ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 225, 2009

*Alina Jędrzejczak**

COMPARING INCOME DISTRIBUTIONS – METHODS AND THEIR APPLICATION TO WAGE DISTRIBUTIONS IN POLAND

Abstract

Rankings of income distributions are usually based on comparisons of social welfare. Assuming more or less general form of a social welfare function we can compare income distributions in time and in space. Income inequality can be compared by means of the well known Lorenz curve but the results will be ambiguous when the Lorenz functions of the considered populations intersect. Generalized Lorenz curves and quantile functions are more useful tools for ranking income distributions but in many situations it is necessary to make additional assumptions concerning social preferences reflected in a social welfare function. In the paper we present the methods useful for ranking income distributions and their application to the analysis of wage distributions in Poland. As a theoretical distribution the Dagum type-I model has been used.

Key words: income distribution, economic distance, social welfare, inequality.

1. Introduction

Comparing income distribution in time and in space we should take into consideration the average level of income and the inequality of its distribution. The Gini ratio is said to be the best synthetic measure of income inequality. It can be expressed in terms of Gini mean difference being the measure of absolute dispersion within a distribution. Economic distance between two populations can be evaluated by means of the coefficients proposed by Dagum which are connected with the concept of Gini mean difference between distributions. To order income distributions from a point of view of social welfare the methods based on Lorenz curves, generalized Lorenz curves and quantile functions could be useful. In the paper we present the methods and their applications to the analysis of wage distributions in Poland in 1999 and 2003.

* Dr., Chair of Statistical Methods, University of Łódź.

2. Income distribution model

In many situations it seems reasonable to use theoretical income distributions, which show high consistency with the empirical ones. First, such an approach allows for the flattening of irregularities in empirical distributions coming from the method of gathering information. Second, the use of a theoretical model simplifies and accelerates the analysis because all distribution characteristics can be expressed by the same parameters. Moreover, the maximum likelihood and ordinary least squares estimates of inequality measures can be provided easily, given the mathematical form of a density function or a cumulative distribution function.

A variety of probability functions has been suggested as suitable in describing the distributions of income by size. The lognormal distribution has been widely used in wage and income distribution analysis for many years. The advantage of this distribution is its simplicity; a disadvantage, however, is its poor fitting to the data, especially in the tails.

Unlike the lognormal, the Dagum model was based on empirical observations of income distributions made in many countries. Dagum (1977) and Dagum and Lemmi (1989) noted that the function describing income elasticity of a cumulative distribution function of income is convex, decreasing and bounded. It can be described by the following differential equation:

$$\varepsilon(y,F(y))=\frac{d\ln F(y)}{d\ln y}=\beta_1\left[1-[f(y)^{\beta_2}]\right] \tag{1}$$

for $y\geq 0, \beta_1, \beta_2>0$.

The cumulative distribution function of the Dagum model is the solution of the equation given by formula (1). It can be written as follows:

$$F(y)=(1+\lambda y^{-\delta})^{-\beta}, y>0 \tag{2}$$

for $\beta, \lambda, \delta>0$,

where:

$\beta=1/\beta_1$, $\delta=\beta_1\beta_2$, $\lambda=\exp c$,

c – a constant of integration resulting from the solution of equation (2).

Parameters β and δ are inequality parameters of the Dagum distribution while λ is a parameter of scale.

The moments of order r about the origin corresponding to the model (2) known as Dagum type I distribution, are specified by the equation:

$$\mu_r=\beta\lambda^{r/\delta}B(1-r/\delta, \beta+r/\delta) \text{ for } r<\delta \tag{3}$$

where:

$B(1-r/\delta, \beta+r/\delta)$ – the beta function with parameters $(1-r/\delta, \beta+r/\delta)$.

It follows from equation (3) that the moments of order r exist only for $r<\delta$. Hence, the moments of orders $r \geq \delta$ are infinite.

The Lorenz curve corresponding to the cumulative distribution function (2) can be written as follows:

$$L(p) = B^*\left[p^{1/\beta}; \beta+1/\delta, 1-1/\delta\right] \quad (4)$$

for $\delta > 1, 0 \leq p \leq 1$,

where:

$B^*\left[p^{1/\beta}; \beta+1/\delta, 1-1/\delta\right]$ – the incomplete beta function.

The Gini concentration coefficient obtained on the basis of equation (4) has the form:

$$G = -1 + B(\beta, \beta) / B(\beta, \beta+1/\delta) \quad (5)$$

where:

$B(\ldots)$ – the beta function.

Table 1

ML estimates of the Dagum model parameters

Province	Dagum model parameters					
	1999			2003		
	λ	β	δ	λ	β	δ
1. Dolnośląskie	0.1736	2.1976	3.6100	0.7568	1.0293	3.5446
2. Kujawsko-pomorskie	0.0768	3.6259	3.5400	0.5348	1.1027	3.9000
3. Lubelskie	0.0745	3.9759	3.5233	0.5088	1.1140	3.5602
4. Lubuskie	0.0676	4.7569	3.4409	0.7072	0.9367	4.1117
5. Łódzkie	0.0815	4.3026	3.4712	0.5581	1.0780	3.7171
6. Małopolskie	0.1621	2.1841	3.9523	0.8791	0.8353	4.1483
7. Mazowieckie	0.0795	6.8175	2.8504	0.3500	2.3970	2.9125
8. Opolskie	0.1120	2.7867	4.0521	0.8945	0.8211	4.1355
9. Podkarpackie	0.1242	1.7153	4.6150	0.6148	1.0668	4.4677
10. Podlaskie	0.1091	2.6964	3.8365	0.6473	1.1874	3.7817
11. Pomorskie	0.0813	5.3130	3.1117	0.5340	1.1746	3.4960
12. Śląskie	0.2728	1.8691	3.5798	0.9569	0.9756	3.7829
13. Świętokrzyskie	0.1740	1.8106	3.9608	0.4274	1.2314	4.2079
14. Warmińsko-mazurskie	0.1731	2.0360	3.8543	0.6656	0.9939	3.8837
15. Wielkopolskie	0.2531	1.4801	3.9566	0.7403	1.0032	3.7227
16. Zachodniopomorskie	0.1201	3.4881	3.3130	0.6177	1.3236	3.8933

Source: author's calculation.

Table 2

Distribution consistency measures

Province	Consistency measures			
	1999		2003	
	S_d	OLM	S_d	OLM
1. Dolnośląskie	0.9442	0.0123	0.9206	0.0174
2. Kujawsko-pomorskie	0.9284	0.0187	0.9475	0.0112
3. Lubelskie	0.9674	0.0076	0.9534	0.0160
4. Lubuskie	0.9776	0.0055	0.9080	0.0188
5. Łódzkie	0.9471	0.0122	0.9217	0.0196
6. Małopolskie	0.9622	0.0091	0.9558	0.0103
7. Mazowieckie	0.9321	0.0153	0.9026	0.0227
8. Opolskie	0.9547	0.0132	0.9210	0.0173
9. Podkarpackie	0.9726	0.0071	0.9544	0.0117
10. Podlaskie	0.9540	0.0126	0.9565	0.0114
11. Pomorskie	0.9323	0.0149	0.3012	0.0208
12. Śląskie	0.9451	0.0125	0.9209	0.0189
13. Świętokrzyskie	0.9454	0.0130	0.9398	0.0124
14. Warmińsko-mazurskie	0.9507	0.0121	0.9219	0.0179
15. Wielkopolskie	0.9216	0.0173	0.9156	0.0182
16. Zachodniopomorskie	0.9340	0.0148	0.9007	0.0214

S o u r c e: author's calculation.

Wage distributions in Poland by regions in 1999 and 2003 were approximated by means of the Dagum type-I model. The basis for the calculations was continuous data obtained from the Labour Force survey conducted by Polish Central Statistical Office. The parameters of the theoretical distributions were estimated by means of the maximum likelihood method. To find the maximum of the logarithm of the likelihood function for the Dagum model an individual numerical procedure has been applied. The results of the estimation are presented in Tables 1 and 2. In Table 1 there are distribution consistency measures: the overlap measure (OLM) and the standard deviation of relative frequencies (*Sd*). They were calculated to compare the goodness-of-fit of the Dagum distribution and the lognormal distribution with the corresponding wage distributions in Poland. The coefficient of distribution similarity, called the overlap measure, was proposed by V i e l r o s e (1960). It can be calculated as a sum of smaller frequencies, taking into account empirical and theoretical frequencies for the same class boundary:

$$OLM = \sum_{i=1}^{k} \min[\gamma_i; \hat{\gamma}_i] \tag{6}$$

where:

γ_i – empirical frequency,

$\hat{\gamma}_i$ – theoretical frequency,

k – number of income intervals.

The bigger the value of OLM, the higher the consistency of compared distributions. Analyzing both the measures one can easily notice that the distributions estimated by means of the Dagum function show generally high consistency with the empirical distributions. It is worth mentioning that the goodness-of-fit with lognormal model was poor for all the distributions under consideration.

3. Comparisons of wage distributions in Poland by means of economic distance measures

Comparing two populations of economic units, differing by socio-economic characteristics, we can investigate the degree of income inequality within each of these populations, using for instance the Gini coefficient. We can also evaluate the degree of affluence of one population with respect to another using the statistics introduced by D a g u m (1980) and called economic distances.

The economic distance d_0 between the income distribution X with the probability density function $f_1(x)$ and the cumulative distribution function $F_1(x)$, and the distribution Y with the density function $f_2(y)$ and cumulative distribution function $F_2(y)$ is defined as probability that income Y is greater than income X, given that $E(Y)$ is greater than $E(X)$:

$$d_0 = P\{Y > X \mid E(Y) > E(X)\} = \int_0^{\infty}\int_0^{y} dF_1(x)dF_2(y) = E[F_1(Y)] \tag{7}$$

where:

$E(Y)$, $E(X)$ – expected values of random variables X and Y.

The economic distance d_1 between income distributions $f_1(x)$ and $f_2(x)$ is defined as the weighted sum of the income difference Y–X for all Y>X given that $E(Y)$ is greater than $E(X)$:

$$d_1 = \int_0^{\infty}\int_0^{y} (y-x)dF_1(x)dF_2(y) = E[YF_1(Y)] + E[XF_2(X)] - E(X) \tag{8}$$

Measure d_1 can be also generalized in a non-linear form:

$$d_r = \left[\int_0^\infty \int_0^y [(y-x)^r dF_1(x) dF_2(y)] \right]^{1/r}, \quad r \neq 0 \tag{9}$$

All the members of equation (9) are weighted averages of income differences $y - x$ for $y > x$, given that $E(Y) > E(X)$. Hence, when $r = 1$ (9) is a conditional arithmetic mean, when $r = 2$ – a conditional quadratic mean and for $r = -1$ it can be regarded as a conditional harmonic mean. Taking big values of d_r, we enhance greater income differences, but when r tends to minus infinity, d_r will be dominanted by small income differences. In developing countries using economic distances of higher order which reflect bigger differences between income distributions could be useful.

The normalized forms of d_0 and d_1 are the following measures called economic distance ratios:

$$D_0 = 2d_0 - 1$$

$$D_1 = [E(Y) - E(X)]/[2d_1 - E(Y) + E(X)] \tag{10}$$

The economic distance ratios measure the proportion by which the more affluent population is better off than the other. The values „0" are taken when income variables X and Y are independent and identically distributed. This implies that there is no economic distance between the two populations. D_0 and D_1 take their maximum value „1" when the population Y is by 100% better than X (the two populations do not overlap). That means that each member of the more affluent population Y has higher income than any member of the population X. The economic distance ratio D_1 measures not only the frequency but also the amount by which the incomes of the two populations differ. Therefore, it is sensitive to any changes in means, variances and asymmetry of the compared distributions.

Table 3 contains maximum likelihood estimates of expected values and Gini ratios for the wage distributions by regions in 1999 and 2003. Moreover, in the table there are ratios D_0 and D_1 measuring the economic affluence of wage distributions in 2003 with respect to 1999. All the distribution characteristics were calculated on the basis of the Dagum model parameters (see (3), (5), (10)).

The highest values of economic distances were observed for "podkarpackie" and "podlaskie" provinces. For example wages in "podkarpackie" were better in 2003 by 31,88%, taking into account only the probability of getting higher income, and by 47.53% better taking into account also the amount of income differences.

Table 3

Characteristics of wage distributions by regions in 1999 and 2003 [a]

Province	Expected value (in thousands PLN)		Gini ratio		Economic distance	
	1999	2003	1999	2003	D_0	D_1
1. Dolnośląskie	0.9230	1.0694	0.3188	0.3895	**0.1613**	**0.2763**
2. Kujawsko-pomorskie	0.8610	0.9847	0.2338	0.2505	**0.1933**	**0.2715**
3. Lubelskie	0.8783	0.9848	0.2334	0.2740	**0.1277**	**0.2232**
4. Lubuskie	0.9029	0.9925	0.2367	0.2474	**0.1549**	**0.1939**
5. Łódzkie	0.9235	0.9929	0.2360	0.2643	**0.0914**	**0.1449**
6. Małopolskie	0.9054	1.0038	0.2188	0.2531	**0.1338**	**0.2149**
7. Mazowieckie	1.0991	1.2318	0.2861	0.2995	**0.1336**	**0.1983**
8. Opolskie	0.8854	1.0022	0.2073	0.2552	**0.1540**	**0.2605**
9. Podkarpackie	0.8040	0.9953	0.1930	0.2202	**0.3188**	**0.4753**
10. Podlaskie	0.8689	1.0665	0.2204	0.2542	**0.2652**	**0.4081**
11. Pomorskie	0.9971	1.0210	0.2626	0.2759	**0.0406**	**0.0448**
12. Śląskie	0.9927	1.1018	0.2481	0.2660	**0.1341**	**0.2009**
13. Świętokrzyskie	0.8697	0.9612	0.2243	0.2262	**0.1595**	**0.2195**
14. Warmińsko-mazurskie	0.8998	1.0045	0.2269	0.2579	**0.1553**	**0.2239**
15. Wielkopolskie	0.8974	1.0430	0.2322	0.2684	**0.1691**	**0.2923**
16. Zachodniopomorskie	0.9703	1.0876	0.2519	0.2412	**0.1949**	**0.2291**

[a] all statistical characteristics were calculated on the basis of the Table 1.

Source: author's calculations.

On the other hand, the wage distributions in 1999 and 2003 for "pomorskie" are almost identical. It is connected with similar levels of means and concentration coefficients.

4. Application of ranking procedures based on a social welfare function

Using economic distance measures introduced above we assume, that the population with greater mean income is "better" than the population with the lower mean. This approach is connected with a widely accepted social preference for efficiency also called poverty aversion. It is well known that social welfare depends not only on the total mass of income but on its distribution between economic units as well. Comparing income distributions from a point of view of social welfare we must take into consideration both the level of mean income and the level of income inequality. Assuming more or less general form of a social welfare function it is possible to find compromise between efficiency and equity preference of a population of income receivers.

Let us suppose that a social ordering of income distributions can be represented by the following welfare function:

$$W = \int_0^\infty U(y)f(y)dy \tag{10}$$

where: $f(y)$ denotes a density function of income and $U(y)$ represents a utility function of income, usually assumed to be increasing and concave.

The partial ordering of income distributions can be based on the following theorem (Atkinson, 1970).

Theorem 1. Let $f_A(y)$ and $f_B(y)$ denote the density functions of income distributions A and B, $L_A(p)$ and $L_B(p)$ their corresponding Lorenz curves, μ_A and μ_B their mean incomes. For any strictly concave utility function $U(y)$:

$$\text{If } \mu_A = \mu_B \text{ then } \int_0^\infty U(y)f_A(y)dy \geq \int_0^\infty U(y)f_B(y)dy \Leftrightarrow L_A(p) \geq L_B(p)$$

for all $p \in \langle 0,1 \rangle$.

The distribution *A* dominates *B* if and only if the Lorenz curve for *A* lies above the Lorenz curve for *B*. When the curves intersect it is impossible to make decision without further assumptions on a utility function.

Better ordering tools can be based on generalized Lorenz curves obtained by scaling up the ordinary Lorenz curve by the mean income (Shorrocks, 1983). They enable to compare distributions with different means, taking into account widely accepted "efficiency preference". The generalized Lorenz dominance criterion is equivalent to the second-order stochastic dominance.

Theorem 2. Let $GL_A(p)$ and $GL_B(p)$ denote generalized Lorenz curves corresponding to the density functions $f_A(y)$ and $f_B(p)$. For any increasing and strictly concave utility function:

$$\int_0^\infty U(y)f_A(y)dy \geq \int_0^\infty U(y)f_B(y)dy \Leftrightarrow GL_A(p) \geq GL_B(p)$$

for all $p \in \langle 0,1 \rangle$.

Theorem 2 provides the partial ordering of income distributions with different means on condition that generalized Lorenz curves do not intersect. For complete ordering, a cardinal social welfare function that assigns numerical values to all possible social states could be useful.

More basic and less restrictive dominance principle, based on strong Pareto law, was proposed by Saposnik (1981). It is called rank dominance and is equivalent to the first – order stochastic dominance.

Theorem 3. Let $Y_A(p)$ and $Y_B(p)$ denote the quantile functions of income distributions A and B. For any increasing and anonymous welfare function W:

$$\int_0^\infty U(y) f_A(y) dy \geq \int_0^\infty U(y) f_B(y) dy \Leftrightarrow Y_A(p) \geq Y_B(p)$$

for all $p \in \langle 0,1 \rangle$.

Bishop, Formby and Thistle (2) showed that much of power contained in generalized Lorenz dominance criterion is contained in comparisons of quantile functions.

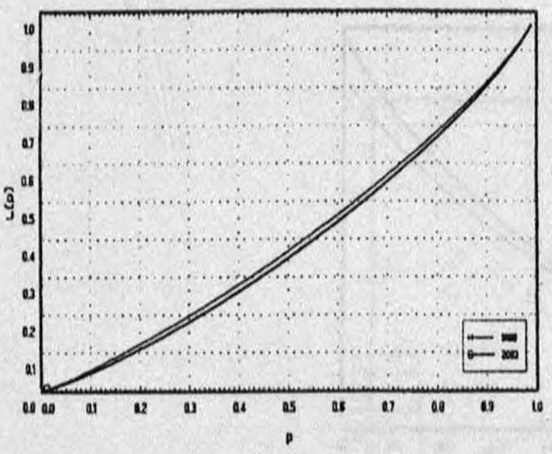

Fig. 1a. Lorenz curves based on wage distributions for podkarpackie

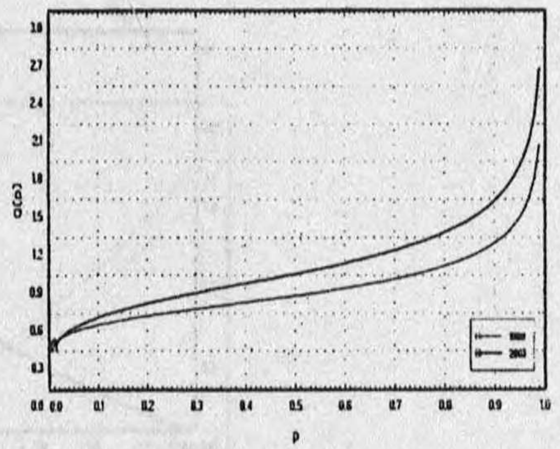

Fig. 1b. Quantile functions based on wage distributions for podkarpackie

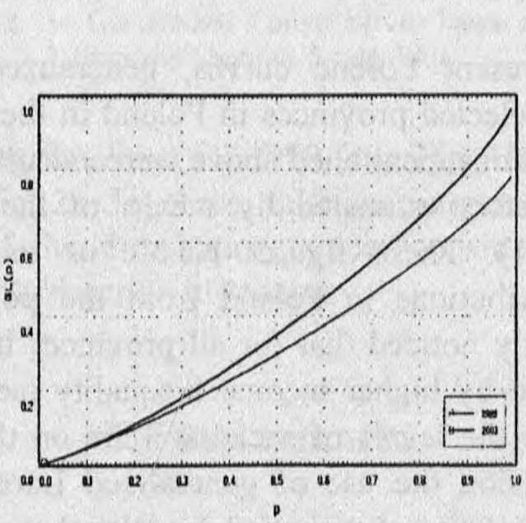

Fig. 1c. Generalized Lorenz curves based on wage distributions for podkarpackie

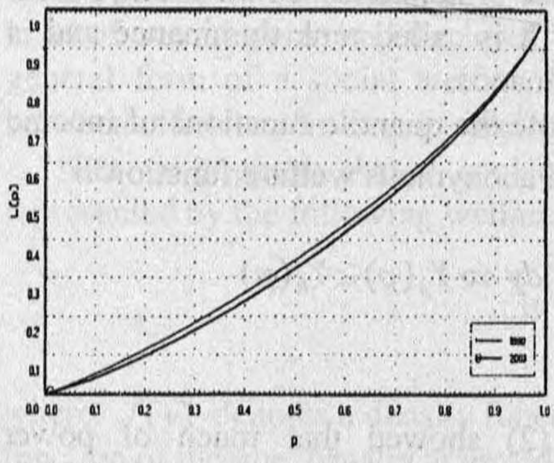

Fig. 2a. Lorenz curves based on wage distributions for warmińsko-mazurskie

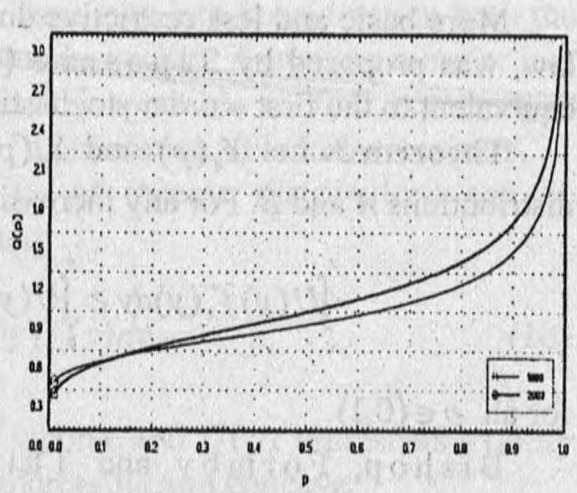

Fig. 2b. Quantile functions based on wage distributions for warmińsko-mazurskie

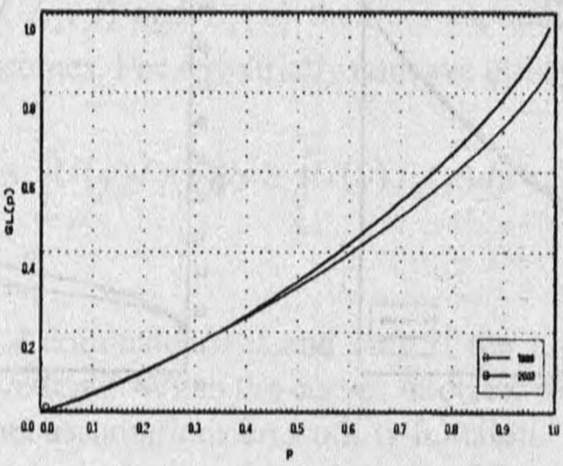

Fig . 2c. Generalized Lorenz curves based on wage distributions for warmińsko-mazurskie

Figures 1a–3c present Lorenc curves, generalized Lorenc curves and quantile function for selected provinces in Poland in the years 1999 and 2003. The values of the functions mentioned above were calculated on the basis of the Dagum model parameters estimated by means of the maximum likelihood method (see: Table 1). These figures enable us to analyze the changes concerning wage distributions in Poland from the point of view of social welfare. It can be easily noticed that for all provinces higher mean income in 2003 was accompanied by higher income inequality (see Table 3). Thus it is impossible to compare the levels of social welfare on the basis of the Lorenz curves. To make decision the use of generalized Lorenz curves or quantile functions (inverse cumulative distribution functions) was necessary. For most provinces the generalized Lorenc curves and quantile functions for 2003 lie

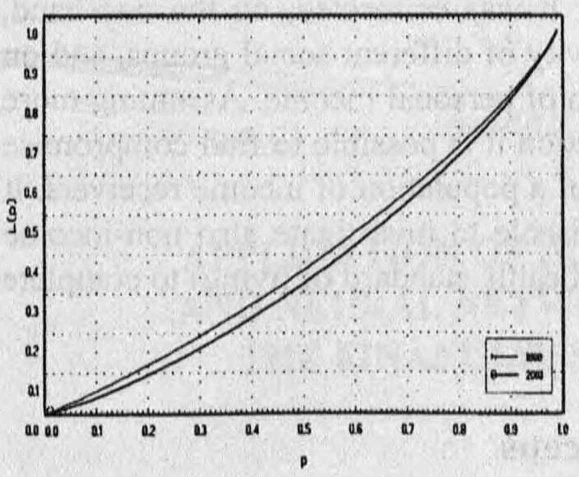

Fig. 3a. Lorenz curves based on wage distributions for opolskie

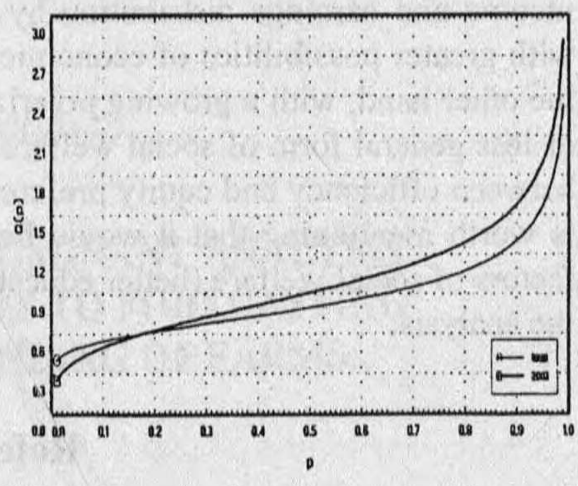

Fig. 3b. Quantile functions based on wage distributions for opolskie

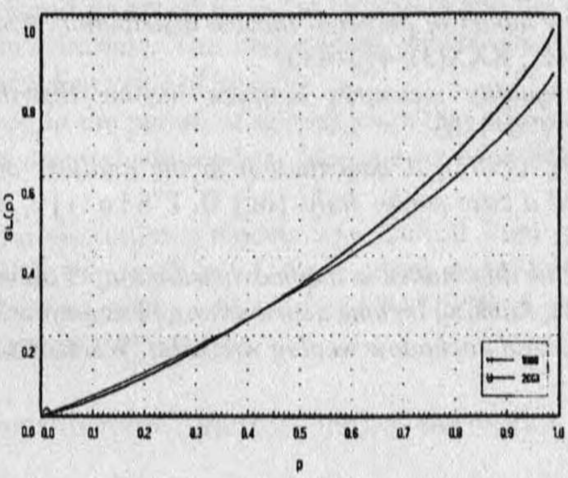

Fig. 3c. Generalized Lorenz curves based on wage distributions for opolskie

above the corresponding functions for 1999 (see. Fig. 1b–1c). Nevertheless, for some provinces (see Fig. 2a–3c) these functions intersect so it is impossible to derive the ordering of the distributions under consideration without further assumptions on the utility function of income.

5. Final remarks

Ranking of income distributions based on social welfare functions can be very useful in the analysis of wage and income distributions in Poland. The period of economic transformation (since 1990) was characterized by a series of fundamental changes in economy. The changes influenced, among other things, income and earnings distribution by size. It was connected, on the one hand, with greater possibilities of economic activity of different social groups, and on the other hand, with a growing polarization of personal income. Assuming more or less general form of social welfare function it is possible to find compromise between efficiency and equity preference of a population of income receivers. It is worth mentioning that it would be advisable to investigate also non-income factors of social welfare (better education, health, standard of living) to complete the analysis.

Referencens

Atkinson A. B. (1970), *On the measurement of inequality*, "Journal of Economic Theory", **2**, 244–263.

Bishop J. A., Formby J. P., Thistle P. D., (1991), *Rank dominance and international comparisons of income distributions*, "European Economic Review", **35**, 1399–1410.

Dagum C. (1977), *A new model of personal income distribution. specification and estimation*, "Economicé Appliqueé", **XXX**(3), 413–436.

Dagum C. (1980), *Inequality measures between income distributions with application*, "Econometrica", **48**, 1970–1803.

Dagum C., Lemmi A. (1989), *A contribution to the analysis of income distribution and income inequality and a case study: Italy*, [in:] D. I. Slottje, *Advances in econometries*, Yai Press, Greenwich.

Saposnik R. (1981), *Rank dominance in income distribution*, "Public Choice", **36**, 147–151.

Shorrocks A. F. (1983), *Ranking income distributions*, "Economica", **50**, 3–17.

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

Alina Jędrzejczak

Porównywanie rozkładów dochodów – metody i ich wykorzystanie do analizy rozkładów płac w Polsce

Porównywanie rozkładów dochodów może być przeprowadzane na podstawie różnych kryteriów. Jedną z metod jest zastosowanie współczynników dystansu ekonomicznego. Porównując rozkłady dochodów z punktu widzenia zamożności rozważanych populacji, bierzemy pod uwagę dwa aspekty tego zagadnienia – różnice w średnich poziomach dochodów oraz różnice w ich nierównomierności. Porównywanie nierównomierności może się odbywać za pomocą krzywych Lorenza – wyniki będą jednak niejednoznaczne, np. gdy funkcje Lorenza przecinają się. Bardziej uniwersalnym narzędziem do rangowania rozkładów dochodów są uogólnione funkcje Lorenza oraz funkcje kwantylowe. Wymagają one jednak przyjęcia pewnych dodatkowych założeń dotyczących funkcji dobrobytu, która jest wyrazem preferencji społecznych. W artykule przedstawiono wyniki zastosowania powyższych metod dla rozkładów płac w Polsce. Jako rozkład teoretyczny wykorzystany został rozkład Daguma pierwszego typu.