ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 3(302), 2014

*Justyna Wilk**

THE MEASUREMENT OF THE ECONOMIC DISTANCE ON THE BASIS OF SYMBOLIC DATA

Abstract. The economic distance defines a dissimilarity level between objects functioning in the economic space. It is one of the most important issues of spatial econometrics. However, its measurement is difficult due to the definition, description and estimation problems. The objective of the paper is to indicate the role of symbolic data in describing the economic distance and also the way of its measurement using symbolic data analysis methods. A significance of the economic distance, measurement problems, symbolic data concept and dissimilarity measures, and also an empirical example were presented in the paper.

Keywords: economic distance, dissimilarity measure, symbolic data, regional research

I. INTRODUCTION

A distance is very often referred to the physical distance. In statistical meaning, the distance determines the level of dissimilarity between patterns, objects or units. Therefore we can distinguish the cultural, social, political, economic and time distances. A significant role in modeling socio-economic phenomena (carrying out comparative studies, discovering regularities etc.) is played by the economic distance.

The measurement of the economic distance is difficult due to occurring of a few methodological problems such as relevant definition, complete description and terms of calculation. These decisions result from the research objective but also from the data availability, quality and selection and also assumptions and restrictions of statistical methods.

The objective of the paper is to indicate the role of symbolic data in describing of the economic distance and also the way of its measurement on the basis of symbolic data analysis methods. A significance of the economic distance, measurement problems, symbolic data concept and dissimilarity measures, and an empirical example were presented in the paper.

* Ph.D., Department of Econometrics and Computer Science, Wrocław University of Economics.

II. SIGNIFICANCE OF THE ECONOMIC DISTANCE

In general, the economic distance identifies a dissimilarity level between managing entities (e.g. companies, households, self-government units), offered products or services (e.g. cars, computer programs, credits) and the other objects (e.g. market segments, portfolios, social classes). For example, countries are classified as undeveloped (pre-industrial, almost entirely agrarian), developing (underdeveloped industrial base, low living standard) and developed (post-industrial) economies.

The development of spatial statistics and econometrics makes the economic distance one of the most important issues in examining the relations between territorial units, e.g. regions, cities, metropolises, countries (see Cliff and Ord (1981), Anselin (1988), Zeliaś (1991) and Suchecki *et al.* (2010)).

In the era of globalization, technological progress and other socio-economic changes, the economic distance affects relations between territorial units much more than geographical distance. For example, international trade is considerably more determined by transport costs and economic dissimilarities between countries than by the physical distance between them. Migration decisions are made by comparing the socio-economic situation (e.g. economic situations of enterprises, labour demand and supply, the costs of living, offered services etc.) of a destination region against an origin residence (see Matusik, Pietrzak and Wilk (2012)).

III. MEASUREMENT OF THE ECONOMIC DISTANCE

In many empirical studies the economic distance between territorial units is identified on the basis of Gross Domestic Product (GDP). However, the measurement of the economic distance is much more complicated. Thus, many social, economic, political and cultural issues should be considered to determine relations between territorial units. For example, socio-economic situations of regions are affected by the service sector development, investment size, labour market situation, inflow of foreign capital, access to services (see Bal-Domańska and Wilk (2011)). Therefore, the measurement of the economic distance is a research problem in the field of multivariate data analysis (see Everitt and Dunn (2001), Hair *et al.* (2006)).

Another problem concerns the complex nature of compared units. These units are usually not internally homogeneous. For example, regions are composed of sub-regions which may differ in labour market situation, economic profile etc. That is why a comparative study regarding regions' situations should be based on their sub-regions situations. Additional problem is to describe phenomena in a natural way, e.g. expected period of investment performance (e.g. from 15 to 18 months), the structure of household's expenditures (e.g. food – 20%, rent – 10%, clothes – 5%, services – 35%, other – 30%), business profile (e.g. industrial and service company). These problems may be solved with the use of symbolic data analysis (see Gatnar (1998), Bock, Diday *et al.* (2000), Billard and Diday (2006); Diday, Noirhomme-Fraiture *et al.* (2008), Wilk (2010), Gatnar, Walesiak *et al.* (2011)).

IV. SYMBOLIC DATA ANALYSIS

In symbolic data analysis, variables implementations take the form of intervals of values (interval-valued variables), sets of categories or values (multivalued variables), sets of categories with weights, frequencies, probabilities (modal variables) and also logical structures, e.g. taxonomical or hierarchical dependences (dependent variables) (see Bock, Diday *et al.* (2000), Billard and Diday (2006), Diday, Noirhomme-Fraiture *et al.* (2008)).

Therefore, symbolic data analysis offers the possibility of characterizing a situation of higher-level units (e.g. NTS-2 regions) based on the situations of lower-level units (e.g. NTS-4 regions). For example the Dolnośląskie region is composed of the Jeleniogórski, Legnicko-głogowski, Wałbrzyski, Wrocławski subregions and the city of Wrocław. This is applied for disclosing details (e.g. territorial diversity) of higher-level units. Symbolic data results from data aggregation, e.g. determination of quartiles or descriptive statistics (e.g. minimum and maximum, frequencies) on the basis of lower-level units. An approach to the construction of symbolic variables and objects for regional research was presented in Wilk (2011, 2012).

The measurement of the economic distance on the basis of symbolic data requires applying dissimilarity measures proposed in the field of symbolic data analysis. Dissimilarity measures for Boolean symbolic objects, i.e. objects described by interval-valued, multivalued and dependent variables, were presented in Bock, Diday *et al.* (2000), pp. 165–185, Diday, Noirhomme-Fraiture *et al.* (2008), pp. 126–129, Malerba *et al.* (2001), Wilk (2006b). Hausdorff's and also Chavent and Lechevallier's distance measures are applied in examining objects described by interval-valued variables. Gowda and Diday, Ichino and Yaguchi and also de Carvalho proposed measures for comparing objects described by interval-valued and multivalued variables. The majority of de Carvalho's measures also cover logical dependences. All these measures are based on Cartesian meet and join.

A separate group of dissimilarity measures was proposed for probabilistic symbolic objects, i.e. objects described by modal variables. Majority of them were previously applied in the image segmentation and for probability distributions, e.g. Kullback-Leibler divergence, Chernoff's distance, Bhattacharyya coefficient. They were adapted for symbolic data analysis (see Malerba, Esposito and Monopoli (2002), pp. 33–35, Bock, Diday *et al.* (2000), pp. 153–165, Wilk (2006a); Diday, Noirhomme-Fraiture *et al.* (2008), pp. 130–134).

V. EMPIRICAL EXAMPLE

The objective of the study was to compare the economic situations of 16 Polish regions (NTS-2) in 2010 on the basis of symbolic data. The economic profile, industry condition, investment outlays and economic situation of enterprises were considered in the investigation. Four symbolic interval-valued variables served to determine the economic distances between regions (see Table 1). They were defined on the basis of minimum and maximum values noted by subregions (NTS-3) of each region.

Table 1. The set of symbolic variables

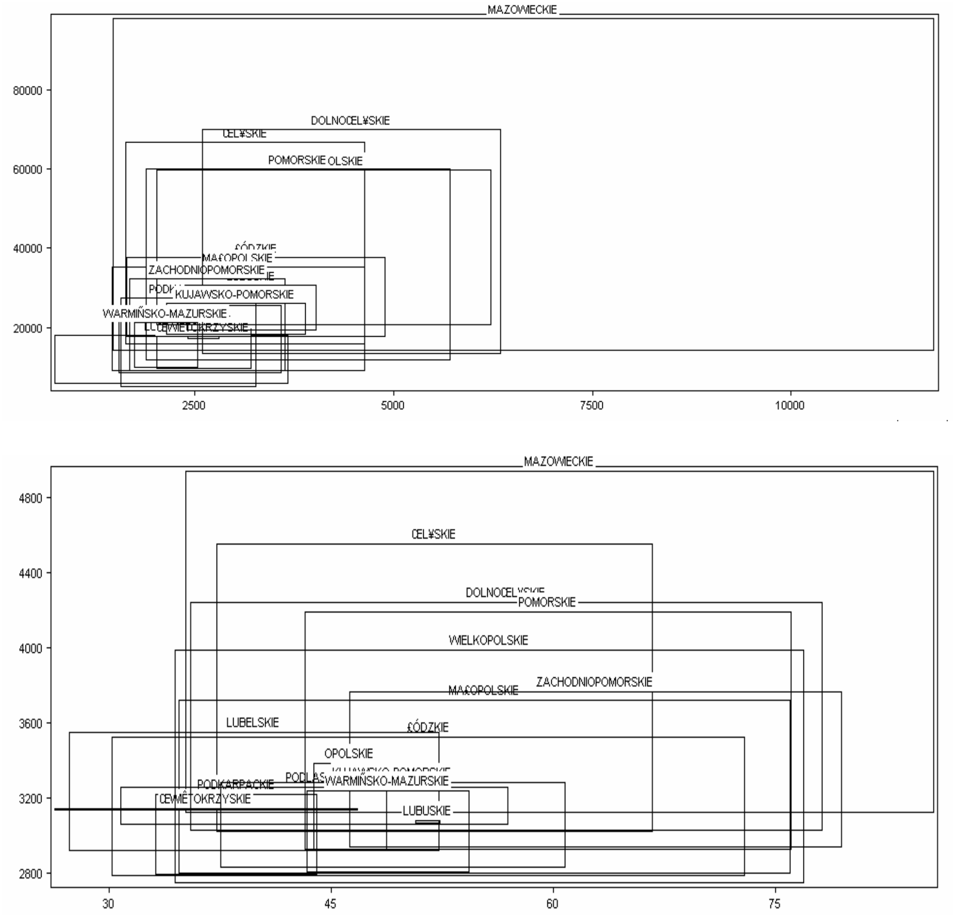
Abbreviation	Variable name	Variable implementation
Investments	Investment outlays in enterprises *per capita* [PLN]	[729.00, 11 798,00]
Services and trade	The share of people employed in services and trade to the total employed population (%)	[26.30, 85.61]
Industry	Sold industrial production *per capita* (PLN)	[5 052.00, 97 766.00]
Wages and salaries	Average monthly gross wages and salaries [PLN]	[2 746.13, 4 936.36]

Source: own elaboration based on data provided by Local Data Bank of the Central Statistical Office of Poland.

The highest territorial disparities regarding the economic situation are exhibited by the Mazowieckie region, while internally the most homogeneous but weakly developed is the Świętkorzyskie region (see Figure 1).

The Łódzkie region noted 92.6% of national average of GDP *per capita* (34 063 PLN), while the Dolnośląskie region presented 112.0% (41 194 PLN) in 2010. However GDP only partially shows the economic situations of these regions. Although both regions significantly differ in sold industrial production *per capita* and average monthly gross wages and salaries, they are very similar in respect of the economic profile (see Figure 2).

Normalized Ichino-Yaguchi distance measure was applied to determine economic disparities between Polish regions. The measure takes the values in [0, ∞], where 0 means identical objects. The shortest economic distance (0.14) is exhibited by two pairs of regions: the Łódzkie and Małopolskie regions and also the Opolskie and Warmińsko-mazurskie; while the Mazowieckie and Świętokrzyskie regions are economically the most distant (see Table 2).



a) “Industry” (the axis of ordinates) and “Investments” (the axis of abscissa)

a) “Wages and salaries” (the axis of ordinates) and “Services and trade” (the axis of abscissa)

Figure 1. Implementations of symbolic variables defining economic situations of regions

Source: own elaboration based on data provided by Local Data Bank of the Central Statistical Office of Poland.

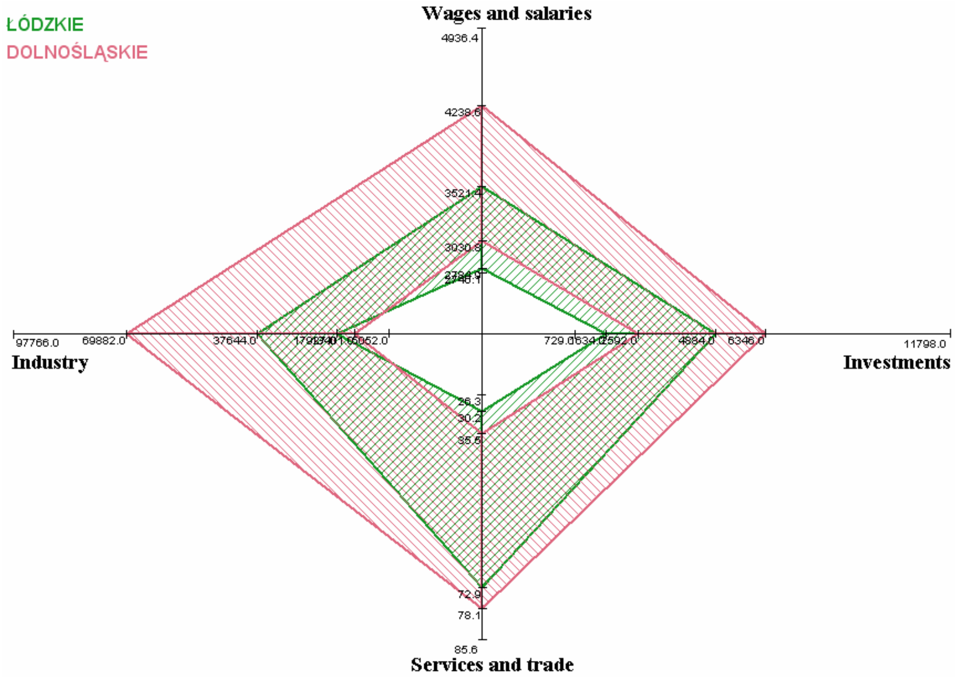


Figure 2. The comparison of economic features of the Łódzkie and Dolnośląskie regions

Source: own elaboration based on data provided by Local Data Bank of the Central Statistical Office of Poland.

VI. CONCLUSIONS

Symbolic data analysis offers a possibility to define the economic distance between complex objects, e.g. territorial units. The measurement results serve in determining the dissimilarities between objects (e.g. regional disparities). They may also be applied in multivariate data analysis methods which are based on distance matrix (e.g. cluster analysis, multidimensional scaling). They are also significant in the field of spatial econometrics to examine spatial dependences and construct the adjacency matrix and also to examine the conditions of socio-economic phenomena as an explanatory variable in gravity model.

Table 2. Distance matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	0.89														
3	**0.14**	0.85													
4	0.49	0.60	0.46												
5	0.39	1.12	0.42	0.64											
6	0.44	1.16	0.45	0.71	0.25										
7	0.35	1.09	0.40	0.63	0.21	0.23									
8	0.49	**1.20**	0.59	0.73	0.28	0.25	0.20								
9	0.55	1.14	0.57	0.66	0.41	0.34	0.34	0.35							
10	0.29	0.67	0.32	0.34	0.64	0.64	0.61	0.73	0.68						
11	0.32	0.89	0.20	0.49	0.57	0.60	0.50	0.70	0.47	0.41					
12	0.46	0.54	0.42	0.26	0.71	0.75	0.65	0.78	0.72	0.22	0.45				
13	0.49	1.11	0.47	0.62	0.26	0.21	0.27	0.32	0.23	0.65	0.41	0.70			
14	0.28	1.06	0.31	0.58	0.31	0.29	0.19	0.35	0.29	0.46	0.39	0.60	0.25		
15	0.39	0.63	0.31	0.26	0.66	0.70	0.62	0.80	0.63	0.18	0.32	0.17	0.53	0.52	
16	0.46	1.17	0.43	0.68	0.29	0.26	0.21	0.34	0.26	0.64	0.41	0.73	**0.14**	0.21	0.57

🟨 Long distance (low similarity) 🟦 Short distance (high similarity)

Explanations: 1 – Łódzkie, 2 – Mazowieckie, 3 – Małopolskie, 4 – Śląskie, 5 – Lubelskie, 6 – Podkarpackie, 7 – Podlaskie, 8 – Świętokrzyskie, 9 – Lubuskie, 10 – Wielkopolskie, 11 – Zachodniopomorskie, 12 – Dolnośląskie, 13 – Opolskie, 14 – Kujawsko-pomorskie, 15 – Pomorskie, 16 – Warmińsko-mazurskie.

Source: own estimation in symbolicDA package (Dudek, Pełka and Wilk 2013) of R-CRAN.

REFERENCES

Anselin L. (1988), *Spatial econometrics: methods and models*, Kluwer Academic, Dordrecht.

Bal-Domańska B., Wilk J. (2011), Gospodarcze aspekty zrównoważonego rozwoju województw – wielowymiarowa analiza porównawcza, *Przegląd Statystyczny*, Volume 58, Number 3–4, pp. 300–322.

Billard L., Diday E. (2006), *Symbolic Data Analysis. Conceptual Statistics and Data Mining*, Wiley, Chichester.

Bock H.H., Diday E. (Eds.) (2000), Analysis *of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Springer-Verlag, Berlin-Heidelberg.

Cliff A.D., Ord J.K. (1981), *Spatial Processes: Models and Applications*, Pion, London.

Diday E., Noirhomme-Fraiture M. (Eds.) (2008), *Symbolic Data Analysis and the SODAS Software*, Wiley, Chichester.

Everitt B.S., Dunn G. (2001), *Applied Multivariate Data Analysis*, Arnold, London.

Gatnar E. (1998), *Symboliczne metody klasyfikacji danych*, PWN, Warszawa.

Gatnar E., Walesiak M. (red.) (2011), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, C.H. Beck, Warszawa.

Hair J.F., Black W.C., Babin B.J, Anderson R.E., Tatham R.L. (2006), *Multivariate Data Analysis*, Pearson Prentice Hall, New Jersey.

Malerba D., Esposito F, Giovalle V., Tamma V. (2001), Comparing Dissimilarity Measures for Symbolic Data Analysis, In: P. Nanopoulos (Ed.), *New Techniques and Technologies for Statistics and Exchange of Technology and Know-how*, pp. 473–481.

Malerba D., Esposito F., Monopoli M. (2002), Comparing dissimilarity measures for probabilistic symbolic objects, In: A. Zanasi, C.A. Brebbia, N.F.F. Ebecken, P. Melli (Eds.), *Data Mining III*, series Management Information Systems, Volume 6, WIT Press, Southampton, pp. 31–40.

Matusik S., Pietrzak M., Wilk J. (2012), Ekonomiczne-społeczne uwarunkowania migracji wewnętrznych w Polsce w świetle metody drzew klasyfikacyjnych, *Studia Demograficzne*, Number 2(162), pp. 3–28.

Suchecki B. (Ed.) (2010), *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, C.H. Beck, Warszawa.

Wilk J. (2006a), Miary odległości obiektów opisanych zmiennymi symbolicznymi z wagami, In: K. Jajuga, M. Walesiak (Eds.), Taksonomia 13. Klasyfikacja i analiza danych – teoria i zastosowania, *Research Papers of University of Economics in Wrocław*, Number 1126, Wrocław, pp. 224–236.

Wilk J. (2006b), Problemy klasyfikacji obiektów symbolicznych. Symboliczne miary odległości, In: J. Garczarczyk (Ed.), Ilościowe i jakościowe metody badania rynku. Pomiar i jego skuteczność, *Research Papers of University of Economics in Poznań*, Number 71, University of Economics in Poznań Publishing House, Poznań, pp. 69–83.

Wilk J. (2010), Metody analizy danych symbolicznych, In: J. Dziechciarz (Ed.), Ekonometria 29. Zastosowania metod ilościowych, *Research Papers of Wrocław University of Economics*, Number 141, Wrocław, pp. 29–38.

Wilk J. (2011), Taksonomiczna analiza rynku pracy województw Polski – podejście symboliczne, In: J. Dziechciarz (Ed.), Ekonometria 34. Zastosowania metod ilościowych, *esearch Papers of Wrocław University of* Economics, Number 200, Wrocław, pp. 26–37.

Wilk J. (2012), Symbolic approach in regional analyses, *Statistics in Transition – new series*, Volume 13, Number 3, pp. 581–600.

Zeliaś A. (1991), *Ekonometria przestrzenna*, PWE, Warszawa.

Justyna Wilk

POMIAR ODLEGŁOŚCI EKONOMICZNEJ NA PODSTAWIE DANYCH SYMBOLICZNYCH

Odległość ekonomiczna określa poziom niepodobieństwa obiektów funkcjonujących w przestrzeni ekonomicznej. Stanowi jedno z najważniejszych zagadnień ekonometrii przestrzennej. Jej pomiar jest jednak utrudniony ze względu na problemy definiowania, opisu i szacowania. Celem artykułu jest wskazanie roli danych symbolicznych w opisie odległości ekonomicznej oraz sposobu jej pomiaru z wykorzystaniem metod analizy danych symbolicznych. W artykule zaprezentowano znaczenie odległości ekonomicznej, problemy jej pomiaru, koncepcję danych symbolicznych i miary odległości, a także przykład empiryczny.