ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 131, 1993

Alina Jędrzejczak[*]

APPLICATION OF THE DAGUM DISTRIBUTION IN THE ANALYSIS OF INCOME DISTRIBUTIONS IN POLAND

The paper presents an attempt of approximation of wages distributions in Poland by means of the Dagum distribution. We compare the results of the calculations for Dagum distribution with the results we obtained by approximation of the distributions using the lognormal curve.

First we should define the main economic variables below considered as wages, income and wealth. And so wage is payment made or received for work, while income is all money received during a given period of time (as wage, receipts from trade, interests from investment etc.). Personal income is divided between consumption and saving. Accumulated past saving is called wealth.

The problem of specifying the theoretical distribution describing empirical wages and income distributions has been the aim of scientific researches for many years. There have been made attempts of application such curves like the Pareto curve, the Pearson curves, the normal and lognormal curve and the beta or gamma functions to solve this problem. Most of the above mentioned curves were chosen only because they are asymmetric probability density functions.

So far the approximation of wages and income distributions has been most often applied by means of the lognormal distribution. Its probability density function has the form:

---

[*] Assistant in the Institute of Econometrics and Statistics, University of Łódź.

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2\sigma^2}(\ln x - \mu)^2\right], & x > 0 \\ 0 & x \le 0 \end{cases} \quad (1)$$

where:

μ - the arithmetic mean of the logarithms of a random variable X,

σ^2 - the variance of the logarithms of a random variable X.

The lognormal distribution has a rather high consistency with empirical wages and income distributions, proved by researches conducted in Poland and abroad for many years (f.i.: Aitchison, Brown 1957; Èltètö 1962; Walter 1962; Vielrose 1960; Kordos 1968, 1976; Kordos, Stroińska 1971, Domański 1975). The lognormal distribution depends only on two parameters: μ and σ^2, which, on the one hand simplifies the calculations, and on the other hand, decreases the opportunity of perfect fitting the theoretical curve to the empirical points. Moreover the lognormal distribution does not apply, generally, to wealth distribution analysis. The reason is that its probability density function is unimodal while the observed wealth distributions are zeromodal[1] - we do not usually observe any concentration of frequency in the central intervals.

The characteristics of the lognormal distribution given above have induced many scholars to further search for an explanatory model of income distributions. The empirical investigations performed by Dagum (1977) have shown, that the income elasticity of the cumulative distribution function of income can be specified in the form of the differential equation:

$$\varepsilon(y, F) = \frac{d\ln F(y)}{d\ln y} = \beta_1[1 - [F(y)]^{\beta_2}], \quad \text{for } y \ge 0;\ \beta_1,\ \beta_2 > 0 \quad (2)$$

From this equation we deduce that for the whole income range, the income elasticity ($\varepsilon(y, F)$) of the cumulative distribution function ($F(y)$) is a decreasing and bounded function of F. Solving

[1] Income distributions can be zeromodal, too. In particular it concerns developing countries with a population polarized into a small number of the very wealthy and a large number of the very poor.

this equation, Dagum got the cumulative distribution function of income. Its mathematical form is (see: Figure 1):

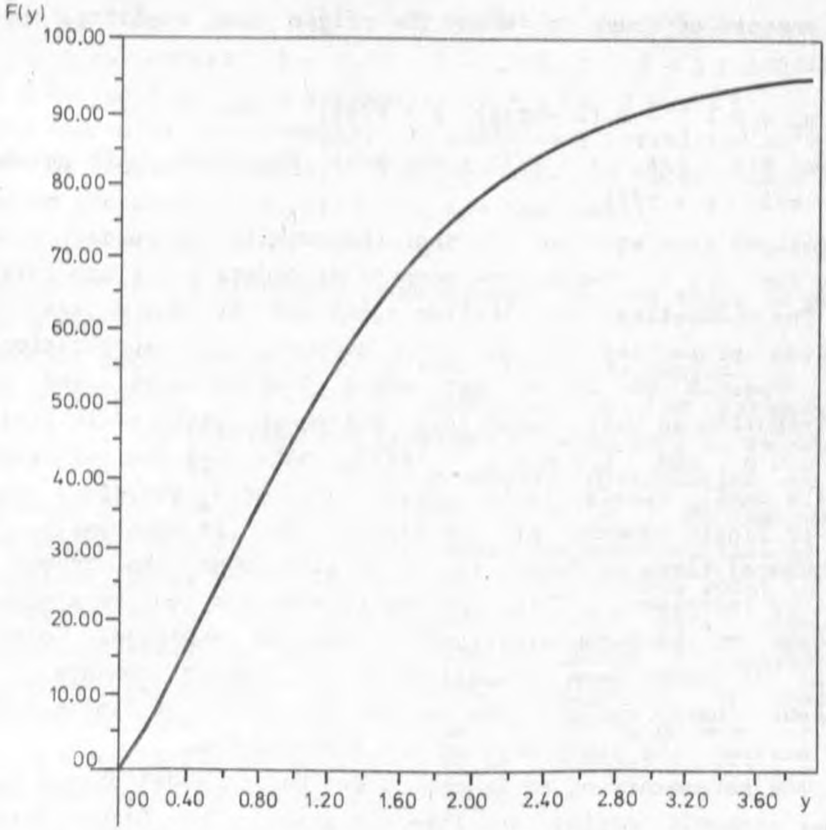

Fig. 1. Dagum distribution function

$$F(y) = \begin{cases} (1 + \lambda y^{-\delta})^{-\beta}, & y > 0 \\ 0 & y \leq 0 \end{cases} \qquad \text{for } \beta, \lambda, \delta > 0 \tag{3}$$

where:

$\beta = 1/\beta_2$,

$\delta = \beta_1 \beta_2$,

$\lambda = \exp.c$,

c - the constant of integration arising from the solution of equation (2).

The probability density function corresponding to the cumulative distribution function given in equation (3) is:

$$f(y) = \begin{cases} \beta \lambda \delta y^{-\delta-1}(1 + \lambda y^{-\delta})^{-\beta-1}, & y > 0 \\ 0 & y \leq 0 \end{cases} \tag{4}$$

The moments of order r about the origin are specified by the equation:

$$\mu_r = \beta \lambda^{r/\delta} B(1 - r/\delta, \ \beta + r/\delta), \quad \text{for} \ r < \delta \tag{5}$$

where: $B(1 - r/\delta, \ \beta + r/\delta)$ - the beta functions with parameters $(1 - r/\delta, \ \beta + r/\delta)$.

It follows from equation (5) that the moments of order r exist only for $r < \delta$. Hence, the moments of orders $r > \delta$ are infinite.

The theoretical distribution specified by Dagum has interesting properties taking into account its application in the research of income and wages distributions, and wealth distribution as well. Empirical evidences gathered in Italy by Dagum and Lemmi (1987), show that the estimates of δ are small; usually in the neighborhood of 4. Therefore the number of finite moments of the distribution is also small, being in general three or four. (It is smaller when the income inequality increases). This fact can be admitted to be a great advantage of the Dagum distribution, because empirical distributions of income have a small number of finite moments - they present "heavy tails". The smaller the number of finite moments, the heavier the right tail of the distribution.

The parameters of an income distribution model should have a clear economic meaning and interpretation. The Dagum distribution is characterized by three parameters: β, λ and δ, where β and δ are parameters of inequality, and λ is a parameter of scale, being a function of the monetary unit of measurement. Such an interpretation of λ has a great practical meaning - we have the possibility to compare income distributions of different countries or the distributions of the same country in different periods of time, replacing in equation (3) the estimated value of λ by λ^*:

$$\lambda^* = \lambda k^{-\delta} \tag{6}$$

where k can be for example the exchange rate of a "new" currency.

Parameters β and δ are inequality parameters. The Lorenz curve and the Lorenz coefficient of concentration are functions of β and δ. The Lorenz curve corresponding to the cumulative distribution function (3) can be written in the form:

$$l(F) = \frac{B^*[F^{1/\beta};\ \beta + 1/\delta,\ 1 - 1/\delta]}{B[\beta + 1/\delta,\ 1 - 1/\delta]}, \quad \text{for} \quad \delta > 1, \quad 0 \le F \le 1 \tag{7}$$

where: $B^*[F^{1/\beta};\ \beta + 1/\delta,\ 1 - 1/\delta]$ - the uncomplete beta function with parameters $\beta + 1/\delta,\ 1 - 1/\delta)$, $B[\beta + 1/\delta,\ 1 - 1/\delta]$ - the beta function with parameters $(\beta + 1/\delta,\ 1 - 1/\delta)$.

The degree of its convexity is positively correlated with the degree of income inequality. The concentration coefficient obtained on the basis of equation (7) has the form;

$$K = -1 + B(\beta, \beta) / B(\beta, \beta + 1/\delta) \tag{8}$$

The value of K is inversely proportional to the value of parameters β and δ.

Another important feature of the Dagum distribution is that it can be unimodal or zeromodal, depending on the parameters. Specifically, the distribution is unimodal, when $\beta\delta > 1$, zeromodal when $0 < \beta\delta < 1$. Hence, using the presented distribution we can approximate both the income distributions ($\beta\delta > 1$) and the wealth distributions ($0 < \beta\delta < 1$), which varies the possibilities of its application. It is worth noting, that the starting point in the construction of the above presented distribution were economic and empirical premises. Many other models applied in analysis of this problem, for example the lognormal distribution or Pearson curves, do not have such foundations. So the Dagum model of income possesses a set of basic properties important to analyze and to interpret observed income distributions.

In the following part of the paper we compare the consistency of the Dagum distribution and the lognormal distribution with empirical distributions of wages in Poland in the year 1988. To do this we estimated the parameters of the Dagum function by means of the maximum likelihood method. The lognormal distribution parameters have been estimated by means of two methods: the maximum likelihood method and the method of quantiles. The general form of the logarithm of likelihood function is following:

$$\ln L = \sum_{i=1}^{k} \ln \gamma_i\ [F(y_i) - F(y_{i-1})] \tag{9}$$

where:

y_i, y_{i-1} - limits of income intervals,

γ_i - frequency.

Given y_i and γ_i, the likelihood function for the Dagum distribution is the function of three variables: λ, β and δ; for the lognormal distribution function of two variables: μ and σ. To get its maximum the individual numerical procedure programmed on the IBM PC has been applied in both cases. (The methods of getting maximum of likelihood functions are clearly described by M i l o 1990).

T a b l e 1

Estimated values of Dagum and lognormal distribution parameters in the branches of national economy

Branches of national economy	Lognormal distribution				Dagum distribution		
	method of quantiles		method of maximum likelihood				
	μ	σ	μ	σ	λ	β	δ
Total	1.60	0.39	1.60	0.37	786.4	1.26	4.38
Material production	1.65	0.38	1.66	0.37	960.0	1.32	4.39
Industry	1.72	0.39	1.73	0.37	922.6	1.44	4.27
Building	1.67	0.38	1.68	0.38	1 239.9	1.22	4.43
Agriculture	1.58	0.36	1.57	0.35	13 528.2	0.74	5.71
Forestry	1.53	0.38	1.52	0.35	11 433.9	0.73	5.76
Transportation	1.70	0.34	1.70	0.32	11 641.2	0.99	5.50
Telecomunication	1.62	0.31	1.61	0.30	37 248.0	0.78	6.28
Trade	1.46	0.35	1.47	0.32	973.4	1.43	5.03
Communal management	1.58	0.37	1.59	0.36	473.7	1.58	4.31
Beyond the material production	1.43	0.39	1.44	0.35	358.4	1.44	4.47
Flat management	1.53	0.35	1.53	0.34	2 395.0	1.08	5.14
Science and technics	1.75	0.35	1.75	0.33	7 264.8	1.12	5.17
Education	1.38	0.40	1.37	0.36	79.8	2.17	3.97
Culture	1.36	0.38	1.36	0.35	154.3	1.91	4.39
Health and social care	1.48	0.35	1.48	0.32	3 071.8	1.03	5.44
Phisical culture, tourism and recreation	1.50	0.37	1.51	0.34	735.6	1.38	4.70
State administration	1.43	0.44	1.42	0.41	7.3	6.60	3.06
Finance and insurance	1.57	0.33	1.55	0.31	103 304.1	0.64	6.90

S o u r c e: Author's calculations.

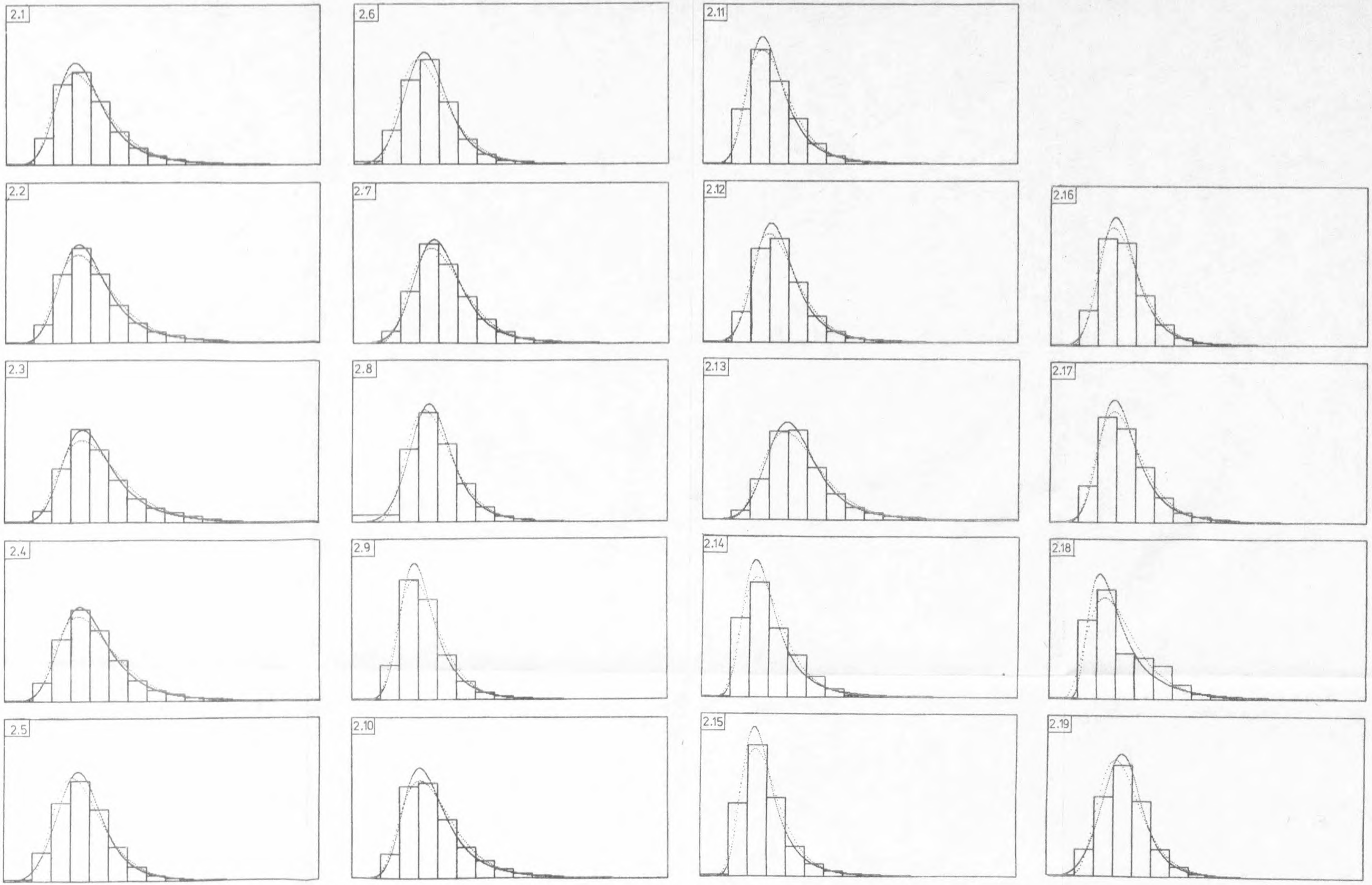

Fig. 2. Empirical wages distributions and their approximation by means of Dagum and lognormal distribution: 2.1 - total, 2.2 - material production, 2.3 - industry, 2.4 - building, 2.5 - agriculture, 2.6 - forestry, 2.7 - transportation, 2.8 - telecommunication, 2.9 - trade, 2.10 - communal management, 2.11 - beyond the material production, 2.12 - flat management, 2.13 - science and technics, 2.14 - education, 2.15 - culture, 2.16 - health and social care, 2.17 - phisical culture, 2.18 - state administration, 2.19 - finance and insurance. Lognormal distribution - dashed line, Dagum distribution - continuous line

T a b l e 2

Measures characterizing consistency of empirical and theoretical distributions in the branches of national economy

Branches of national economy	Lognormal distribution				Dagum distribution	
	method of quantiles		method of max. likelihood			
	C_s	S(x)	C_s	S(x)	C_s	S(x)
Total	0.962	0.0096	0.972	0.0069	0.984	0.0035
Material production	0.953	0.0112	0.957	0.0104	0.991	0.0019
Industry	0.932	0.9162	0.939	0.0143	0.972	0.0055
Building	0.949	0.0113	0.953	0.0104	0.990	0.0020
Agriculture	0.951	0.0107	0.960	0.0108	0.976	0.0057
Forestry	0.919	0.0198	0.936	0.0162	0.985	0.0035
Transportation	0.956	0.0121	0.971	0.0084	0.976	0.0059
Telecommunication	0.969	0.0076	0.975	0.0072	0.974	0.0069
Trade	0.928	0.0187	0.952	0.0121	0.994	0.0015
Communal management	0.954	0.0109	0.954	0.0107	0.974	0.0054
Beyond the material production	0.951	0.0132	0.979	0.0060	0.955	0.0099
Flat management	0.958	0.0118	0.974	0.0068	0.981	0.0042
Science and technics	0.931	0.0175	0.941	0.0137	0.983	0.0035
Education	0.946	0.0140	0.949	0.0152	0.925	0.0190
Culture	0.936	0.0181	0.960	0.0107	0.976	0.0072
Health and social care	0.947	0.0148	0.981	0.0049	0.978	0.0047
Phisical culture tourism and recreation	0.957	0.0105	0.981	0.0047	0.964	0.0085
State administration	0.878	0.0311	0.883	0.0316	0.896	0.0249
Finance and insurance	0.931	0.0165	0.942	0.0158	0.973	0.0067

N o t e: C_s - coefficient of distribution similarity (see: K o r d o s, 1976, p. 74); S(x) - standard deviation of relative frequencies (see: K o r d o s, 1976, p. 116).

S o u r c e: As Table 1.

The obtained results of calculations are presented in tables and on figures. Table 1 presents estimates of Dagum and lognormal distribution parameters, while in Table 2 there are consistency measures calculated for both distributions - the standard deviation of relative frequencies and the coefficient of distribution similarity. The coefficient of distribution similarity was constructed 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 income groups:

$$W_p = \sum_{i=1}^{k} \min.(\gamma_i;\ \gamma_i'), \qquad 0 \leq W_p \leq 1 \tag{10}$$

where:

γ_i - empirical frequency,

γ_i' - theoretical frequency.

The bigger the value of W_p, the higher the consistency of compared distributions. Analyzing both measures one can easily notice that the wages distributions estimated by means of the Dagum function show generally greater consistency with the empirical distributions, than the same ones estimated using the logarithmic - normal curve. The consistency is much greater in the case of Dagum distribution, when the lognormal distribution parameters have been estimated by means of the method of quantiles (only in the case of "Education" we have observed a lower level of consistency). The Dagum distribution, when estimating the parameters of lognormal distribution by means of maximum likelihood method, proved lower level of consistency only in three branches: "The sphere beyond the material production", "Education" and "Physical education, tourism and recreation". It is worth mentioning that the degree of consistency of empirical distributions with the lognormal distribution is satisfactory only for ten out of nineteen examined wages distributions, and high in only three of them (the standard deviation of differences <0,006). In other cases, deviations between empirical and theoretical distributions can be considered as significant (the standard deviation of differences >0,01). On the other hand, for the Dagum distribution there is high consistency with the empirical wages distribution in Poland and only in two branches - "Education" and "State administration" the standard deviation of differences >0,01.

Summing up, one can say, that the Dagum distribution is better for approximation of the analyzed wages distributions than the lognormal model. The short scope of this research should be however a reason for a careful generalization of our conclusions. Nevertheless, the presented results are a strong argument for carrying on and depening the researches concerning the Dagum distribution in the approximation of wages and income distributions in Poland.

REFERENCES

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

Dagum C. (1977), *A New Model of Personal Income Distribution: Specification and Estimation,* "Economie Appliquèe", No. XXX(3).

Dagum C., Lemmi A., (1987), *A Contribution to the Analysis of Income Distributions and Income Inequality, and a Case Study: Italy,* Econometric Society Meeting, Chicago.

Domański Cz. (1975), *Rozkład liczby zatrudnionych pracowników fizycznych i umysłowych w gałęziach gospodarki narodowej w 1967 r.* [The distribution of manual and office workers in the branches of national economy in the year 1967], "Zeszyty Naukowe Uniwersytetu Łódzkiego", Ser. III nr 1.

Èltètö O. (1962), *Analysis of the Types and Characteristics of the Income Distribution,* [in:] *The Standard of Living,* Budapest.

Kordos J. (1968), *Matematyczne metody analizy rozkładów dochodów osobistych ludności* [Mathematical methods of studying and analysis of personal income distributions], Warszawa.

Kordos J. (1976), *Metody analizy i prognozowanie płac oraz rozkładów dochodów osobistych ludności* [Methods of analysis and forecasting of wages and personal income distributions], Warszawa.

Kordos J., Stroińska Z. (1971), *Statystyczne metody analizy płac i dochodów ludności* [Statistical methods of analysis of wages and personal income distributions], "Z Prac Zakładu Badań Statystyczno-Ekonomicznych", nr 62.

Milo W. (1990), *Nieliniowe modele ekonometryczne* [Non-linear econometric models], PWN, Warszawa.

Rocznik statystyczny 1989 [Statistical year-book 1989], Warszawa.

Vielrose E. (1960), *Rozkłady dochodów* [The size of distribution of income], Warszawa.

Walter J. (1962), *Lognormal Distribution and Its Application in the Analysis of Households Stratified by Income,* [in:] *The Standard of Living,* Budapest.

Alina Jędrzejczak

ZASTOSOWANIE ROZKŁADU DAGUMA DO BADANIA ROZKŁADÓW PŁAC I DOCHODÓW W POLSCE

W artykule przedstawiliśmy próbę aproksymacji rozkładów płac w Polsce za pomocą rozkładu Daguma. Rozkład ten, zaproponowany w 1977 r., nie był dotąd wykorzystywany do badania płac i dochodów w naszym kraju. Praca zawiera prezentację modelu oraz wyniki, które otrzymaliśmy aproksymując rozkłady płac wg działów gospodarki narodowej w 1988 r. Dla porównania oszacowaliśmy także parametry rozkładu logarytmiczno-normalnego, który był dotąd najczęściej stosowany do badania płac i dochodów ludności w Polsce. Obliczone miary zgodności rozkładów empirycznych z teoretycznymi wskazują jednoznacznie, że rozkład Daguma lepiej aproksymuje badane rozkłady płac niż rozkład logarytmiczno-normalny. Możliwość zastosowania rozkładu Daguma do badania rozkładów zamożności, a także przejrzysta interpretacja ekonomiczna jego parametrów są dodatkowym argumentem skłaniającym nas do dalszych badań przydatności tego rozkładu do aproksymacji rozkładów płac i dochodów w Polsce.