M A T H E M A T I C A L   E C O N O M I C S

No. 9(16) 2013

A STUDY OF CHANGES IN THE DEVELOPMENT LEVEL OF WATER AND SEWERAGE INFRASTRUCTURE AND ENVIRONMENTAL PROTECTION IN THE COUNTIES OF THE SILESIAN VOIVODSHIP – A DYNAMIC INTERPRETATION

Anna Czopek

Abstract. In this paper the author presents an analysis of the development of the water and sewerage infrastructure and environmental protection in the counties of the Silesian voivodship on the basis of designated synthetic variables in the dynamic approach. Selected years before and after joining the European Union are considered and some interpretative hypotheses are verified by empirical research. To achieve this purpose, the problem of the selection of diagnostic variables and the descriptive characteristics to be used, as well as the determination of the of the synthetic variables values based on selected aggregation formulas, also in dynamic terms, are discussed and suitable solutions are proposed.

Keywords: classification methods, cluster analysis, agglomeration methods, diagnostic variables, synthetic variables, stimulants, destimulants.

JEL Classification: C1, C38.

DOI: 10.15611/me.2013.9.01.

1. Introduction

The main purpose of this article is an analysis of the development level of the water and sewerage infrastructure and environmental protection in a Silesian voivodship on the basis of designated synthetic variables in dynamic interpretation (selected years before and after joining the European Union). The author also presents the important problem of choosing diagnostic variables, discusses the formation of the values of the characteristics of diagnostic descriptive variables in a dynamic

Anna Czopek

The author Anna Czopek is a student of the first year of PhD studies at the University of Economics in Katowice under the guidance of prof. dr hab. Jerzy Mika.

Department of Mathematics, University of Economics in Katowice, ul. 1 Maja 50, 40-287 Katowice, Poland.

E-mail: anna_czopek@o2.pl

interpretation, and debates the received values of the chosen characteristics of synthetic descriptive variables describing the level of development of water and sewerage infrastructure and environmental protection in a dynamic interpretation.

The empirical research that was conducted allows to verify the validity of the following hypotheses:

1. The level of development of the water and sewerage infrastructure and environmental protection in the counties of the Silesian voivodship has risen on average during the researched years.
2. Since joining the EU there has been a significant increase in the disparities in the level of water and the sewerage infrastructure and environmental protection in the analyzed counties.
3. The vast majority of the counties are characterized by a lower level of development of water and sewerage infrastructure and environmental protection than the average level.

The article is divided into five chapters, ending with a summary of the study. First, the author provides an overview of the problem of the selection of diagnostic variables, along with a presentation of the final set of variables. Then the author presents selected descriptive characteristics of the variables as well as a discussion about their values over the selected years. Next, the author presents the problem of determining the values of synthetic variables based on selected aggregation formulas of diagnostic variables. Finally, the last chapter is devoted to a discussion about the obtained values of selected descriptive characteristics for a set of synthetic variables in dynamic terms. The summary refers to the hypothesis of the study.

2. The first selection of diagnostic variables

2.1. Introductory remarks

The source of the collected data is the local CSO Data Bank[1]. With the use of Microsoft Excel, a computerized data base was set up related to the researched project. It includes statistical data in the spatial (county) and dynamic (2003, 2004, 2007, 2010) terms. All analyses were performed independently using Microsoft Excel and STATISTICA PL. The values entered into the database include the accuracy with which they have been taken from the local CSO Data Bank. The analysis covered 17 counties and 19 cities with county rights of Silesia. The study deliberately selected the

[1] See: http://www.stat.gov.pl/bdr_n/app/strona.indeks, as at 5. 03. 2012 r.

values from 2003, 2004, 2007 and 2010 to show the state of the water and sewerage infrastructure and the environmental protection of Silesian counties in the year before accession to the European Union, and to obtain conclusions about the changes that occurred after accession to the EU. Not all attributes were recorded in the index, so it was necessary to change some values in order to obtain comparable data.

2.2. Choosing the initial list of variables

Citing the opinion Obrębalski contained in the book edited by Strahl (2006), and taking into account the work of Zeliaś (1983), we can distinguish three of the most general groups of criteria for the factors of regional development: substantive, formal and statistical.

The substantive selection of factors should include knowledge about the economy, industrial economics, modern economics of regions. The experience and intuition of the researcher is also essential.

For the formal criteria, the following issues should be included:

1. Measurability
2. Ensuring the comparability of the objects in space and time
3. Complete data for all objects and periods of the study.

The most important statistical criteria are:

1. A large spatial and spatial-temporal variability (coefficient of variation for the *j*-th variable $v_j \geq 10\%$)
2. Asymmetric distribution
3. No excessive correlation.

After completing the initial elimination of variables by examining the formal criterion, the fulfillment of the completeness of the data was a big problem. Despite the fact that there are many important variables in terms of content which would assess accurately the level of the development of the water and sewerage infrastructure and environmental protection, unfortunately, due to the large amount of missing data in the database of the local CSO Data Base for the selected years, they could not be taken into account. The remaining variables fulfilled the condition of measurability or comparability of characteristics. The initial set of variables is given in Table 1.

In a group of such selected variables, the statistical criteria should have been checked. All the variables were characterized by a variability greater than 10%, and the asymmetry of the distribution. Unfortunately, the problem concerned an excessive correlation of the variables. By analyzing the correlation between variables in 2003, 2004, 2007 and 2010, it can be said that in 2003 the correlations between variables were slightly different than in the other years. The direction and strength of correlation was presented for 2003 and 2010 in Table 2 and Table 3, respectively (in 2004 and 2007 the correlations between the variables were similar to those from 2010, which is why the matrix of correlation of variables from these years was not presented). The values of significant correlations are in bold.

Table 1. The initial set of variables included in the data base

Symbol of the variable	Names of variables
X_1	Number of people using the water supply network per 1000 people in the county
X_2	Number of people using the sewerage network per 1000 people in the county
X_3	Number of people using the natural gas network per 1000 people in the county
X_4	Length of water distribution network in km per 100 km^2 area of the county
X_5	Length of sewerage distribution network in km per 100 km^2 area of the county
X_6	Length of the gas distribution network in km per 100 km^2 area of the county
X_7	Number of people benefiting from treatment plant per 1000 people in the county
X_8	Dust emission in tons per 1 km^2 area of the county (per year)
X_9	Air pollution due to emissions of sulphur dioxide from plants which are especially noxious, in tons per 1 km^2 area of the county (per year)
X_{10}	Air pollution due to emissions of carbon dioxide from plants which are especially noxious, in tons per 1 km^2 area of the county (per year)
X_{11}	Gaseous pollutants retained or neutralized in equipment to reduce pollution in tons per 1000 tons of pollutants produced (per year)
X_{12}	Number of people per one sewage treatment plant
X_{13}	The amount of industrial and municipal wastewater treated per year in dm^3 to 100dm^3 of the sewage discharged.
X_{14}	Percentage of waste produced and accumulated (excluding municipal waste) recovered relative to the total number of waste generated per year in 1000 tons.
X_{15}	Waste storage area not reclaimed in ha per 1000 ha of the county area
X_{16}	Water consumption in households of one beneficiary / recipient in m^3

Source: Local CSO Data Bank.

Table 2. Correlation matrix for variables in 2003

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
X_1	1.00	**0.61**	**0.50**	**0.55**	0.40	0.43	0.42	0.32	0.22	0.25	0.29	–0.16	0.08	0.32	0.41
X_2	**0.61**	1.00	**0.75**	**0.76**	**0.72**	**0.72**	**0.80**	0.38	0.19	0.23	0.45	0.02	0.31	0.39	**0.65**
X_3	**0.50**	**0.75**	1.00	**0.68**	**0.56**	**0.74**	**0.62**	0.25	0.04	0.12	0.37	0.04	0.31	0.25	**0.71**
X_4	**0.55**	**0.76**	**0.68**	1.00	**0.91**	**0.90**	**0.74**	0.37	0.27	0.37	**0.57**	–0.02	0.10	0.42	**0.53**
X_5	0.40	**0.72**	**0.56**	**0.91**	1.00	**0.90**	**0.74**	0.26	0.17	0.25	**0.67**	–0.02	–0.02	0.27	**0.51**
X_6	0.43	**0.72**	**0.74**	**0.90**	**0.90**	1.00	**0.69**	0.26	0.15	0.23	**0.52**	0.00	0.10	0.28	**0.55**
X_7	0.42	**0.80**	**0.62**	**0.74**	**0.74**	**0.69**	1.00	0.41	0.23	0.27	**0.52**	0.13	0.09	0.33	0.46
X_8	0.32	0.38	0.25	0.37	0.26	0.26	0.41	1.00	**0.80**	**0.86**	0.24	–0.02	0.13	0.20	0.06
X_9	0.22	0.19	0.04	0.27	0.17	0.15	0.23	**0.80**	1.00	**0.93**	0.08	0.03	0.09	0.22	–0.07
X_{10}	0.25	0.23	0.12	0.37	0.25	0.23	0.27	**0.86**	**0.93**	1.00	0.17	0.08	0.05	0.26	–0.05
X_{12}	0.29	0.45	0.37	**0.57**	**0.67**	**0.52**	**0.52**	0.24	0.08	0.17	1.00	–0.25	–0.14	–0.12	0.42
X_{13}	–0.16	0.02	0.04	–0.02	–0.02	0.00	0.13	–0.02	0.03	0.08	–0.25	1.00	–0.13	–0.06	0.06
X_{14}	0.08	0.31	0.31	0.10	–0.02	0.10	0.09	0.13	0.09	0.05	–0.14	–0.13	1.00	0.17	0.22
X_{15}	0.32	0.39	0.25	0.42	0.27	0.28	0.33	0.20	0.22	0.26	–0.12	–0.06	0.17	1.00	0.05
X_{16}	0.41	**0.65**	**0.71**	**0.53**	**0.51**	**0.55**	0.46	0.06	–0.07	–0.05	0.42	0.06	0.22	0.05	1.00

Source: author’s own calculations.

Table 3. Correlation matrix for variables in 2010

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
X_1	1.00	**0.61**	**0.52**	**0.53**	0.43	0.44	0.47	0.34	0.26	0.25	0.21	–0.06	0.06	0.25	0.07
X_2	**0.61**	1.00	**0.77**	**0.79**	**0.73**	**0.73**	**0.81**	0.47	0.30	0.24	0.27	0.09	0.45	0.27	0.03
X_3	**0.52**	**0.77**	1.00	**0.75**	**0.62**	**0.77**	**0.62**	0.37	0.16	0.08	0.32	0.03	0.49	0.15	0.11
X_4	**0.53**	**0.79**	**0.75**	1.00	**0.88**	**0.90**	**0.57**	0.27	0.29	0.17	0.44	–0.08	0.26	0.14	0.20
X_5	0.43	**0.73**	**0.62**	**0.88**	1.00	**0.91**	**0.68**	0.20	0.35	0.20	0.45	0.04	0.25	0.05	0.15
X_6	0.44	**0.73**	**0.77**	**0.90**	**0.91**	1.00	**0.62**	0.22	0.20	0.12	0.43	–0.02	0.35	0.04	0.24
X_7	0.47	**0.81**	**0.62**	**0.57**	**0.68**	**0.62**	1.00	0.40	0.29	0.30	0.33	0.13	0.32	0.26	–0.05
X_8	0.34	0.47	0.37	0.27	0.20	0.22	0.40	1.00	**0.52**	**0.61**	0.12	–0.01	0.24	0.07	–0.13
X_9	0.26	0.30	0.16	0.29	0.35	0.20	0.29	**0.52**	1.00	**0.88**	0.26	0.15	0.20	0.16	0.18
X_{10}	0.25	0.24	0.08	0.17	0.20	0.12	0.30	**0.61**	**0.88**	1.00	0.30	0.21	0.20	0.17	0.10
X_{12}	0.21	0.27	0.32	0.44	0.45	0.43	0.33	0.12	0.26	0.30	1.00	0.13	0.02	–0.04	–0.04
X_{13}	–0.06	0.09	0.03	–0.08	0.04	–0.02	0.13	–0.01	0.15	0.21	0.13	1.00	0.15	–0.08	–0.02
X_{14}	0.06	0.45	0.49	0.26	0.25	0.35	0.32	0.24	0.20	0.20	0.02	0.15	1.00	0.00	0.01
X_{15}	0.25	0.27	0.15	0.14	0.05	0.04	0.26	0.07	0.16	0.17	–0.04	–0.08	0.00	1.00	0.10
X_{16}	0.07	0.03	0.11	0.20	0.15	0.24	–0.05	–0.13	0.18	0.10	–0.04	–0.02	0.01	0.10	1.00

Source: author’s own calculations.

Zeliaś (1983) writes that the variables, forming the collection of potential diagnostic variables, characterizing complex economic phenomena, most of the time are essentially correlated, so they convey similar information. One way to eliminate this disadvantage is, according to Stanisz (2007), the use of cluster analysis, which can lead to a significant reduction of strongly correlated variables.

In order to make the final selection of diagnostic variables, the set of attributes should be divided into sub-sets of similar variables, and then different groups of representatives should be selected. This division takes place in such a way that in the same group the variables are highly correlated, and in different groups the variables are weakly correlated.

2.3. Selection of the final list of variables

Cluster analysis[2] is a set of methods for the isolation of homogeneous subsets of the population of objects. The purpose of this analysis is the division of objects into a number of groups (clusters) so that objects within a group were most similar to each other, and the ones belonging to different groups were most different from each other. There are many procedures for carrying out such a grouping, however, in this analysis three agglomeration methods[3] were used: single linkage method, complete linkage method, weighted pair-group method using the centroid average – WPGMC, to compare obtained results.

As written in Grabiński et al. (1982) and Pociecha et al (1988), to measure a distance between the variables, Euclid's distance should not be used. Since the correlation coefficient r is treated as a measure of the similarity of variables, so the distance measure may be adopted:

$$d\left(X_i, X_j\right) = 1 - r, \quad i, j = 1, \ldots, m) \tag{1}$$

thanks to that, the clustering results that were obtained will be correct with a minimum degree of mutual correlation of traits.

The results of the clustering by the method of single linkage are presented in Figure 1. Other methods confirmed the obtained results in the corresponding years.

[2] Information about cluster analysis can be found in (Marek 1989), (Stanisz 2007).

[3] The description of these methods can be found in (Pociecha et al. 1988), (Stanisz 2007), (Zeliaś 2000).

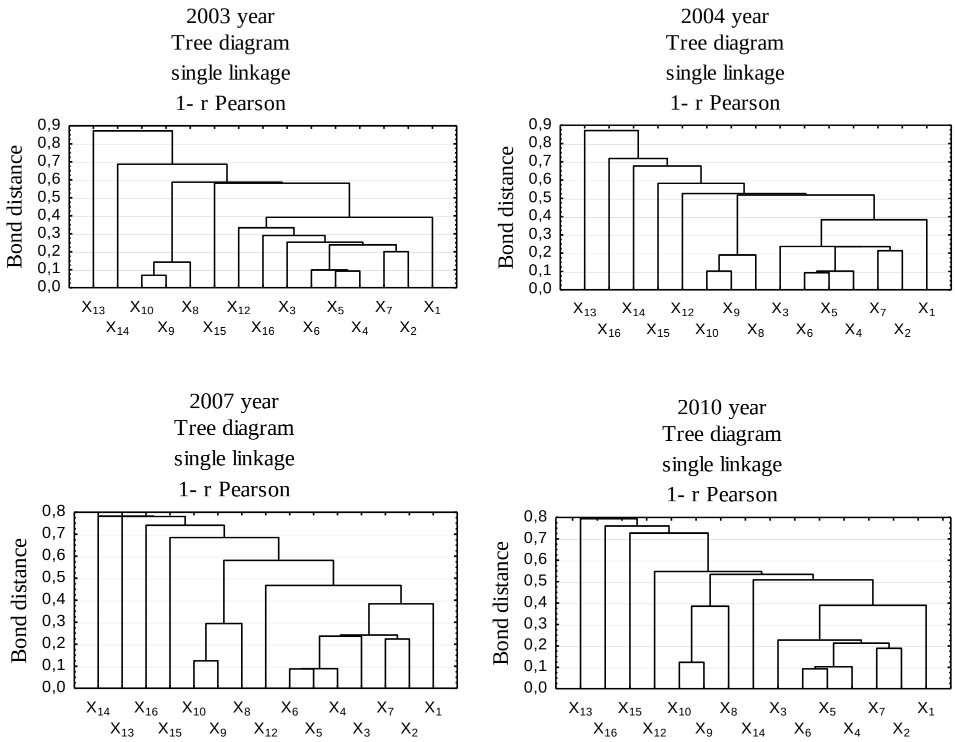

Fig. 1. Clustering results by the method of single linkage, respectively for 2003, 2004, 2007 and 2010

Source: author's own analysis using STATISTICA PL.

In 2003, five groups of variables emerged. Three one-piece groups consisted of variables X_{13}, X_{14}, X_{13} and X_{15}. The fourth group was a group of three pieces and it consisted of variables X_8, X_9, X_{10}, and the fifth one, the largest group, consisted of X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_{12}, X_{16}. In 2004, 2007 and 2010, variables were classified into seven groups. Variables X_{12}, X_{13}, X_{14}, X_{15}, X_{13}, and X_{16} formed one-piece groups. The sixth group consisted of variables X_8, X_9, X_{10}, and the seventh consisted of variables X_1, X_2, X_3, X_4, X_5, X_6, X_7.

The next step is to select representatives of each group. Here it is worth using a well-known method of choosing representatives, which can be found in (Pluta 1977), (Pociecha et al. 1988). This method states that the representative of a single-element group is precisely this variable, however in multiple groups (more than two elements) the sum of each variable's distance from the others is calculated by:

$$S_i = \sum_{j=1}^{l} d_{ij}, \quad (i = 1, \ldots, l)$$

where:

l – number of attributes in the group.

d_{ij} – the distance between i-th and j-th variable.

The attribute for which the sum of the distances is the smallest is the assumed representative of the group. To determine the distance between the variables formula (1) was used. Such specified distance ensures the proper selection of explanatory variables, since the minimum distance (equal to 0) corresponds to the maximum values of correlation coefficients (r close to 1) and vice versa. To confirm the results, the Chebyshev distance measure was used. The results are presented in Table 4.

Table 4. The sum of distance between each variables

Measure of distance		$1-r$				Chebyshev			
Group	Attributes	2003	2004	2007	2010	2003	2004	2007	2010
I	X_1	4.38	3.06	3.08	3.01	26.56	19.56	19.36	19.24
	X_2	2.55	1.62	1.70	**1.57**	18.84	13.98	14.06	13.46
	X_3	3.06	2.08	2.11	1.96	18.66	13.20	13.63	13.07
	X_4	**2.36**	**1.47**	**1.54**	1.59	**15.63**	**10.71**	**10.66**	12.29
	X_5	2.59	1.80	1.92	1.76	17.33	12.32	12.61	**12.23**
	X_6	2.54	1.63	1.65	1.63	17.84	12.27	12.35	12.69
	X_7	3.01	2.07	2.32	2.23	20.16	15.46	17.50	18.74
	X_{12}	4.18	-	-	-	23.73	-	-	-
	X_{16}	3.77	-	-	-	20.45	-	-	-
II	X_8	0.34	0.47	0.71	0.87	4.15	5.53	6.46	6.90
	X_9	0.26	0.38	0.54	0.61	3.86	5.17	**5.24**	5.49
	X_{10}	**0.21**	**0.29**	**0.42**	**0.51**	**2.94**	**4.12**	5.31	**4.49**

Source: author's own calculations.

Both methods showed that within some years in the first group the smallest sum of distances mostly belonged to the variable X_4, while in the second group to the variable X_{10}.

Thus, a set of diagnostic variables in 2003 were formed by variables X_4, X_{10}, X_{12}, X_{13}, X_{14}, X_{15}, while in 2004, 2007 and 2010 the variables X_4, X_{10}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16} belonged to the set of diagnostic variables. In order to carry out this analysis further, it is required to have the same set of final variables from 2003, 2004, 2007 and 2010, therefore, ultimately, to test the level of the development of water and sewerage infrastructure and environment, the variables X_4, X_{10}, X_{13}, X_{14}, X_{15} were used.

3. Descriptive characteristics of variables

After analyzing $X_{4,}$ it can be deducted that the average length of the water distribution network in the analyzed period increased by 14%. Right-side asymmetry indicates a larger number of counties in which the length of the water distribution network is less than the average length. There is a high variation between counties in the whole analyzed period (ca. 53%). In 2010, the longest water supply network in km per 100 km^2 were in the city counties: Chorzów (539.7 km), Świętochłowice (537.2 km) and Tychy (511.7 km), while the shortest networks were in the land counties of Żywiec (56.5 km), Lubliniec (81.7 km) and Częstochowa (83.9 km).

Table 5. Descriptive characteristics of diagnostic variables

Variable symbol	Descriptive characteristics	2003	2004	2007	2010
X_4	$\bar{x}$	248.99	252.54	264.25	283.44
	Vs (%)	53.3252	53.3588	53.8444	53.9298
	As	0.45384	0.42960	0.46679	0.21793
X_{10}	$\bar{x}$	7587.81	8307.00	8597.77	7661.58
	Vs (%)	181.1806	183.7854	186.5355	184.2988
	As	2.84422	2.77605	2.33068	2.48544
X_{13}	$\bar{x}$	86.49	87.83	88.40	88.49
	Vs (%)	21.0108	21.5038	21.0039	22.0456
	As	–2.73981	–2.85411	–2.97171	–2.22295
X_{14}	$\bar{x}$	86.06	85.54	83.35	87.84
	Vs (%)	27.6095	29.2821	32.6202	25.8735
	As	–2.62116	–2.42432	–1.98378	–2.98398
X_{15}	$\bar{x}$	5.00	4.67	3.93	3.38
	Vs (%)	140.3512	136.4119	152.9570	167.4343
	As	2.18965	2.03922	2.00015	2.18947

Source: author's own calculations.

Considering the descriptive characteristics of the variable X_{10} it can be said that from 2003 to 2007 the average carbon dioxide emissions increased by over 13%, but already by 2010 they decreased by 11%, and in consequence returned to a level similar level to that of 2003. A very high coefficient of variation (over 180%) clearly demonstrates the in-depth difference between the counties. There are several counties that emit enormous amounts of carbon dioxide compared with the rest. In 2010, it was quite dramatic in this regard in the city counties: Rybnik (57713.61 tons per 1km^2 per year), Jaworzno (46716.24 tons per 1km^2 per year) and Dąbrowa Górnicza (41787.81 tons per 1km^2 per year). The lowest emissions of this chemical in 2010 were in the land counties: Kłobuck (16.36 tons per 1km^2 per year) and Lubliniec (29.30 tons per 1km^2 per year). Nevertheless, high right-side asymmetry highlights the fact that the majority of counties emits less carbon dioxide than the average emission.

In the period of 2003-2010 there was a slight average increase (just over 2%) in the amount of industrial and municipal wastewater treated per year in dam^3 per 100 dam^3 sewage discharged. Differentiation between counties is maintained at a constant level and is not too high. The very strong left-side asymmetry proves that most counties have more treated wastewater from the average amount of treated sewage. In 2010, the city counties Chorzów, Świętochłowice, Jatrzębie Zdrój, Tychy, Żory and the land counties Cieszyn, Kłobuck were treating 100% of sewage, and such city counties as Siemianowice Śląskie and the land counties Rybnik and Wodzisław Śląski properly cleared only 25.99 dam^3, 34.18 dam^3 and 34.90 dam^3 per 100 dam^3 effluent discharged.

Considering the variable X_{14}, it can be seen that the percentage of waste produced and accumulated (excluding municipal waste) which had been recovered, relative to the total amount of waste generated per year from 2003 to 2010, underwent slight fluctuations. First, from 2003 to 2007 it declined slightly, then began to increase. Finally, the average amount of recycled waste from 2003 to 2010 in the voivodship of Silesia increased by 2%. The coefficient of variation has also changed. Up to 2010 the lowest level of differentiation can be seen between counties in comparison with other years, although counties with extreme values of this variable can be seen, for example: the city county Siemianowice Śląskie recycled 100% waste, and the land county Kłobuck was not concerned about the problem of recycling. Persistent high left-side asymmetry indicated that the greater number of counties recycled more waste than the average.

A good phenomenon is that in 2003-2010 the average size of non-reclaimed waste storage decreased (by 32%). The very high coefficient of variation throughout the study period demonstrates significant differences between counties. What is more, these differences are steadily widening. In 2010, the city counties, such as Jastrzębie Zdrój, Ruda Śląska and Gliwice had not reclaimed waste storage areas equal to: 20.65 ha, 19.98 ha, 17.96 ha, respectively. These cities really stand out from the rest because of the tested variable. High right-side asymmetry indicates that the majority of counties have less storage space for non-reclaimed waste than the average.

4. Synthetic variables

4.1. Introductory remarks

The level of the development of water and sewerage infrastructure and environmental protection can be studied by analyzing the synthetic variable (aggregate variable), which allows movement from a multi-dimensional description of the analyzed objects into a one-dimensional. Variables belonging to the final set of diagnostic attributes are the basis for construction of a synthetic variable. This allows for the possibility of comparing and arranging objects studied from the viewpoint of the analyzed phenomenon.

4.2. Changing destimulants into stimulants

Prior to the designation of synthetic variables, it is necessary to transform the destimulant variables into stimulant variables. If the higher values of the attribute allow to qualify an object (county) as a better one in accordance with the adopted general criterion, then such an attribute is called a stimulant. However, if larger values of the attribute imply a worse assessment of the object in view of the general criterion, then such an attribute is called a destimulant[4]. The set of diagnostic variables that is taken for analysis consists both of stimulants and destimulants The stimulants are: X_4, X_{13}, X_{14}, while the destimulants are: X_{10}, X_{15}.

Destimulants have been converted into stimulants by the following formula:

$$`x_{ij} = c_j - x_{ij}$$

[4] Information about stimulants and destimulants can be found in (Strahl 2006), (Zeliaś 2000).

where:

x_{ij} – realization of the variable destimulant

c_j – a constant satisfying the condition

$$c_j \geq \max_i \{x_{ij}\}.$$

4.3. Normalization of variables

The next step is to achieve comparability between the variables. In this work the variables were normalized using the formula (Pociecha et al. 1988, Zeliaś 2000):

$$z_{ij} = \frac{x_{ij}}{x_{0j}}, \quad (i = 1, \ldots, 36;\ j = 1, \ldots, 5)$$

where:

z_{ij} – normalized value of j-th variable for i-th county

x_{ij} – value of j-th variable for i-th county

x_{0j} – point of reference for j-th diagnostic variable.

In this study, two variants of the reference points were adopted (Zeliaś 2000). In the first variant, the reference point is the average value of the j-th diagnostic variable in the t year, which is selected in subsequent years (the variable pattern). However, in the second variant, the reference point is the average value of the diagnostic variable X_j, in the first year of analysis, i.e. in 2003 (the fixed pattern).

4.4. Determination of values of synthetic variables

After determining the matrix of the standardized values of diagnostic variables (two matrices due to the two variants of the reference points), the value of synthetic variables was calculated using the formula:

$$z_i = \sum_{j=1}^{5} z_{ij}, \qquad (i = 1, \ldots, 36).$$

Both synthetic variables determining patterns were used in order to compare the obtained results.

The calculated values of synthetic variables in 2003, 2004, 2007 and 2010 with the use of the variable pattern and fixed pattern are shown in Table 6.

Table 6. The values of synthetic variables in 2003, 2004, 2007 and 2010

No.	Land Counties (a) / City Counties (b)	Variable pattern				Fixed pattern			
		2003	2004	2007	2010	2003	2004	2007	2010
		z_i	z_i	z_i	z_i	z_i	z_i	z_i	z_i
1	Bielsko-Biała (a)	5.3445	5.3807	5.5348	5.0810	5.3445	5.5419	5.3341	5.0083
2	Cieszyn (a)	5.0518	5.0651	5.1625	5.0924	5.0518	5.2254	4.9388	4.9785
3	Żywiec (a)	4.7520	4.7174	4.8674	4.7647	4.7520	4.8711	4.6163	4.6003
4	Bielsko-Biała (b)	5.8561	5.8443	6.3481	6.0672	5.8561	6.0100	6.2111	6.1221
5	Lubliniec (a)	4.7406	4.6505	4.5239	4.6277	4.7406	4.8066	4.2847	4.4706
6	Tarnogóra (a)	4.4307	4.3915	4.6649	4.5943	4.4307	4.5457	4.4299	4.4502
7	Bytom (b)	5.7305	5.9407	6.0441	5.9108	5.7305	6.0932	5.9068	5.9597
8	Piekary Śląskie (b)	5.2939	5.3804	5.4512	5.4264	5.2939	5.5322	5.2873	5.3772
9	Częstochowa (a)	4.5808	4.6178	4.8935	4.4166	4.5808	4.7726	4.6483	4.2572
10	Kłobuck (a)	3.8724	3.8410	3.8497	3.7954	3.8724	4.0055	3.6476	3.6297
11	Myszków (a)	3.9660	3.9045	3.8993	4.7134	3.9660	4.0678	3.6955	4.5745
12	Częstochowa (b)	5.8944	5.8115	5.8685	5.7571	5.8944	5.9765	5.6987	5.7521
13	Gliwice (a)	4.5709	4.6460	4.6772	4.6017	4.5709	4.7877	4.4765	4.4889
14	Gliwice (b)	4.7465	4.6827	3.6222	4.3893	4.7465	4.8034	3.5643	4.4241
15	Zabrze (b)	5.7559	5.7803	5.9013	5.6176	5.7559	5.9384	5.7312	5.6263
16	Chorzów (b)	5.1369	4.8952	5.1617	5.8272	5.1369	5.0095	5.1460	5.9632
17	Katowice (b)	5.2703	5.3020	5.2275	5.5007	5.2703	5.4520	5.0564	5.5249
18	Mysłowice (b)	5.0156	5.0894	4.9892	5.9941	5.0156	5.2414	4.7972	6.0602
19	Ruda Śląska (b)	4.3131	4.4134	4.6062	4.5188	4.3131	4.5054	4.5716	4.6066
20	Siemianowice Śląskie (b)	4.9399	4.7759	5.1499	5.1579	4.9399	4.9384	4.9533	5.1445
21	Świętochłowice (b)	5.7940	5.8687	5.8070	6.2724	5.7940	6.0267	5.7417	6.3519
22	Racibórz (a)	4.9203	4.8743	4.3651	4.5989	4.9203	5.0342	4.1567	4.4577
23	Rybnik (a)	4.2953	4.2414	4.2897	3.3651	4.2953	4.3810	4.0672	3.2501
24	Wodzisław (a)	5.3946	4.8750	4.8172	4.4663	5.3946	5.0278	4.6526	4.4044
25	Jastrzębie-Zdrój (b)	4.6710	5.1698	5.0030	4.5467	4.6710	5.2840	4.9684	4.6617
26	Rybnik (b)	4.3843	4.3901	4.7307	4.6258	4.3843	4.4697	4.6750	4.7549
27	Żory (b)	5.3949	5.6267	5.4747	5.5747	5.3949	5.7921	5.2898	5.5372
28	Będzin (a)	5.0265	5.0922	5.0528	4.9994	5.0265	5.2413	4.8560	4.9335
29	Zawiercie (a)	4.8224	4.7856	4.7830	4.8808	4.8224	4.9405	4.5410	4.7354
30	Dąbrowa Górnicza (b)	4.8879	4.7065	4.4073	4.3185	4.8879	4.8372	4.2774	4.3236
31	Jaworzno (b)	4.3533	4.3146	3.8176	3.7814	4.3533	4.4098	3.7670	3.8523
32	Sosnowiec (b)	6.1283	5.9238	6.0201	5.8939	6.1283	6.0876	5.8920	5.9674
33	Mikołów (a)	4.6639	4.6333	4.6055	4.5206	4.6639	4.7527	4.4988	4.5605
34	Pszczyna (a)	5.0107	5.0418	5.0743	5.0163	5.0107	5.1938	4.8692	4.9271
35	Bieruń-Lędzin (a)	4.8574	5.2383	5.2169	5.1067	4.8574	5.3904	5.0331	5.0363
36	Tychy (b)	6.1326	6.0875	6.0920	6.1783	6.1326	6.2554	5.9423	6.2548

Source: author's own calculations.

5. Discussion about the results of empirical research

In order to examine what changes have occurred in the development level of the water and sewerage infrastructure and environmental protection in 2003, 2004, 2007 and 2010 due to the selected variables, Table 7 contains important descriptive characteristics of the synthetic variables describing the issue in question.

Table 7. Descriptive characteristics of synthetic variables

Descriptive characteristics	Variable pattern				Fixed pattern			
	2003	2004	2007	2010	2003	2004	2007	2010
Arithmetic mean	5.00000	5.00000	5.00000	5.00000	5.00000	5.14581	4.83955	4.97300
Median	4.93007	4.88508	4.99610	4.94010	4.93007	5.03101	4.82660	4.84100
Minimum	3.87243	3.84096	3.62219	3.36507	3.87243	4.00554	3.56429	3.25012
Maximum	6.13257	6.08752	6.34808	6.27243	6.13257	6.25536	6.21110	6.35194
Gap	2.26015	2.24656	2.72589	2.90736	2.26015	2.24982	2.64681	3.10181
The standard deviation	0.57855	0.58909	0.68205	0.71453	0.57855	0.59884	0.68614	0.76091
Coefficient of variation	11.5710	11.7818	13.6410	14.2907	11.5710	11.6375	14.1777	15.3007
Asymmetry	0.24661	0.14737	–0.0035	–0.0394	0.24661	0.17041	0.08193	0.05991

Source: author’s own calculations.

Based on the results obtained after applying the variable pattern as shown in Table 7, it can be concluded that the development of water and sewerage infrastructure and environmental protection in Silesia analyzed by selected variables in selected years on average did not change. This conclusion is confirmed by the results obtained using the fixed pattern (here the average varies slightly). Even Poland’s accession to the European Union has not changed this situation.

It is also noticeable that the range of variability of the value of synthetic variable (gap, standard deviation, coefficient of variation) increased during this period. Over the years, following the permanent increase in diversity of the Silesian counties due to the studied variables, the divergence between the counties is increasing. The results obtained from both patterns are unanimous that in 2003 and 2004 the most developed city county was Tychy, Bielsko-Biała in 2007 and Świętochłowice in 2010. The lowest level of development in 2003 and 2004 was registered in the land county Kłobuck, in 2007 the city county Gliwice and in 2010 in a the land county Rybnik.

In 2003, the asymmetry coefficient equal to $As = 0.24661$, demonstrated that a slightly larger number of the counties was characterized by a level of the development of water and sewerage infrastructure and environmental protection lower than average (19 counties of the 36 taken for analysis). However, in subsequent years (as proved by the results for the fixed pattern) asymmetry is very small (slightly deflected from zero), it can be said that the synthetic variables were characterized by a moderate synthetic distribution.

6. Summary

In reference to the hypotheses presented at the beginning of the article, following the above analysis it can be stated that:

1. The level of the development of water and sewerage infrastructure and environmental protection in the counties of the Silesian voivodship has not, on average, increased in the analyzed years. It was relatively constant.

2. Since joining the EU there has been a significant increase in the differentiation in the development level of water and sewerage infrastructure and environmental protection in the analyzed counties. The divergence of regional development is constantly widening.

3. Only in 2003 a greater number of counties showed a lower development level of water and sewerage infrastructure and environmental protection than the average. In subsequent years it can be seen that synthetic variables were characterized by a moderate synthetic distribution.

References

Grabiński T., Wydymus S., Zeliaś A. (1982). *Metody doboru zmiennych w modelach ekonometrycznych*. PWN. Warszawa.

Grabiński T. (1984). *Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk ekonomicznych*. Zeszyty Naukowe Akademii Ekonomicznej nr 61. Wydawnictwo Akademii Ekonomicznej. Kraków.

Jajuga K. (1993). *Statystyczna analiza wielowymiarowa*. PWN. Warszawa.

Kudrycka I. (2009). *Wpływ funduszy unijnych na wzrost gospodarczy i rozwój regionów w Polsce*. Vizja Press& IT. Warszawa.

Marek T. (1989). *Analiza skupień w badaniach empirycznych. Metody SAHN*. PWN. Warszawa.

Młodak A. (2006). *Analiza taksonomiczna w statystyce regionalnej*. Difin. Warszawa.

Pluta W. (1977). *Wielowymiarowa analiza porównawcza w badaniach ekonomicznych*. Państwowe Wydawnictwo Ekonomiczne. Warszawa.

Pociecha J., Podolec B., Sokołowski A., Zając K. (1988). *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. PWN. Warszawa.

Stanisz A. (2007). *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*. T. 3. StatSoft Polaka. Kraków.

Strahl D. (Ed.). (2006). *Metody oceny rozwoju regionalnego*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu. Wrocław.

Wydymus S. (1984). *Metody wielowymiarowej analizy rozwoju społeczno-gospodarczego*. Zeszyty Naukowe Akademii Ekonomicznej nr 62. Wydawnictwo Akademii Ekonomicznej. Kraków.

Zeliaś A. (Ed.). (1983). *Globalne prognozy rozwoju społeczno-gospodarczego*. PWN. Warszawa.

Zeliaś A. (Ed.). (2000). *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*. Wydawnictwo Akademii Ekonomicznej. Kraków.