels of the variables for the selected regions, respectively.

The voivodeships in cluster no. 1 are characterized by moderate values of the poverty head-count ratio, high values of the poverty gap index and the Gini index, a low dispersion index and high average equivalent income.

The voivodeships in cluster no. 2 are described by high values of the poverty head-count ratio, the poverty gap index and the Gini index, as well as a low dispersion index and low average equivalent income.

The voivodeships in cluster no. 3 are characterized by an average level of all the indicators studied, except for the Wielkopolskie Voivodeship, which has higher values of the poverty head-count ratio and the poverty gap index compared to other voivodeships in this cluster.

The voivodeships in cluster no. 4 are described by low values of the poverty head-count ratio, poverty gap index and the Gini index, high dispersion index and high average equivalent income.

The voivodships included in cluster no. 5 are characterized by low values of poverty head-count ratio, the poverty gap index and the Gini index, as well as a fairly high dispersion index and low average equivalent income.

The conclusions drawn from the heatmap are supported by the values of the descriptive statistics presented in Table 1 and Figure 4.

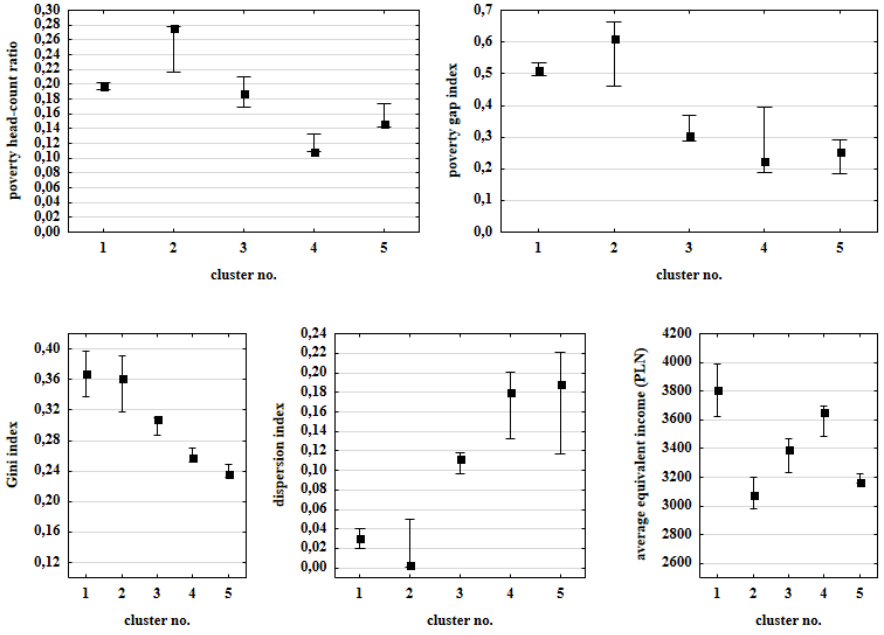


Figure 4: Comparison of clusters by poverty level and income inequality (median with range (min-max))

Source: authors' calculations

On average, the highest values of the poverty head-count ratio, the poverty gap index and the Gini index are observed for the voivodeships in cluster no. 2 (Opolskie, Pomorskie and Wielkopolskie voivodships); these voivodeships are also characterized by the lowest equivalent income on average.

On average, the highest values of the Gini index with a relatively high poverty gap index, are observed for voivodeships belonging to cluster no. 1 (Dolnośląskie and Mazowieckie voivodeships); these voivodeships are also characterized by the highest average equivalent income.

Table 1: Cluster characteristics – descriptive statistics

Variable	Measure	Cluster no.				
		1	2	3	4	5
		n=2	n=3	n=5	n=3	n=3
Poverty head-count ratio	mean	0.20	0.26	0.19	0.12	0.15
	SD	0.01	0.04	0.02	0.01	0.02
	CV	3.5%	13.7%	9.4%	11.1%	11.0%
Poverty gap index	mean	0.51	0.58	0.32	0.27	0.24
	SD	0.03	0.11	0.04	0.11	0.06
	CV	6.0%	18.1%	11.2%	40.7%	22.8%
Gini index	mean	0.37	0.36	0.30	0.26	0.24
	SD	0.04	0.04	0.01	0.01	0.01
	CV	11.5%	10.3%	3.1%	3.7%	3.7%
Dispersion index	mean	0.03	0.02	0.11	0.17	0.18
	SD	0.01	0.03	0.01	0.03	0.05
	CV	48.3%	156.4%	9.9%	20.4%	30.5%
Average equivalent income	mean	3808.76	3088.70	3380.18	3612.22	3185.19
	SD	258.74	109.09	95.87	114.84	37.94
	CV	6.8%	3.5%	2.8%	3.2%	1.2%

Note: SD - standard deviation; CV – coefficient of variation

Source: authors' calculations

The voivodeships in cluster no. 3 (Kujawsko-Pomorskie, Lubelskie, Łódzkie, Podlaskie and Warmińsko-Mazurskie) are characterized by an average level of all the examined variables.

The voivodeships belonging to cluster no. 4 (Lubuskie, Małopolskie and Śląskie voivodeships) and cluster no. 5 (Podkarpackie, Świętokrzyskie and Zachodniopomorskie voivodeships) are characterized by, on average, the lowest values of the poverty head-count ratio, the poverty gap index and the Gini index, as well as the highest values of the dispersion index. The compared clusters are distinguished from each other by the level of average equivalent income (high in cluster no. 4 and low in cluster no. 5).

The approximation of income distributions for voivodeships, carried out by means of the Dagum model, and the application of formulas (7)-(8) did not affect the classification of voivodeships, however, it smoothed out irregularities and shed light on differences in the location and shape of the compared subpopulations. Figure 5 shows the empirical and theoretical distributions (determined by formula (5)) for voivodeships representing individual clusters. Table 2 contains the parameters of the distributions adjusted to the empirical data, while Table 3 values of fit measures calculated on the basis of formulas (9)-(11).

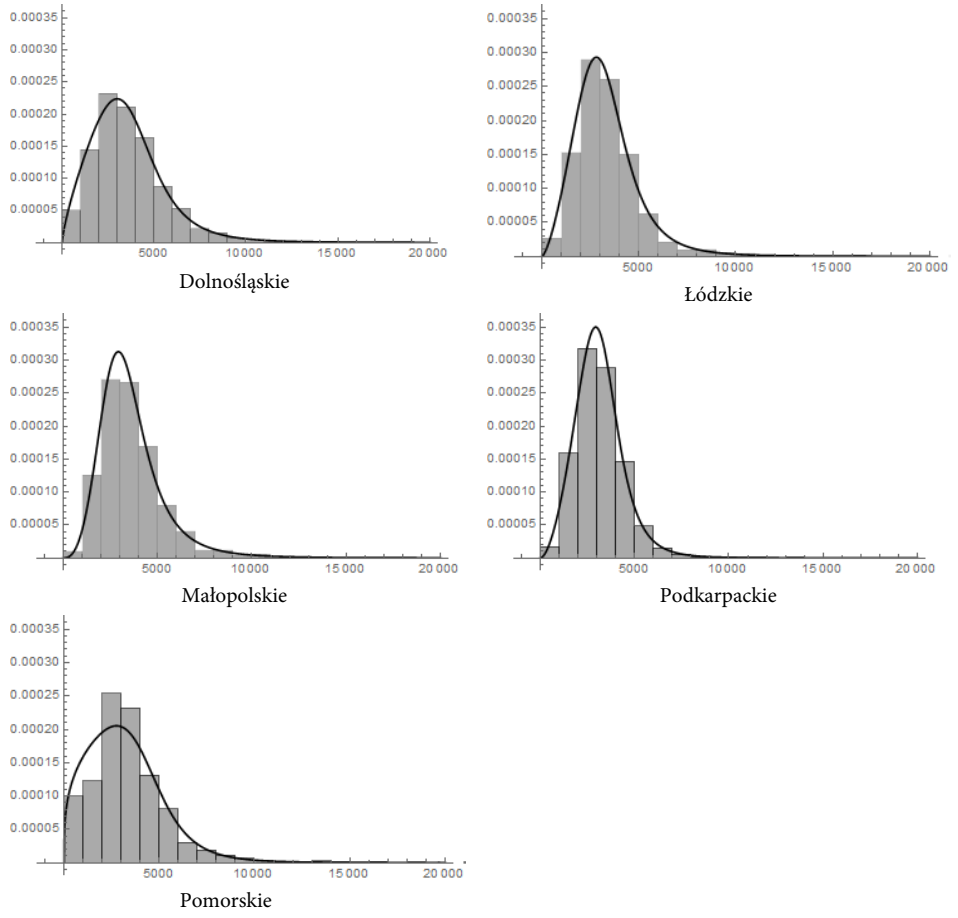


Figure 5: Income distributions in selected Polish voivodeships and fitted Dagum distributions

Source: authors' calculations

Table 2: Dagum distribution parameters for cluster representatives

Voivodeship	Dagum distribution parameters		
	a	***b***	***p***
Dolnośląskie	4.402	4667.61	0.403
Łódzkie	4.456	3775.38	0.554
Małopolskie	4.093	3456.44	0.868
Podkarpackie	5.815	3719.32	0.469
Pomorskie	4.964	4981.48	0.265

Source: authors' calculations

Table 3: Goodness-of-fit measures for cluster representatives

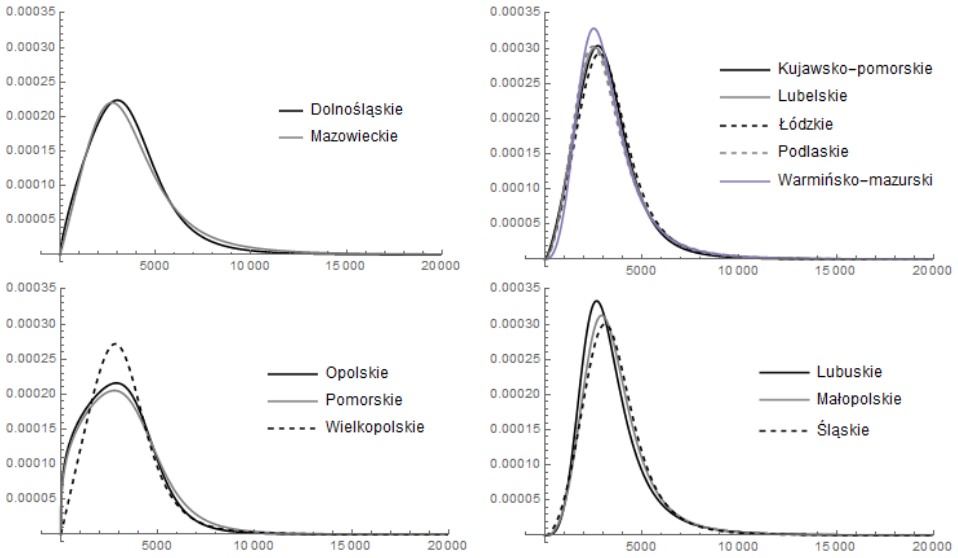
Voivodeship	Goodness-of-fit measure		
	Wps	A_1	A_2
Dolnośląskie	0.996	0.008	0.437
Łódzkie	0.984	0.031	1.572
Małopolskie	0.995	0.008	0.659
Podkarpackie	0.959	0.082	0.127
Pomorskie	0.930	0.139	0.228

Source: authors' calculations

On the basis of the goodness-of-fit measures presented in Table 3, it can be noted that the consistency of the empirical and theoretical distributions is satisfactory – the similarity coefficient of the distributions generally exceeds 0.95, i.e. the compared empirical and theoretical frequencies are very similar. The high goodness-of-fit of empirical and theoretical distributions is also confirmed by the low values of the A1 and A2 measures. The worst results were obtained for the Pomorskie voivodeship, which can also be seen in Figure 5.

Both the values of goodness-of-fit measures and the plots of empirical and theoretical distributions suggest good consistency of the income distributions of the representatives of each cluster with the Dagum model. Analogous results were obtained for the other voivodeships.

The plots of the Dagum densities estimated for the voivodeships constituting individual clusters are presented in Figure 6.



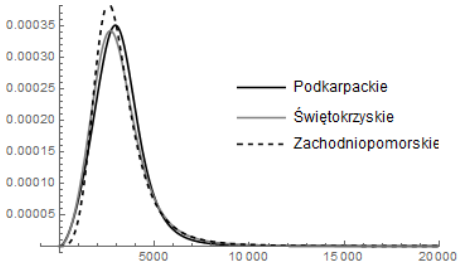


Figure 6. Dagum distributions estimated for voivodeships within the obtained clusters

Source: authors' calculations

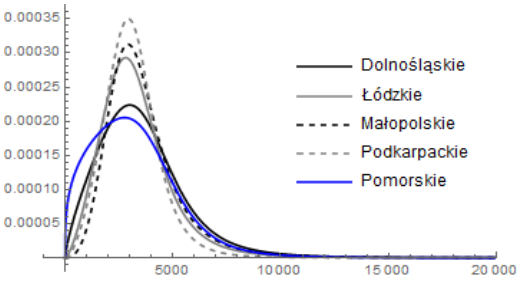


Figure 7: Dagum distributions estimated for cluster representatives

Source: authors' calculations

Figure 6 shows that the differences between the estimated income distributions for the members of each cluster are relatively small, but the differences between the income distributions for voivodeships from different clusters are substantial, as can be seen in Figure 7.

4. Conclusion

The aim of the paper was to reveal regional differences in income distributions in Poland, with a special attention paid to income inequality and poverty. The statistical analysis, based on individual data from the Household Budget Survey, applied selected non-parametric and parametric methods for estimating inequality and poverty measures, as well as the complete-linkage clustering for grouping the Polish voivodeships. The approximation of empirical distributions performed using the three-parameter Dagum model turned out to be an appropriate tool for both parametric estimation and visualization of the results. The estimated distribution characteristics enabled the effective segmentation of voivodeships, which was additionally confirmed by very similar Dagum density curves obtained for each cluster. On the contrary, the

differences between clusters are substantial. For the richest cluster, which consists of Dolnośląskie and Mazowieckie voivodeships, the highest poverty gap has been observed, which can be connected with discrepancies located at the left tail of income distributions. The most difficult situation was observed for the cluster containing: Opolskie, Pomorskie and Wielkopolskie - the cluster represents highest values of the average poverty head-count ratio, the poverty gap index and the Gini index and is also characterized by the lowest equivalent income. In the cluster including: Podkarpackie, Świętokrzyskie and Zachodniopomorskie, the poverty indices are small due to relatively small overall inequality, even if the average income level for this cluster is low.

The results of the calculations confirmed the assumption that income distributions in Poland differ significantly across regions. The obtained clusters allowed detecting groups of regions that may require separate social policies aimed at increasing household income, or rather at reducing income inequality.

References

Barca, F., (2009). An agenda for a reformed cohesion policy: A place-based approach to meeting European Union challenges and expectations, Independent Report prepared at the request of Danuta Hübner. *Commissioner for Regional Policy, EU Commission*, Brussels.

Brzezińska, J., (2018). Statistical Analysis of Economic Poverty in Poland Using R. *Econometrics. Advances in Applied Data Analysis*, 22(2), pp. 45–53.

Bonesmo Fredriksen, K., (2012). Income Inequality in the European Union. *OECD Economics Department Working Papers*, n. 952, *OECD Publishing*, Paris.

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

Dańska-Borsiak, B., (2024). Ocena adekwatności PKB per capita jako miary poziomu życia w powiatach. *Wiadomości Statystyczne. The Polish Statistician*, 69(7), pp. 1–21.

Deininger, K., Squire, L., (1998). New Ways of Looking at Old Issues: Inequality and Growth. *Journal of Development Economics,* 57(2).

Dudek, A., (2020). Silhouette Index as Clustering Evaluation Tool, (in) Jajuga K. et al. (eds.), Classification and Data Analysis, Studies in Classification, Data Analysis, and Knowledge Organization. *Springer*, pp. 19–33.

Eurostat, (2014). *Statistics in Focus 12/2014*. https://ec.europa.eu/eurostat/statistics-explained/.

Fei J., Ranis G., Kuo S., (1978). Growth and the Family Distribution of Income by Factor Components. *Quarterly Journal of Economics*, 92, pp. 17–53.

Gastwirth, J. L., (1975). Statistical Measures of Earning Differentials. *American Statistician*, 29, pp. 32–35.

Gini, C., (2012). Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche. *Studi Economico-Giuridici*, *Facoltà di Giurisprudenza della Regia Universit`a di Cagliari*, anno III, parte II, Cuppini, Bologna.

Gini, C., (1914). Sulla Misura Della Concentrazione e Della Variabilita` dei Caratteri, [w:] *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti. Anno Accademico 1913–1914*, Tomo LXXIII – Parte Seconda.

GUS, (2024|). Budżety gospodarstw domowych w 2023 roku. *Zakład Wydawnictw Statystycznych*, Warszawa. https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2023-roku,9,22.html.

Li, H., Squire, L. and Zou H., (1998). Explaining International and Intertemporal Variations in Income Inequality. *Economic Journal*, 108 (446), pp. 26–43.

OECD, (2008). Growinq Unequal: Income Distribution and Poverty in OECD Countries. *OECD Publishing*, Paris. https://www.oecd-ilibrary.org/social-issues-migration-health/growing-unequal_9789264044197-en.

OECD, (2011). Divided We Stand. Why Inequality Keeps Rising. *OECD Publishing*, Paris. https://www.oecd-ilibrary.org/social-issues-migration-health/the-causes-of-growing-inequalities-in-oecd-countries_9789264119536-en.

OECD, (2015). In It Together: Why Less Inequality Benefits All. *OECD Publishing*, Paris. https://www.oecd-ilibrary.org/employment/in-it-together-why-less-inequality-benefits- all_9789264235120-en.

OECD, (2024). Income and wealth inequalities, [in:] *Society at a Glance 2024: OECD Social Indicators.*

Jędrzejczak, A., (2011). Metody analizy rozkładów dochodów i ich koncentracji. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Jędrzejczak, A., (2015). Regional Income Inequalities in Poland and Italy. *Comparative Economic Research*, 18(4), pp. 27–45.

Jędrzejczak, A., Kubacki, J., (2017). Analiza rozkładów dochodu rozporządzalnego według województw z uwzględnieniem czasu, [in:] Klasyfikacja i analiza danych.

Teoria i zastosowanie. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia*, 29(469), pp. 69–81.

Jędrzejczak, A., (2023). Klasyczne i nieklasyczne metody analizy nierówności dochodowych. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Jędrzejczak, A., Pekasiewicz, D., (2020). Teoretyczne rozkłady dochodów gospodarstw domowych i ich estymacja. *Wydawnictwo Uniwersytetu Łódzkiego*, Łódź.

Jędrzejczak, A., Pekasiewicz, D., (2022). Nierównomierność ekwiwalentnych dochodów gospodarstw domowych w województwie łódzkim. *Wiadomości Statystyczne*, 67(6), pp. 29–51.

Kaufman, L., Rousseeuw, P. J., (1990). Finding groups in data: an introduction to cluster analysis. *John Wiley*, New York.

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

Malina, A., (2020). Analiza przestrzennego zróżnicowania poziomu rozwoju społeczno-gospodarczego województw Polski w latach 2005–2017. *Nierówności Społeczne a Wzrost Gospodarczy*, No. 61(1), pp. 138–155.

Panek, T., (2011). Ubóstwo, wykluczenie społeczne i nierówności. Teoria i praktyka pomiaru. *Oficyna Wydawnicza SGH*, Warszawa.

Sen, A., (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44(2), pp. 219–231.

Walesiak, M., Dudek, A., (2020). The Choice of Variable Normalization Method in Cluster Analysis, In Soliman KS (ed.), *Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development During Global Challenges*, pp. 325–340, ISBN 978-0-9998551-4-1.

Walesiak, M., (2016). Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R. *Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu*, Wrocław.

Zenga, M. M., Pasquazzi, L. and Zenga, M., (2010). Rapporto n. 188: First applications of a new three-parameter distribution for non-negative variables. Mediolan: *Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano Bicocca*.