ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 141, 1997

*Alina Jędrzejczak**

APPLICATION OF DAGUM ECONOMIC DISTANCES TO THE ANALYSIS OF WAGE DISTRIBUTIONS IN POLAND

Abstract. Studies in the field of wage and income distribution concentrate on measuring income inequality.

We can investigate both the inequality within the population and the inequality between the populations. The latter can be called the economic distance of one population with respect to another. In the paper we present economic distance ratios. D_0 and D_1 introduced by Dagum. They were calculated for the empirical distributions of wages in Poland (nonparametric form) and for the corresponding theoretical ones derived from the Dagum function (parametric form).

Key words: income distribution, inequality, approximation.

1. INTRODUCTION

In the comparative analysis of wage and income distributions in time and in space synthetic inequality measures are often used. These measures are very popular mainly because of simple economic interpretation. Using them we can evaluate in which groups of economic units income concentrates and it may be a starting point for the construction of global poverty indices. It is well known that the level of social welfare depends not only on the total mass of income but also on its distribution among economic units.

In the paper we present the inequality measures introduced by Dagum (1980) with the application to the analysis of wage distributions in Poland.

* University of Łódź, Chair of Statistical Methods.

2. INEQUALITY MEASURES OF INCOME DISTRIBUTION

Since Pareto started his formal quantitative research on personal income distribution, studies in this field have been mainly concentrated on the following two areas:

– creation of an accurate and elementary mathematical model describing personal income distributions,

– measurement of the degree of income inequality.

Most often used inequality measures were derived from the Lorenz curve. They differ from one another in the method of measuring the area of concentration, that is the area between the line of equal share and the Lorenz curve given by the equation:

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

where:

p – fraction of the lowest elements of the population,

F^{-1} – the population p^{th} quantile,

μ – the expected value of income distribution.

Many authors (for example Morgan (1962), Gastwirth (1972)) consider the Gini coefficient to be the best single measure of income inequality. Gini defined his index (see Gini 1912) as the relation of the Gini mean difference Δ to the double expected value of income distribution:

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

where Δ can be defined as:

$$\Delta = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x - y| dF(x)dF(y) \tag{3}$$

or:

$$\Delta = \sum_{i \neq j}\sum \gamma_i \gamma_j |\mu_i - \mu_j| \tag{4}$$

where:

X, Y – two income variables identically distributed,

μ_i, μ_j – middles of income itervals,

γ_i, γ_j – frequencies in the intervals.

As we see Gini mean difference is the weighted arythmetic mean of all possible absolute values of income differencies. The weighting factor is the joint probability density function $f(x)$ $f(y)$. The Gini index described by formula (2) is the quotient of the measure of dispersion (Δ) to the arithmetic mean. That means that G can be also regarded as a measure of relative dispersion.

The Gini coefficient of concentration can be also interpreted from the point of view of interpersonal comparisons (Pyatt 1976). In this sense Gini index is the avarage gain to be expected if each individual has the choice of being himself or some other member of the population drawn at random, expressed as the proportion of the avarage level of income. Sen (1973) interprets the Gini index in the similar way but he underlines not possible benefits as rather existing deprivation of the poor in comparison with the rich. This approach was the base for construction of Sen's well known poverty index (Sen 1976).

In order to count the Gini index from grouped data the following formula, introduced by Morgan (1962), is often used:

$$G = 1 - \sum_{i=1}^{k} [F(y_{i+1}) - F(y_i)][L(y_i) + L(y_{i+1})] \tag{5}$$

where:

$F(y)$ – cumulative distribution function

$L(y)$ – Lorenz function given by the equation (1).

The Gini coefficient of concentration takes values in the unit interval. It takes value "0" when all units have the same level of income (concentration does not exist), and value "1" when all members of the population except one have zero income.

3. INEQUALITY MEASURES BETWEEN INCOME DISTRIBUTIONS

Comparing two populations of economic units, differing with 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 affluance of one population with respect to another using the statistics introduced by Dagum (1980) and called economic distances or inter income inequality measures.

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 the cumulative distribution function $F_2(y)$ is defined as the probability that income Y is greater than income X, given that $E(Y)$ is greater than $E(X)$:

$$d_0 = P\{Y > X | E(Y) > E(X)\} = \int_0^{\infty}\int_0^{y} dF_1(x)dF_2(y) = E[F_1(Y)] \quad (6)$$

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

The economic distance d_0 between income distributions $f_1(x)$ and $f_2(y)$ is defined as the weighted sum of the income difference $Y-X$ for all $Y > X$ given that $E(Y)$ is greater than $E(X)$:

$$d_1 = \int_0^{\infty}\int_0^{y} (y-x)dF_1(x)dF_2(y) = E[YF_1(Y)] + E[XF_2(X)] - E(X) \quad (7)$$

Measures d_0 and d_1 should be dimensionless and normalized in the unit interval. The economic distance d_0 fulfills the first but not the second property while d_1 fulfills none of them. d_0 is dimensionless but has the lowest value 1/2, d_1 has the dimension of income and its maximum value is the Gini mean difference. (That means that the maximum of d_1 is the unconditional expectation of the absolute values of income differences).

D_0 and D_1 are the economic distance ratios which satisfy the above mentioned properties:

$$D_0 = 2d_0 - 1 \quad (8)$$

$$D_1 = [E(Y) - E(X)]/[2d_1 - E(Y) + E(X)] \quad (9)$$

Measures D_0 i D_1 are dimensionless and take values in the unit interval. The value "0" is taken when income variables X and Y are independent and identically distributed. This implies that there is no economic distance between the two populations. The value "1" is taken when the two populations do not overlap.

4. THE APPLICATION OF THE GINI COEFFICIENT AND ECONOMIC DISTANCE RATIOS TO THE ANALYSIS OF INCOME INEQUALITY IN POLAND

The Gini coefficient and the economic distance ratios D_0 and D_1 can be calculated from the observed income distributions or derived from an income distribution model. So we can obtain not only the empirical values of these statistics but the theoretical ones as well. The theoretical values of G, D_0 and D_1 are called parametric because they depend on the parameters of an income distribution model.

We have calculated the above mentioned statictics for the distributions of men and women wages in the branches of the Polish economy. The first problem was to find the theoretical distribution presenting a satisfactory goodness of fit with the theoretical ones. We decided to approximate the empirical distributions using two kinds of theoretical models: the lognormal distribution and the Dagum distribution. The lognormal distribution has presented a rather high consistency with empirical wages in Poland for many years. Its density function has the form:

$$f(y) = \frac{1}{y\sigma\sqrt{2\Pi}} \exp\{-\frac{1}{2\sigma^2}(\ln y - \mu)^2\}, \quad y > 0 \tag{10}$$

where:

μ – the expected value of the logarithms of a random variable Y,

σ^2 – the variance of the logarithms of a random variable Y.

The Dagum distribution was introduced in the year 1977 (see Dagum 1977). The researches conducted some years ago (see Jędrzejczak 1993) showed that the Dagum model was even better fitted to the wages distributions in Poland than the lognormal. The cumulative distribution function of the Dagum distribution is the following:

$$F(y) = (1 + \lambda y^{-\delta})^{-\beta} \quad y > 0; \quad \lambda, \beta, \delta > 0 \tag{11}$$

where: λ, β, δ – parameters of the distribution.

To compare the consistency of the Dagum distribution and the lognormal distribution with the empirical distributions of wages in Poland in 1992 we estimated the parameters of both distributions using the maximum likelihood method. Application of this method in estimation of the lognormal distribution parameters has been presented by Aitchison and Brown (1957). In the case of the Dagum model the logarithm of the likelihood function for grouped data has been used:

$$\ln L = \sum_{i=1}^{k} \ln \gamma_i [F(y_i) - F(y_{i-1})], \tag{12}$$

where:

γ_i – empirical frequencies,

$F(y_i)$ – the Dagum distribution function.

Tables 1 and 2 present the results of the calculations. Table 1 shows the consistency measures calculated for the lognormal distribution (the standard deviation of relative frequencies and the coefficient of distribution similarity (Velrose 1960 has been used). Table 2 presents the same measures of consistency for the Dagum model.

The consistency is high only in the case of the Dagum distribution; in the case of the lognormal one the level of consistency is rather low. So we could say that the lognormal distribution can no longer be used for the approximation of wages in Poland. We have calculated G, D_0 and D_1

Table 1

Measures characterizing consistency of empirical wage distributions with lognormal distribution

Branches of national economy	Total		Men		Women	
	C_s	$S(x)$	C_s	$S(x)$	C_s	$S(x)$
Total	0.9288	0.0114	0.9278	0.0126	0.9542	0.0077
Industry	0.9188	0.0130	0.9208	0.0133	0.9635	0.0063
Building	0.9272	0.0126	0.9269	0.0131	0.9233	0.0122
Agriculture	0.9390	0.0106	0.9377	0.0116	0.9259	0.0153
Forestry	0.9638	0.0063	0.9604	0.0065	0.9317	0.0154
Transportation	0.9394	0.0111	0.9376	0.0112	0.9461	0.0101
Telecommunication	0.9022	0.0164	0.8760	0.0203	0.9134	0.0152
Trade	0.9057	0.0164	0.8955	0.0173	0.9198	0.0149
Communal management	0.9636	0.0055	0.9642	0.0057	0.9462	0.0085
Flat management	0.9260	0.0132	0.8868	0.0193	0.9684	0.0059
Science and technics	0.9123	0.0152	0.8592	0.0240	0.9553	0.0072
Education	0.9368	0.0135	0.9395	0.0093	0.9347	0.0160
Culture	0.9519	0.0081	0.9410	0.0109	0.9586	0.0070
Health care	0.9040	0.0153	0.8520	0.0250	0.9320	0.0114
Physical culture	0.8022	0.0322	0.8067	0.0336	0.7906	0.0321
State administration	0.8834	0.0210	0.8545	0.0275	0.8981	0.0181
Administration of justice	0.7512	0.0374	0.7243	0.0538	0.7620	0.0343
Finance and insurance	0.8804	0.0233	0.8269	0.0389	0.8916	0.0202

Note: C_s – coefficient of distribution similarity (see: Velrose 1960),
$S(x)$ – standard deviation of relative frequencies (see: Kordos 1973).

Table 2

Measures characterizing consistency of empirical wage distributions with Dagum distribution

Branches of national economy	Total		Men		Women	
	C_s	$S(x)$	C_s	$S(x)$	C_s	$S(x)$
Total	0.9860	0.0023	0.9842	0.0024	0.9769	0.0044
Industry	0.9746	0.0037	0.9615	0.0056	0.9847	0.0028
Building	0.9813	0.0033	0.9809	0.0032	0.9748	0.0043
Agriculture	0.9873	0.0020	0.9862	0.0023	0.9690	0.0068
Forestry	0.9705	0.0048	0.9592	0.0069	0.9508	0.0107
Transportation	0.9833	0.0027	0.9713	0.0048	0.9807	0.0035
Telecommunication	0.9515	0.0077	0.9579	0.0074	0.9704	0.0050
Trade	0.9689	0.0056	0.9700	0.0055	0.9727	0.0051
Communal management	0.9659	0.0069	0.9643	0.0072	0.9709	0.0054
Flat management	0.9627	0.0067	0.9439	0.0091	0.9679	0.0067
Science and technics	0.9766	0.0044	0.9637	0.0061	0.9579	0.0076
Education	0.9077	0.0158	0.9484	0.0090	0.8962	0.0174
Culture	0.9723	0.0043	0.9740	0.0051	0.9685	0.0053
Health care	0.9735	0.0044	0.9689	0.0057	0.9517	0.0086
Physical culture	0.9193	0.0128	0.9342	0.0102	0.9234	0.0125
State administration	0.9409	0.0101	0.9407	0.0120	0.9417	0.0104
Administration of justice	0.8821	0.0171	0.8658	0.0198	0.9003	0.0151
Finance and insurance	0.9713	0.0044	0.9785	0.0035	0.9766	0.0041

Note: C_s – coefficient of distribution similarity (see: Velrose 1960),
$S(x)$ – standard deviation of relative frequencies (see: Kordos 1973).

for the empirical distributions and for the approximated Dagum distributions, presenting satisfactory goodness of fit. The parametric forms of G, d_0 and d_1 are (see Dagum 1977, Dagum, Lemmi 1989):

$$G = B(\beta, \beta)/B(\beta, \beta + 1/\delta) \tag{13}$$

where:

$B(\beta, \beta)$ – the beta function with the parameters β, β,

β, δ – the Dagum distribution parameters (see Dareto 1897)

$$d_0 = \beta_2 \int_0^1 (1-t)^{\tau}[(1-t)^{\delta} + at^{\delta}]^{-\beta_1} dt \tag{14}$$

where:

$\tau = \beta_1\delta + \beta_2 - 1,$

$\delta = \delta_1/\delta_2,\ a = \lambda_1\lambda_2^{-\delta}$

λ_1, β_1, δ_1 – the Dagum distribution parameters

$$d_1 = \beta_2\lambda_2^{1/\delta_2}\int_0^1 t^{-1/\delta_2}(1-t)^{b}[(1-t)^{\delta} + at^{\delta}]^{-\beta_1}dt + \\ + \beta_1\lambda_1^{1/\delta_1}\int_0^1 t^{-1/\delta_1}(1-t)^{b}[(1-t)^{\delta} + at^{\delta}]^{-\beta_2}dt - E_1(X) \tag{15}$$

where: $b = \beta_1\delta + \beta_2 + 1/\delta_2 - 1$; $\delta = \delta_1/\delta_2$; $a = \lambda_1\lambda_2^{-\delta}$, (in the second component of the equation we permute the subscripts 1 and 2).

λ, β, δ – the Dagum distribution parameters.

Tables 3 and 4 contain the results of the calculations for the observed and estimated distributions of wages in Poland in 1992. In Tab. 3 we present expected values, the Gini coefficients and economic distance ratios D_0 and D_1, comparing wages of men and women in each branch of the national economy. Table 4 comprises expected values and Gini coefficients for each branch of national economy (for men and women together), and ratios D_0 and D_1 measuring economic distance between each of these branches and the branch "Total".

Analysing the obtained results of the calculations we can easily notice some regularities. The theoretical values of G are almost always greater than the corresponding empirical ones. It is connected with the fact that using grouped data we ignore concentration within income groups. The economic distance D_1 is very sensitive to the lower level of consistency between empirical and theoretical distribution. On the other hand, the economic distance ratio D_0 does not react even when the observed and estimated distributions differ considerably.

Table 3

Observed and Estimated Values of Means, Gini Indices and Economic Distance Ratios

Branches of national economy	$E(X)$		G		D_0	D_1
	Men	Women	Men	Women		
Total	3.3676 3.4666	2.7308 2.7489	0.2335 0.2619	0.2083 0.2225	0.2726 0.2672	0.4407 0.4526
Industry	3.5553 3.6470	2.5303 2.5516	0.2335 0.2670	0.1899 0.2059	0.4377 0.4143	0.6608 0.6493
Building	3.2583 3.3234	2.7915 2.7616	0.2331 0.2530	0.2385 0.2515	0.2273 0.2414	0.3155 0.3556
Agriculture	2.5444 2.5373	2.3529 2.4141	0.2141 0.2299	0.2019 0.2458	0.1058 0.0930	0.1861 0.1052
Forestry	2.6545 2.7592	2.3807 2.3478	0.1802 0.1997	0.1714 0.1797	0.1851 0.2197	0.3004 0.4094
Transportation	3.0026 2.9478	2.6393 2.6021	0.1868 0.1859	0.1709 0.1676	0.2137 0.2033	0.3458 0.3414
Telecommunication	3.4238 3.3918	3.5509 3.5646	0.1547 0.1420	0.1505 0.1456	0.1082 0.1174	0.1183 0.1719
Trade	2.2916 2.9856	2.4644 2.5233	0.2556 0.2764	0.2208 0.2484	0.2155 0.1974	0.3670 0.3154
Communal management	3.6147 2.7321	2.9648 2.9920	0.1867 0.2095	0.1914 0.2000	0.3320 0.3306	0.4792 0.4982
Flat management	3.2279 3.1514	2.7617 2.7778	0.1883 0.1725	0.1962 0.2092	0.2637 0.2587	0.3854 0.3188
Science and technics	3.7270 3.9183	3.2296 3.4132	0.2239 0.2559	0.2069 0.2393	0.1768 0.1612	0.3213 0.2793
Education	3.2352 3.2731	2.6943 2.8262	0.1952 0.2094	0.1804 0.2207	0.2767 0.2530	0.4541 0.3282
Culture	3.1500 3.1747	2.8291 2.8795	0.2290 0.2461	0.2103 0.2333	0.1316 0.1254	0.2399 0.2025
Health care	3.4727 3.6625	2.7931 2.9314	0.2485 0.2888	0.1882 0.2124	0.2429 0.2036	0.4529 0.4341
Physical culture	3.6387 4.3073	3.1139 3.2324	0.2846 0.3891	0.2797 0.3127	0.1783 0.1958	0.2678 0.4242
State administration	4.0297 4.7024	3.3248 3.4249	0.2533 0.3516	0.2030 0.2110	0.2124 0.1914	0.3938 0.5396
Administration of justice	4.7265 6.3202	3.3183 3.3558	0.2360 0.4203	0.2762 0.3052	0.4377 0.4661	0.5867 0.7580
Finance and insurance	4.3118 5.6129	3.7099 4.0154	0.2529 0.4126	0.2440 0.3056	0.1716 0.1835	0.2912 0.4782

The first row for each branch contains nonparametric values the second corresponding parametric ones.

Table 4

Observed and Estimated Values of Means, Gini Indices and Economic Distance Ratios

Branches of national economy	$E(X)$		G		D_0	D_1
	Nonpar.	Param.	Nonpar.	Param.		
Finance and insurance	3.8141	4.2859	0.2486	0.3191	0.2548 0.4128	0.2635 0.5333
Administration of justice	3.7096	4.1524	0.2486	0.3191	0.1404 0.1470	0.3391 0.4667
State administration	3.5555	3.6924	0.2284	0.2506	0.2027 0.2277	0.2986 0.3365
Telecommunication	3.5071	3.6136	0.1527	0.1543	0.3031 0.3593	0.3110 0.3480
Science and technics	3.4736	3.4966	0.2189	0.2299	0.1871 0.1935	0.2567 0.2420
Communal management	3.4552	3.5992	0.1932	0.2200	0.2245 0.2509	0.2259 0.3031
Physical culture	3.3415	3.5957	0.2859	0.3419	0.0046 0.0236	0.1508 0.2790
Industry	3.2040	3.2586	0.2384	0.2640	0.0425 0.0365	0.0798 0.0920
Building	3.1792	3.2479	0.2368	0.2587	0.0353 0.0636	0.0442 0.0864
Total	3.0859	3.1101	0.2311	0.2478	0.0000 0.0000	0.0000 0.0000
Flat management	3.0483	3.0301	0.1987	0.1975	0.0209 0.0362	0.0282 0.0586
Culture	2.9594	3.0009	0.2204	0.2408	0.0425 0.0383	0.0924 0.0738
Transportation	2.9138	2.9226	0.1850	0.1892	0.0278 0.0181	0.1360 0.1418
Health care	2.9116	2.8898	0.2061	0.2067	0.0609 0.0598	0.1317 0.1615
Education	2.8199	2.9374	0.1891	0.2239	0.0702 0.0450	0.2107 0.1216
Trade	2.6311	2.5676	0.2387	0.2443	0.2497 0.2623	0.3244 0.3746
Forestry	2.6012	2.6663	0.1800	0.1956	0.1955 0.1630	0.3925 0.3369
Agriculture	2.4932	2.4964	0.2116	0.2327	0.2901 0.2886	0.4472 0.4327

Generally, the inequality of earnings is higher for men. It is connected with greater differentiation of qualifications and occupations within this group. The economic situation of men is better than the situation of women regardless of the branch. The biggest economic distance between men and women is in industry and administration of justice. In industry men's wages exceed women's wages by 65% ($D_1 = 0.6493$) Agriculture and telecommunication are the branches in which the economic situation of men and women is similar – in agriculture $D_0 = 0.0930$, $D_1 = 0.1052$. Analyzing the results of the calculations placed in table 4 we can evaluate the economic distance of each branch in relation to the branch "Total", representing the average level of earnings. The branches lying above it have better economic situation while in the branches placed below the situation is worse, being the worst in agriculture where $D_1 = 0.4327$. This means that the earnings in Poland taken as the whole are higher by 43% than the wages in agriculture.

REFERENCES

Aitchison I., Brown I. A. C. (1957): *The Lognormal Distribution*, Cambridge.

Dagum C. (1977): *A New Model of Personal Income Distribution. Specification and Estimation*, „Economie Apliqueé", XXX(3), p. 413–436.

Dagum C. (1980): *Inequality Measure Between Income Distribution with Application*, „Econometrica", 48, p. 1790–1803.

Dagum C., Lemmi A. (1989): *A Contribution to the Analysis of Income Distribution and Income Inequality and a Case Study: Italy*, [in:] D. I. Slottjee, *Advances in Econometrics*, Yai Press, Greenwitch.

Gastwirth J. L. (1972): *The Estimation of the Lorenz Curve and The Gini Index*, „Reviev of Economics and Statistics", 54, p. 300–316.

Gini C. (1912): *Variabilitá e Mutabilitá*, „Studi Economico-giuridici", Universitá di Cagliari, III, 2a.

Jędrzejczak A. (1993): *Application of the Dagum Distribution in the Analysis of Income Distributions in Poland*, „Acta Universitatis Lodziensis", Folia oeconomica, 131.

Kordos J. (1973): *Metody analizy i prognozowania rozkładów płac i dochodów ludności*, Warszawa.

Morgan (1962): *The Anatomy of Income Distribution*, „Review of Economics and Statistics", 44, p. 270–282.

Panek J. (1991): *Zmiany sytuacji dochodowej grup społeczno-ekonomicznych gospodarstw domowych w latach osiemdziesiątych*, „Wiadomości Statystyczne", nr 1, p. 17–21.

Pareto V. (1897): *Cours d'economie politique*, Lozanna.

Pyatt G. (1976): *On the Interpretation and Disaggregation of Gini Coefficient*, „The Economic Journal", 86, p. 243–255.

Sen A. (1973): *On Economic Inequality*, Oxford, Clarendon Press.

Sen A. (1976): *Poverty: an Ordinal Approach to Measurement*, „Econometrica", 44, p. 219–231.

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

Alina Jędrzejczak

ZASTOSOWANIE WSPÓŁCZYNNIKÓW ODLEGŁOŚCI EKONOMICZNEJ DAGUMA DO BADANIA ROZKŁADÓW PŁAC W POLSCE

Studia w dziedzinie rozkładów dochodów koncentrują się nad metodami mierzenia nierównomierności tych rozkładów. Możemy badać zarówno stopień nierównomierności dochodów wewnątrz każdej z populacji jak i porównywać sytuację dochodową dwóch populacji korzystając ze współczynników nierównomierności pomiędzy rozkładami. Współczynniki odległości ekonomicznej zaproponowane przez Daguma pozwalają na ocenę stopnia dominacji ekonomicznej jednej populacji nad drugą, z punktu widzenia ich sytuacji dochodowej. W artykule prezentujemy nieznormalizowane odległości ekonomiczne d_0 i d_1 oraz odpowiednie współczynniki odległości d_0 i d_1 ekonomicznej D_0 i D_1. Współczynnik odległości ekonomicznej D_1 bierze pod uwagę różnice dochodowe pomiędzy wszystkimi jednostkami obu badanych populacji, podczas gdy D_0 mierzy jedynie częstość z jaką dochody populacji o większej średniej są większe od dochodów populacji o mniejszej średniej. Współczynniki te obliczyliśmy dla empirycznych rozkładów płac w Polsce (postać nieparametryczna) oraz dla rozkładów teoretycznych, których stopień zgodności z empirycznymi był wysoki (postać parametryczna).