QUANTITATIVE METHODS IN ECONOMICS
Vol. XV, No. 2, 2014, pp. 167-176

THE USE OF CSISZÁR'S DIVERGENCE TO ASSESS DISSIMILARITIES OF INCOME DISTRIBUTIONS OF EU COUNTRIES

Jarosław Oczki[1], Ewa Wędrowska[2]
Nicolaus Copernicus University in Toruń
[1] Department of Human Resource Management
e-mail: joczki@umk.pl
[2] Department of Econometrics and Statistics
e-mail: ewaw@umk.pl

Abstract: Income distributions can be described by measures of central tendency, dispersion, skewness, kurtosis or by indexes of polarization. In numerous studies, Gini coefficient and Lorenz curve have been used to investigate inequality of incomes. Income distributions can also be analysed in comparison to one another. In the article two measures belonging to Csiszár's divergence class have been used to identify the degree of differentiation of income distributions among the EU countries in 2005 and 2012. Similar and dissimilar countries with respect to distribution of income have been identified and the change of divergence of EU countries income distributions between 2005 and 2012 has been assessed. European Union Statistics on Income and Living Conditions (EU-SILC) dataset has been used.

Keywords: income inequalities, income distribution, Csiszár's divergence

INTRODUCTION

Income levels and income distributions draw attention of researchers, especially those analysing labour markets, social policy and poverty. Many of studies concern the processes of income convergence or divergence, while other explore the properties of income distributions, including income inequalities, e.g. (Jędrzejczak 2012) and (Quintano et al. 2009). In the article, Csiszár's divergence measure have been applied to assess income inequality in EU countries, and to identify the degree of differentiation of income distributions of all EU countries in

2005 and 2012. Similar approach to analyzing dissimilarity of distributions of economic variables has been used in (Tomczyk 2011), (Podolec et al. 2011) and (Wędrowska 2011).

MEASURES OF DIVERGENCE

The importance of measures of distance between probability distributions arises because of the role they play in problems of inference and discrimination [Ullah 1996]. Divergence measures based on the concept of information-theoretic entropy were first introduced in communication theory by Shannon in 1948 and later developed by Wiener in 1949. These types of measures describe the degree of similarity between a pair of probability distributions.

One of the most general probability measures which plays a significant role in information theory is the well known Csiszár's *f*-divergence [Csiszár 1967].

Csiszár's *f*-divergence between a pair of discrete probability distributions: $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ is defined as:

$$C_f(P,Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right), \tag{1}$$

where $f:[0,\infty) \to \Re$ is a convex function satisfying $f(1) = 0$, $f'(1) = 0$, $f''(1) > 0$ and at $x = 0$, $0 \cdot f\left(\frac{0}{0}\right) = 0$ and $0 \cdot f\left(\frac{p}{0}\right) = \lim_{x\to\infty} \frac{f(x)}{x}$ (Menéndez et al. 2003).

A number of information theory measures are merely the particular cases of Csiszár's *f*-divergence. A list of *f*-divergence measures is provided in (Taneja 2004), (Taneja 2008) and (Wędrowska 2012). In the article Jeffreys-Kullback-Leibler divergence (*J*-divergence) and Jensen-Shannon divergence (*JS*-divergence) have been used in order to measure the degree of similarity between a pair or multiple income distributions.

J-divergence is a function of Kullback-Leibler divergence, the *I*-divergence, or the relative entropy, which assesses the dissimilarity between a pair of probability distributions. The *I*-divergence is defined as:

$$I(P,Q) = \sum_{i=1}^{n} p_i log_2\left(\frac{p_i}{q_i}\right), \tag{2}$$

for probability distributions $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$.
It is well known that *I*-divergence is non-negative, additive, but not symmetric. The *I*-divergence is coincident with Csiszár's *f*-divergence for convex function (Wędrowska 2012):

$$f_{KL} = x log_2 x. \tag{3}$$

The sum of the two mentioned divergences is Kullback's symmetric divergence, also known as the *J*-divergence (Cavanaugh 1998). To obtain a symmetric measure, one can define:

$$J(P,Q) = I(P,Q) + I(Q,P) = \sum_{i=1}^{n}(p_i - q_i)log_2\left(\frac{p_i}{q_i}\right). \quad (4)$$

The *J*-divergence coincides with *f*-divergence for convex function (Reid et al. 2009):

$$f_J = (x-1)log_2 x. \quad (5)$$

The properties of *J*-divergence are discussed in (Seghouane et al. 2004), (Lefebvre et al. 2010) and (Taneja 2013).

Lin introduced an information-theory based divergence measure regarding two or more probability distributions (Lin 1991) known as Jensen-Shannon divergence. It is based on the Shannon entropy and is related to the Kullback-Leibler divergence. The *JS*-divergence is defined as:

$$JS(Q,P) = H\left(\frac{P+Q}{2}\right) - \frac{1}{2}(H(P) + H(Q)), \quad (6)$$

where $H(P) = \sum_{i=1}^{n} p_i log_2 p_i$ is the Shannon entropy.

Jensen-Shannon divergence is the difference between the Shannon entropy of the mean density and the mean value of their entropies.

The Jensen-Shannon divergence is a symmetrized and smoothed version of the Kullback-Leibler divergence:

$$JS(P,Q) = \frac{1}{2}\left[I\left(P, \frac{P+Q}{2}\right) + I\left(Q, \frac{P+Q}{2}\right)\right]. \quad (7)$$

It coincides with Csiszár's *f*-divergence for convex function (Taneja 2005):

$$f_{JS}(x) = \frac{x}{2}log_2 x + \frac{x+1}{2}log_2\left(\frac{2}{x+1}\right). \quad (8)$$

Discussion of properties of JS-divergence can be found in (Menéndez et al. 1997), (Lamberti 2008) and (Grosse 2002).

The generalization of JS-divergence is defined as (Lin 1991):

$$JS(Q,P) = H(\pi_1 P + \pi_2 Q) - \pi_1 H(P) - \pi_2 H(Q), \quad (9)$$

where $\pi_1, \pi_2 \geq 0$, $\pi_1, +\pi_2 = 1$ are arbitrary weights for the probability distributions *P* and *Q*. Since *H* is concave function, *JS*(*P*,*Q*) is nonnegative and equal to zero, when $P = Q$. For an arbitrary set of probability distributions $P_1, P_2, \cdots, P_m$ with weights $\pi_1, \pi_2, \cdots, \pi_m \geq 0$, $\sum_{i=1}^{m} \pi_1 = 1$, the Jensen-Shannon divergence is defined by:

$$JS(P_1, P_2, \cdots, P_m) = H\left(\sum_{i=1}^{m} \pi_i P_i\right) - \sum_{i=1}^{m} \pi_i H(P_i). \quad (10)$$

DIVERGENCE OF INCOME DISTRIBUTIONS OF EUROPEAN UNION COUNTRIES

In this part of the article an analysis of divergence of income distributions of the EU countries is carried out. The analysis is preceded by the investigation of income inequalities in the European Union countries in 2005 and 2012. In order to assess countries income inequalities, Gini coefficient – one of the most popular measures of income concentration, as well as Shannon entropy have been presented (table 1). Gini coefficient values have been taken from EU-SILC database (for disposable income after social transfers). Shannon entropy for each country has been calculated for income distributions represented by shares of national disposable income in the relevant decile as percentage of total national disposable income.

The value of disposable income after social transfers is dependent on:

- labour market outcomes, such as: wages of employees or profits of self-employed, which in turn can be a result of labour market institutions (e.g.: minimum wages, flexible employment contracts regulations), dispersion of qualifications, or discrimination, e.g. against immigrants or employees working in flexible employment forms,
- transfers, which are part of countries' tax and social policies.

High level of income inequality can be an effect of increased variation of wages (or profits), and (or) low degree of income redistribution achieved by social transfers, and fiscal policy in general.

Data in Table 1 show that in 2005-2012 period income inequalities decreased in majority of EU countries. Latvia and three Mediterranean countries: Portugal, Spain and Greece can be identified as those with highest income inequalities throughout the whole period. Two least wealthy countries in the EU – Bulgaria and Romania, which joined the block in 2007, both have above-average income inequalities in 2012. High inequalities and low level of average incomes in these countries indicate that there is a threat that substantial groups of their societies, earning incomes in the first several deciles of the income distribution, could be at high risk of poverty. This situation creates a challenge for economic and social policies pursued by the governments of these countries.

Gini coefficients, as well as entropy values also indicate countries with lowest income inequalities throughout the analysed period – two Scandinavian countries: Sweden and Finland, and three Central and Eastern European countries: Slovenia, Slovakia and Czech Republic.

Table 1. Income inequalities in European Union countries in 2005 and 2012

2005			2012		
Country	Entropy	Gini Coefficient	Country	Entropy	Gini Coefficient
Portugal	2.980	38.1	Latvia	3.031	35.7
Lithuania	3.019	36.3	Spain	3.045	35.0

2005			2012		
Country	Entropy	Gini Coefficient	Country	Entropy	Gini Coefficient
Latvia	3.022	36.2	Portugal	3.046	34.5
Poland	3.033	35.6	Greece	3.053	34.3
United Kingdom	3.046	34.6	Bulgaria	3.069	33.6
Estonia	3.057	34.1	United Kingdom	3.074	32.8
Italy	3.071	32.8	Romania	3.078	33.2
Greece	3.074	33.2	Estonia	3.084	32.5
Spain	3.090	32.2	Italy	3.086	31.9
Ireland	3.094	31.9	Lithuania	3.091	32.0
Cyprus	3.136	28.7	Cyprus	3.104	31.0
Belgium	3.145	28.0	Poland	3.106	30.9
Hungary	3.147	27.6	Croatia	3.110	30.5
France	3.149	27.7	France	3.110	30.5
Netherlands	3.157	26.9	Ireland	3.123	29.9
Malta	3.159	27.0	Denmark	3.138	28.1
Germany	3.163	26.1	Germany	3.145	28.3
Luxembourg	3.164	26.5	Luxembourg	3.146	28.0
Austria	3.166	26.2	Austria	3.149	27.6
Czech Republic	3.167	26.0	Hungary	3.158	26.9
Slovakia	3.168	26.2	Malta	3.160	27.1
Finland	3.169	26.0	Belgium	3.163	26.6
Denmark	3.192	23.9	Finland	3.169	25.9
Slovenia	3.192	23.8	Netherlands	3.173	25.4
Sweden	3.199	23.4	Slovakia	3.178	25.3
			Czech Republic	3.181	24.9
			Sweden	3.182	24.8
			Slovenia	3.195	23.7

Source: own calculations based on EU-SILC database

In 2005, Poland, with relatively high value of Ginni coefficient and small Shannon entropy, belonged to the group of the EU countries characterized by the largest income inequalities. After 2005, Gini coefficient in Poland had been falling gradually to reach a level close to EU average in 2012. In the whole period, Poland experienced the largest drop in that index. Also significant decreases in inequalities were observed in Lithuania and Portugal. The downward tendency of the values of Gini coefficient could have been observed in almost all countries with above-average initial levels of income inequalities. On the other hand, the largest increase in inequality between 2005 and 2012 occurred in Denmark, France and Spain.

In the next step of our analysis we identify the degree of divergence between the income decile distributions of EU countries. The Jesnen-Shannon divergence has been calculated for 25 countries in 2005 and for 28 countries in 2012 (27 EU member states in 2012 and Croatia which joined the EU in 2013). Comparison of values of *JS*-divergence (the bottom row of table 2) suggests that in 2005-2012 period the divergence of income distributions of all EU countries decreased, from

JS=0.00521 in 2005 to JS=0.00392 in 2012. The fall in divergence of income distributions between 2005 and 2012 can be attributed to a trend observed in countries with initially high income inequalities towards a more egalitarian distribution of income.

Table 2. Distribution of income deciles and *JS*-divergence for EU countries in 2005 and 2012

2005			2012		
Country	First decile	Tenth decile	Country	First decile	Tenth decile
Portugal	2.5	30.3	Latvia	2.3	27.1
Lithuania	2.2	27.2	Spain	1.5	24.8
Latvia	2.1	27.7	Portugal	2.7	27.3
Poland	2.2	26.9	Greece	1.8	25.1
United Kingdom	2.6	27.1	Bulgaria	2.3	25.4
Estonia	2.4	25.7	United Kingdom	2.7	25.9
Italy	2.5	25.4	Romania	2.1	23.3
Greece	2.5	25.0	Estonia	2.6	24.2
Spain	2.5	23.8	Italy	2.4	24.3
Ireland	3.3	25.2	Lithuania	2.7	23.9
Cyprus	3.5	22.8	Cyprus	3.5	25.1
Belgium	3.8	23.2	Poland	3.1	24.2
Hungary	3.7	23.2	Croatia	2.6	22.8
France	3.8	22.9	France	3.6	25.6
Netherlands	3.2	22.1	Ireland	3.1	23.2
Malta	3.7	21.0	Denmark	2.3	22.2
Germany	3.7	22.1	Germany	3.4	22.4
Luxembourg	3.7	21.6	Luxembourg	3.6	22.2
Austria	3.8	21.9	Austria	3.2	22.1
Czech Republic	4.0	22.2	Hungary	3.7	22.2
Slovakia	3.4	21.5	Malta	3.8	21.8
Finland	4.1	22.1	Belgium	3.5	21.1
Denmark	3.4	19.7	Finland	4.0	21.6
Slovenia	3.9	19.9	Netherlands	3.8	21.3
Sweden	3.9	19.8	Slovakia	3.6	20.3
			Czech Republic	4.1	21.6
			Sweden	3.4	20.0
			Slovenia	3.9	19.6
Jensen-Shannon divergence JS=0.00521			Jensen-Shannon divergence JS=0.00392 (JS*=0.0038901)[1]		

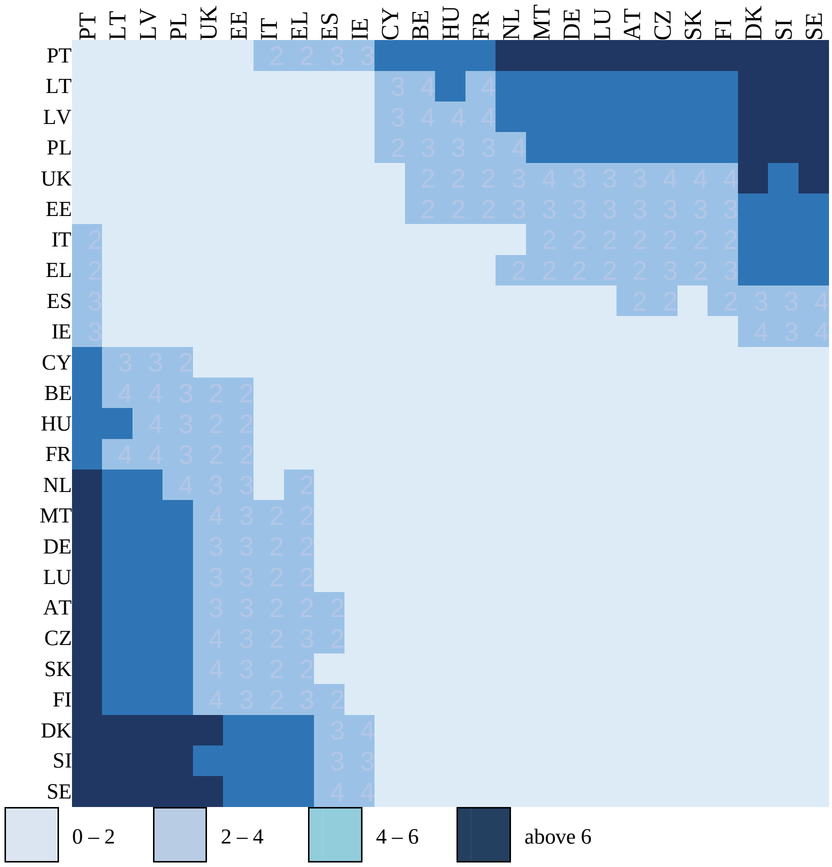
Source: EU-SILC database and own calculations

[1] The value of divergence JS* has been calculated for the same group of countries as JS for 2005, i.e. all EU countries, excluding Bulgaria, Croatia and Romania.

Table 2 also presents the shares of population earning first and tenth decile of income. Analysis of the data leads to a conclusion that the most significant decrease in inequalities measured by shares of population earning top and bottom 10 percent of income was observed in Lithuania and Poland. For example, in Lithuania, the share of population earning bottom 10 percent of income fell from 27.2 percent in 2005 to 23.9 in 2012. In some countries there have been increases of inequalities, especially in Spain where the proportion of population earning top 10 percent of income dropped from 2.5 to 1.5 and Denmark, where the respective proportion fell from 3.4 to 2.3 percent.

In the next step, the degree of dissimilarity between income decile distributions of each pair of countries have been investigated, as measured by Jeffryes-Kullback-Leibler divergence. The results for 2005 and 2012 are presented in Figures 1 and 2, respectively. The darker squares in the figure indicate larger values of divergence between a pair of income distributions of countries representing a particular row and column. The darkest areas are concentrated in top-right and bottom-left corners of the chart simply because income distributions of countries with high income inequalities vary greatly from the distributions of countries with lowest inequalities.

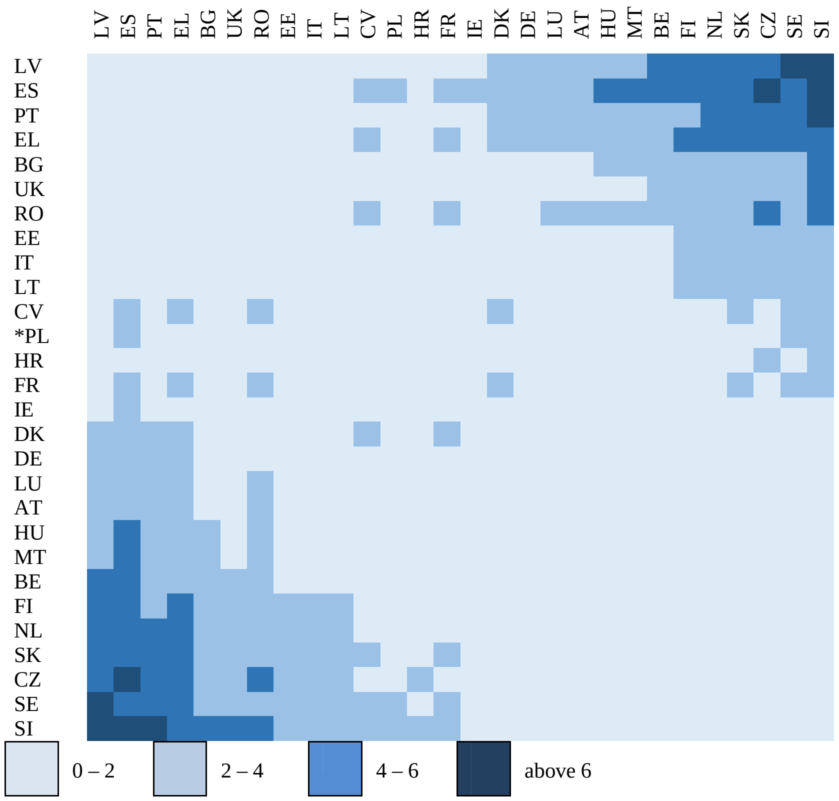
Figure 1. Jeffryes-Kullback-Leibler divergence for pairs of EU countries in 2005



Source: own calculations

The smaller deep-dark areas in top-right and bottom-left corners in Figure 2 in relation to Figure 1, indicate that, in period 2005-2012, EU countries became more similar in their income distributions – in 2012 there were fewer pairs of income distributions for which the value of Jeffryes-Kullback-Leibler divergence exceeded the value of 6. This conclusion confirms the finding mentioned earlier in the article, where Jensen-Shannon divergence values for 2005 and 2012 had been compared. As it has already been discovered earlier, this process of increase in similarity of income distribution patterns is the effect of reduction of income inequalities in countries where they were highest in 2005.

Figure 2. Jeffryes-Kullback-Leibler divergence for pairs of EU countries in 2012



Source: own calculations

CONCLUSIONS

Measures based on entropy can be a useful tool for assessment of income inequalities, as well as divergences between income distributions. Analysis based on Gini coefficient and Shannon entropy concluded that, in period 2005-2012, income inequalities decreased in majority of EU countries, with Poland, Lithuania and Portugal experiencing especially strong moves towards more egalitarian income distributions.

The use of Jensen-Shannon measure has shown that divergence of income distributions of all EU member countries decreased between 2005 and 2012. Income distributions in the EU became more similar mainly as a result of the decline of income inequalities in countries with initially high inequalities. Since disposable income after social transfers has been used as a measure of income, further research is needed in order to assess, to what extent the decline in divergence of distributions was a result of labour market outcomes, and how it had been influenced by tax and social policies.

REFERENCES

Cavanaugh J. (1999) A Large-Sample Selection Criterion Based on Kullback's Symmetric Divergence, Statistics & Probability Letters, vol. 42, p. 333–343.

Csiszár I. (1967) Information-Type Measures of Difference of Probability Functions and Indirect Observations, Studia Sci. Math. Hungar, 2, p. 299–318.

Grosse I., Bernaola-Galva P., Carpena P., Román-Roldan R., Oliver J., Stanley H.E. (2002) Analysis of Symbolic Sequences Using the Jensen-Shannon Divergence, Physical Review E, vol. 65, p. 041905-1 – 041905-16.

Jędrzejczak A. (2012) Estimation of Concentration Measures and Their Standard Errors for Income Distributions in Poland, International Advances in Economic Research, vol. 18, issue 3, p. 287-297.

Lamberti P.W., Majtey A.P., Borras A., Casas M., Plastino A. (2008) Metric Character of the Quantum Jensen-Shannon Divergence, Physical Review A, no. 77, p. 052311-1 – 052311-6.

Lefebvre G., Steele R., Vandal A. (2010) A Path Sampling Identity for Computing the Kullback–Leibler and J Divergences. Computational Statistics & Data Analysis, vol. 54, p. 1719 – 1731.

Menéndez M.L., Pardo J.A., Pardo L., Pardo M.C. (1997) The Jensen-Shannon Divergence, Journal of the Franklin Institute, vol. 134, no. 2, p. 307-318.

Menéndez M.L., Pardo J.A., Pardo L., Zografos K. (2003) On Tests of Homogeneity Based on Minimum ϕ-divergence Estimator with Constraints, Computational Statistics and Data Analysis, 43, 215–234.

Podolec B., Ulman P., Wałęga A. (2011) Próba oceny zróżnicowania spożycia artykułów żywnościowych w Polsce na podstawie wyników Badań Budżetów Gospodarstw Domowych, Acta Universitatis Lodziensis, Folia Oeconomica, vol. 253, p. 225-237.

Quintano C., Castellano R., Regoli A. (2009) Evolution and Decomposition of Income Inequality in Italy, 1991-2004,Statistical Methods and Applications, vol. 18, issue 3, p. 419-443.

Reid M.D., Williamson R.C. (2009) Generalised Pinsker Inequalities, Proc. of the 22nd Annual Conference on Learning Theory (COLT 2009), Montreal – Quebec.

Seghouane A.K., Bekara M. (2004) A Small Sample Model Selection Criterion Based on Kullback' s Symmetric Divergence. IEEE Transactions on Signal Processing, vol. 52, p. 3314 – 3323.

Shannon, C. E. (1948) The Mathematical Theory of Communications. Bell Syst. Tech. Journal, vol. 27, p. 423-467.

Taneja I.J. (2005) Refinement Inequalities Among Symmetric Divergence Measures, The Australian Journal of Mathematical Analysis and Applications, 2 (1), Art. 8, p. 1-23.

Taneja I.J. (2008) On Mean Divergence Measures, Advances in Inequalities from Probability Theory & Statistics (eds.) N.S. Barnett, S.S. Dragomir, Nova Science Publishers, p. 169–186.

Taneja I.J. (2013) Generalized Symmetric Divergence Measures and the Probability of Error. Communications in Statistics: Theory & Methods, vol. 42, p. 1654 – 1672

Taneja I. J., Kumar P. (2004) Relative Information of Type s, Csiszár's f-Divergence, and Information Inequalities, Information Sciences 166, p. 105-125.

Tomczyk E. (2011) Application of Measures of Entropy, Information Content and Dissimilarity of Structures to Business Tendency Survey Data, Przegląd Statystyczny, Zeszyt 1-2, p. 88-101.

Ullah A. (1996) Entropy, Divergence and Distance Measures with Econometric Applications, Journal of Statistical Planning and Inference, 49, p. 137 – 162.

Wędrowska E. (2011) Application of Kullback-Leibler Relative Entropy for Studies on the Divergence of Household Expenditures Structures, Olsztyn Economic Journal, vol. 6, p. 133-142.

Wędrowska E. (2012) Miary entropii i dywergencji w analizie struktur, Wyd. Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.