Milan Stamenković[1]
Mirko Savić[2]

JEL: C38, C43, R11
DOI: 10.5937/industrija45-14483
UDC: 332.133.4(497.11)
005.311.121:330.133.7
Original Scientific Paper

Measuring regional economic disparities in Serbia: Multivariate statistical approach

Article history:
Received: 8 July 2017
Sent for revision: 25 July 2017
Received in revised form: 8 September 2017
Accepted: 8 September 2017
Available online: 10 October 2017

***Abstract**: The identification of regional economic disparities and their extent is an important factor affecting regional development policy formulation. In this work we propose an alternative, multivariate statistical methodology for evaluation of level of economic development of districts in Serbia, and their classification in homogeneous groups, based on five economic indicators. First, the new composite indicator for measuring economic development level (IED) is created using factor analysis, and then the districts were classified according to the obtained IED values. The evaluation of structural quality of thus formed groups was conducted using the non-hierarchical clustering procedure. The approach presented in this paper takes account of statistical assumptions on which the valid application of multivariate methods is based, which makes it advantageous over the current approaches in the literature. The resulting categorization into three district groups clearly confirms the presence of very pronounced regional economic disparities between the less developed districts in southern and eastern part and more developed districts in northern part of Serbia. Districts with city of Belgrade and Novi Sad occupy the dominant positions compared to other districts.*

***Keywords**: multivariate statistical analysis, composite indicator, regional economic development, regional disparities, districts, Serbia*

[1] University of Kragujevac, Faculty of Economics, m.stamenkovic@kg.ac.rs
[2] University of Novi Sad, Faculty of Economics in Subotica

Merenje regionalnih ekonomskih dispariteta u Srbiji: Multivarijacioni statistički pristup

***Apstrakt**: Identifikovanje regionalnih ekonomskih disparitetima i njihovih razmera predstavlja važan faktor pri formulisanju regionalne razvojne politike. U ovom radu, predložena je alternativna, multivarijaciona statistička metodologija za merenje stepena ekonomske razvijenosti okruga u Srbiji i njihovu klasifikaciju u homogene grupe, korišćenjem pet ekonomskih pokazatelja. Nakon kreiranja novog kompozitnog pokazatelja za merenje stepena ekonomske razvijenosti (IED), primenom faktorske analize, izvršena je klasifikacija okruga na bazi njegovih vrednosti. Evaluacija kvaliteta formiranih grupa u pogledu njihove strukture sprovedena je korišćenjem nehijerarhijske procedure grupisanja. Predstavljeni pristup se razlikuje od postojećih u literaturi budući da uvažava statističke pretpostavke na kojima se zasniva validna primena korišćenih multivarijacionih metoda. Rezultirajuća kategorizacija, sačinjena iz tri grupe okruga, jasno potvrđuje prisustvo izraženih regionalnih ekonomskih dispariteta između manje razvijenih okruga u istočnom i južnom delu i više razvijenih okruga u severnom delu Srbije. Pored navedenog, okruzi u čijem sastavu su grad Beograd i Novi Sad zauzimaju vodeće pozicije u poređenju sa ostalim okruzima.*

***Ključne reči**: multivarijaciona statistička analiza, kompozitni pokazatelj, regionalna ekonomska razvijenost, regionalni dispariteti, okruzi, Srbija*

1. Introduction

One of the most important economic policy issues in each country is to ensure a balanced economic development throughout its territory, as the primary condition for successful integration of individual countries and their regions into global economic trends. The Republic of Serbia is characterized by very pronounced regional development disproportions and asymmetries, which constantly tend to increase (Government of the Republic of Serbia, 2007; Molnar, 2013; Vukmirović, 2013). After decades of neglect and marginalization, a significant step in recognizing the importance and necessity of ensuring a balanced and sustainable regional development in Serbia was made in 2007 and 2009, by adopting the Regional Development Strategy and the Law on Regional Development, respectively. This provided an adequate institutional framework for implementation of principles and mechanisms of regional development management based on modern (European) concepts, and effective reduction of highly-present intra-regional and inter-regional inequalities that jeopardize the socio-economic development of Serbia at national and international level.

Objective measurement and understanding of the achieved level of development of territorial units of different administrative levels (i.e. regions, districts, municipalities), with respect for different aspects (dimensions) of development potential / limitation (i.e. economic, social, infrastructural, demographic, environmental) is the key prerequisite on which strategic planning of balanced regional development and effective implementation of regional policy are based (Government of the Republic of Serbia, 2007). Starting from the multidimensional nature of the balanced regional development concept, the research of regional characteristics mainly involves measurement of a large number of indicators of different development dimensions necessary for perception of the degree of general development of territorial units, their comparison and classification, thus creating a suitable basis for formulating appropriate strategies, development policies, and adequate allocation of incentives and other support instruments by the official authorities and institutions. This approach to the measurement of regional asymmetries has led to the shift of analytical framework from one-dimensional monitoring of values of a large number of individual indicators of different dimensions to the development and application of various multidimensional methodologies for construction of specific composite indicators for measuring poverty, human development, social cohesion / exclusion, quality of life, demographic resources, degree of development vulnerability and socio-economic development of territories within the country (National Agency for Regional Development (NARD), 2013). However, in spite of the indisputable importance of all other dimensions of regional development and their indicators, it is necessary to emphasize the key role of economic development indicators in assessing the level of development of territorial units. Being indispensable when it comes to alleviating social inequalities, balancing demographic trends and environmental protection, a balanced regional development is primarily economic issue, since "without economic equality there is no national equality" (Government of the Republic of Serbia, 2007).

In addition to the aforementioned, this study was also motivated by the following: 1) there is a possibility to extend the existing methodological framework in researching the structure and modeling of relations within multidimensional economic phenomena, and thus issues of regional characteristics and economic inequalities by applying multivariate statistical methods, 2) it is necessary to demonstrate the advantages of using composite (multivariate) indicators against individual (univariate) analysis of a large number of separate indicators, 3) there is insufficient transparency regarding the proposed methodologies for construction of composite indicators used in theory and domestic practice, as well as the absence of an empirical and statistically based evaluation of their validity, and 4) there is a lack of research focus on the analysis of the degree of economic development of territorial units at district level (NUTS-3) in Serbia.

Thus, the subject of the research is to examine and present the possibilities of applying the methods of multivariate statistical analysis (MVA) in modeling relations between the relevant indicators of economic development of territorial units at district level in the Republic of Serbia. The first objective of the study is to develop an alternative approach to assessing the degree of economic development of districts in Serbia based on a simultaneous examination of the interdependence of five representative economic indicators and creation of their composite combination in the form of the Index of Economic Development (*IED*), using appropriate MVA methods. The second objective is to classify districts, based on *IED* values, into statistically evaluated, internally homogeneous and externally heterogeneous groups, for identifying the extent of the present regional disparities.

The practical benefit of the created multivariate classification of districts is that it enables a simpler and more practical, but equally informative comparison of territorial units according to their level of economic development in comparison to the individual, isolated interpretation of the individual economic indicators' values. Accordingly, the created "multidimensional snapshot" of economic development of the observed territorial units can represent a suitable basis for subsequent analyzes of regional asymmetries from the angle of the remaining, important development dimensions (social, demographic, infrastructural, ecological) and their interdependence with the achieved economic development, thus providing information useful for decision makers and experts in the field of planning and implementation of measures and policies of the regional development strategy.

2. Research background

The analysis of degree of development of different groups of countries and / or territorial units in their composition, using various combinations of indicators of one or several dimensions of development, and, accordingly, an appropriate classification of observed territories into relatively homogeneous groups, in order to identify the presence and / or the extent of regional disparities, represents an attractive research niche for both scientists and professionals, as evidenced by a significant number of published works, and conducted empirical studies. Most of the proposed approaches in assessing the level of regional development are predominantly based on individual or combined application of appropriate multivariate statistical methods. Accordingly, *Table 1* shows the key methodological determinants of selected empirical research relevant for the research presented in this paper.

Most of the presented papers focus on the use of factor analysis in dimensionality reduction regarding the number of used indicators through the extraction of latent variables, i.e., linear combinations of indicators of different

aspects of regional development (economic, demographic, social, technological, agricultural, quality of life, etc.), and creation of classification of territorial units based on the calculated factor scores and application of cluster analysis (e.g. Cziraky, Puljiz, Jurlin, Maleković & Polić, 2003; Cziraky, Sambt, Rovan & Puljiz, 2006; del Campo, Monteiro & Soares, 2008; Melecky, 2014; Maletić & Bucalo-Jelić, 2016; Rašić-Bakarić, 2006; Soares, Marques & Monteiro, 2003; Vydrova & Novotna, 2012; Xiaoxin, 2012). Individual application of cluster analysis for classification of observed territories into internally homogenous groups, according to the selected group(s) of indicators, is presented in Aumayr (2006), Avram & Postoiu (2016), Brauksa (2013), Kvičalova, Mazalova & Široky (2014), Lepojević, Bošković & Janković-Milić (2015), Michaelides, Economakis & Lagos (2006), Popović & Maletić (2008), Rovan & Sambt (2003).

Table 1: Comparative review of the relevant multivariate empirical studies

Author(s) / (year of publication)	Administrative level of territorial units	State(s)	Development dimension(s)	MVA method(s)
Cziraky et al. (2003)	Municipalities	CRO	Ec./Soc./Dem.	FA/CA
Rovan & Sambt (2003)	Municipalities	SLO	Ec./Soc./Dem./Edu.	CA
Soares et al. (2003)	Municipalities	POR	Ec./Soc./Dem./Edu.	FA/CA
Aumayr (2006)	NUTS-3	EU-25	Ec./Dem.	CA
Cziraky et al. (2006)	Municipalities	CRO/SLO	Ec./Soc./Dem.	FA/CA
Michaelides et al. (2006)	NUTS-2	GRE	Ec.	CA
Ozaslan et al. (2006)	NUTS-2	TUR	Ec./Soc./Dem./Edu.	PCA
Rašić-Bakarić (2006)	Municipalities	CRO	Ec./Dem./Edu.	FA/CA
Kurnoga-Živadinović (2007)	NUTS-3	CRO	Ec./Soc.	CA/FA/DA
del Campo et al. (2008)	NUTS-2	EU-25	Ec./Dem./Edu.	FA/CA
Popović & Maletić (2008)	Municipalities	SRB	Agric.	CA
Goletsis & Chletsos (2011)	NUTS-2	GRE	Ec./Soc.	FA/CA
Pintilescu (2011)	NUTS-2	ROU	Ec.	PCA
Vydrova & Novotna (2012)	NUTS-3	CZE	QL.	FA/CA/MAV
Xiaoxin (2012)	Provinces	PRC	Ec./Soc./Dem.	FA/CA
Brauksa (2013)	Municipalities	LVA	Ec./Soc.	CA
Igić (2014)	NUTS-3	SRB	Ec./Dem./Tech./Env.	FA
Kvičalova et al. (2014)	NUTS-3	CZE	Ec./Soc.	CA
Melecky (2014)	NUTS-2	EU-28	Ec./Soc./Infr./Tech.	FA/CA
Perišić (2014)	NUTS-3 & municipal.	CRO	Ec./Soc./Dem./Edu.	CA/DA
Lepojević et al. (2015)	Municipalities	SRB	Ec./Dem./Edu.	CA
Avram & Postoiu (2016)	NUTS-2	EU	Ec.	CA
Maletić & Bucalo-Jelić (2016)	Municipalities	SRB	Agric.	FA/CA

Meaning of abbreviations in column – MVA methods:
CA (Cluster Analysis), FA (Factor Analysis), DA (Discriminant Analysis), MAV (Multivariate Analysis of Variance), PCA (Principal Component Analysis);
Meaning of abbreviations in column – Dimensions (aspects of regional development):
Ec. (Economic), Soc. (Social), Dem. (Demographic), Infr. (Infrastructure), Tech. (Technological), Edu. (Education), QL. (Quality of Life), Env. (Environmental), Agric. (Agricultural development)

Source: Authors' representations

Meta-analysis of studies presented in *Table 1* suggests that a small number of papers analyze regional economic disparities exclusively. Adding indicators of other dimensions of regional development can blur the general picture in terms of placement of territorial units as well as in terms of clear understanding of particular indicator and/or group(s) of indicators, which conditioned it. In other words, the lag in one dimension can be compensated in other dimensions and therefore result in a distorted picture of the level of socio-economic development of that territory (Perišić & Wagner, 2015). The classifications created in this way can not provide precise ranking of territories on the basis of certain, individual aspects of regional development, but only their aggregate position, since the impact of individual development dimensions is incorporated in general categorization.

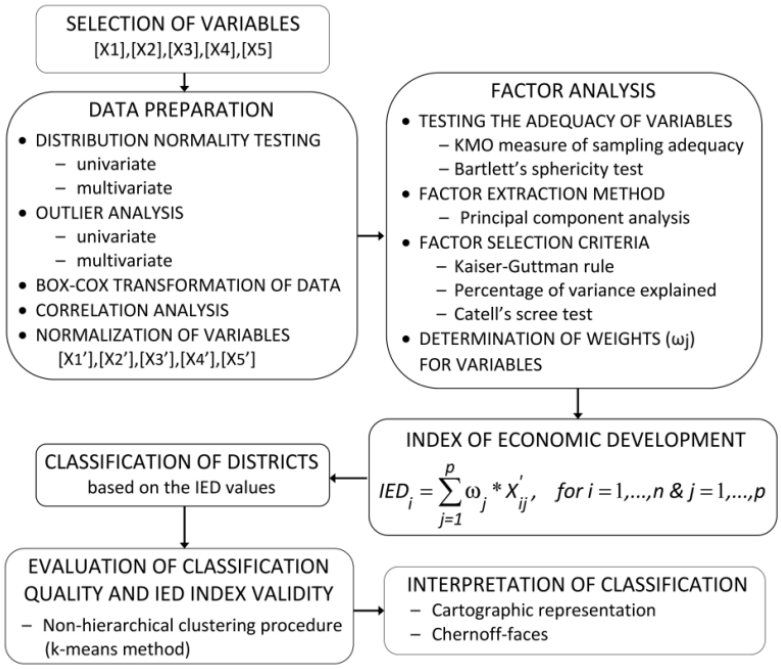
On the other hand, Goletsis & Chletsos (2011), Igić (2014), Ozaslan, Dincer & Ozgur (2006) used different multivariate methodological approach based on the application of factor analysis or principal component analysis for determination of the weights for individual indicators and construction of specific composite indicators of regional development, thus providing precise quantification of the degree of achieved regional development regarding its different aspects. This created the basis for ranking the territorial units according to the calculated values of created composite indicators. The importance and applicative potential of multivariate methods of statistical analysis in the construction of composite indicators, and the creation of a more objective approach in determining the weights of individual indicators that constitute their structure, are presented and explained in detail in Perišić & Wagner (2015). More precisely, through a critical review of the methodological approach in calculating the index of development of Croatian territorial units (Government of the Republic of Croatia, 2010), these authors emphasize the role of MVA methods in eliminating subjectivity in determination of weights and the importance of ensuring the transparency of the complete process of construction of composite indicators. In addition, according to the authors' opinion, and observations made by Cziraky et al. (2006) and Perišić (2014), it is important to emphasize that one of the main disadvantages of most of the previous studies, is the fact that they neglect or incompletely test the fulfillment of appropriate statistical assumptions, which cast shadow on validity of MVA methods application.

3. Research methodology framework

In addition to the traditional univariate statistical methods, we also extensively applied several methods of multivariate statistical analysis aimed at examination of simultaneous interdependencies among individual economic indicators and discovery of an "unknown" structure hidden in the observed set

of multivariate data. A schematic representation of the research methodology framework, complemented by detailed explanations of the implemented univariate and multivariate statistical methods, is presented in *Figure 1*.

Figure 1: Schematic representation of research methodology framework



Source: Authors' representations

After the selection of individual indicators of economic development, in accordance with the detailed explanation of procedure, guidelines and critical activities in construction of composite indicators presented in the relevant literature (Salzman, 2003; OECD, 2008; Perišić & Wagner, 2015), and adequate verification of justification of its application on a particular set of multivariate observations, a factor analysis has been carried out (for details, see: Kovačić, 1994; Hair, Black, Babin & Anderson, 2010; Everitt, 2010), with the principal component analysis as a factor extraction method, and the application of relevant criteria for selection of optimal number of factors. Factor analysis was implemented in order to determine the total and individual values of communalities, i.e. the share of the variance of each individual indicator that was included and explained by the extracted, common factor solution. Communalities were used as a basis for calculating relative weights,

ω_j, for individual *p* variables in the structure of created composite indicator of economic development level (*IED*), using the following expression:

$$\omega_j = \frac{communality_j}{\sum_{j=1}^{p} communality_j}, \quad for\ j = 1, \ldots, p \tag{1}$$

The construction of composite indicator – *Index of Economic Development* (*IED*) was carried out using the method of linear additive aggregation of weighted normalized values of individual economic indicators (OECD, 2008):

$$IED_i = \sum_{j=1}^{p} \omega_j * X'_{ij}, \quad for\ i = 1,\ldots,n\ \ and\ \ j = 1,\ldots,p \tag{2}$$

where *n* and *p* denote the total number of observation units (i.e. objects) and number of variables (i.e. individual indicators), X'_{ij} is a normalized value of j^{th} variable for i^{th} observation unit, ω_j is a weight of j^{th} variable within common factor, and IED_i is a value of composite indicator for i^{th} observation unit.

By ranking the *IED* values and comparing them with the national average (i.e. average *IED* value), the classification of districts in Serbia, according to the degree of economic development, was made. In order to verify the validity of created *IED* classification and provide a statistically based evaluation of quality of composite *IED* indicator, a multivariate classification of districts was performed using the non-hierarchical clustering procedure (for details, see: Kovačić, 1994; Hair et al., 2010), based on the application of *k-means method* on the values of the initial economic indicators. The analysis of possible differences in the composition of clusters identified by these two classification alternatives was carried out using the silhouette coefficient, as a comprehensive indicator of internal homogeneity and external heterogeneity of the formed clusters (Rousseeuw, 1987; Tan, Steinbach & Kumar, 2006):

$$S = \frac{\sum_{i}^{n} s_i}{n}\ for\ s_i = \frac{\left(b_i - a_i\right)}{max\left(a_i, b_i\right)} \tag{3}$$

Within the presented expressions (3), used symbols denote:

s_i – value of the silhouette coefficient for i^{th} object (for i =1,2,…,n);

a_i – average distance of the i^{th} object to all other objects in its cluster;

b_i – minimal average distance of i^{th} object to all objects belonging to other clusters, individually; and

S – value of the overall silhouette coefficient for particular classification.

In addition, the squared Euclidean distance (*Expression 4*) was used as an appropriate distance measure for calculating the values of silhouette coefficient for both classification alternatives:

$$d_{ij}^2 = \sum_{j=1}^{p} \left(x_{ij} - x_{kj} \right)^2, \ \textit{for} \ i = 1,...,n \tag{4}$$

Symbols x_{ij} and x_{kj} denote value of i^{th} observation unit (x_i) for variable X_j, and value of k^{th} observation unit (x_k) for variable X_j, respectively.

This is followed by deeper qualitative analysis and interpretation of identified clusters of districts, from the aspect of original variables' values used in the *IED* construction, along with the corresponding cartographic and multivariate graphical representations (e.g. Chernoff faces). In addition, special attention is devoted to ensuring the fulfillment of the statistical assumptions on which the valid application of presented multivariate analytical procedures is based (for details, see: Tobachnick & Fidell, 2006; Johnson & Wichern, 2007; Hair et al., 2010). Using the appropriate graphical methods and statistical tests, within data preparation phase, the following activities were carried out:

- Univariate distribution normality testing using *Anderson-Darling*, and *Shapiro-Wilk test*, along with univariate outlier analysis using *box plots*;

- Transformation of the original values of variables for which the deviation from normal distribution was identified, using *Box-Cox procedure*, defined as (Osborne, 2010):

$$T(X_{ij}) = \frac{X_{ij}^{\lambda_j} - 1}{\lambda_j}, \ \textit{for} \ \lambda_j \neq 0; \ i = 1,..., n; \ j = 1,..., p \tag{5}$$

where $T(X_{ij})$ denotes the transformed value of i^{th} observation unit for j^{th} variable, X_{ij}, the original value of i^{th} unit for j^{th} variable, and λ_j represents the optimal value of transformation parameter for j^{th} variable;

- Multivariate distribution normality testing using *Mardia's,* and *Henze-Zikler's* test (Ramzan, Zahid & Ramzan, 2013; Korkmaz, Goksuluk & Zararsiz, 2014), multivariate outlier analysis based on the *Mahalanobis distance measure* (Varmuza & Filzmoser, 2009), and examination of multicolinearity and linearity between variables, based on the application of simple linear correlation analysis;

- The normalization of indicators' values by the *min-max transformation technique* (OECD, 2008; Perišić & Wagner, 2015), with the correction of the range of values from 1 to 10, using the following expressions:

$$positive\ coding \rightarrow X_{ij}^{'} = 9^{*}\frac{X_{ij} - X_{j}^{min}}{X_{j}^{max} - X_{j}^{min}} + 1, \quad for\ i = 1,...,n;\ j = 1,...,p, \tag{6}$$

$$inverse\ coding \rightarrow X_{ij}^{'} = -9^{*}\frac{X_{ij} - X_{j}^{min}}{X_{j}^{max} - X_{j}^{min}} + 10, \quad for\ i = 1,...,n;\ j = 1,...,p, \tag{7}$$

where $X_{ij}^{'}$ represents normalized value of j^{th} variable for i^{th} observation unit, X_{ij} is corresponding original value, while, X_j^{min} and X_j^{max} denote minimal, i.e. maximal value of j^{th} variable, respectively. Data analysis and all the necessary statistical calculations were carried out using the following statistical software packages: *IBM SPSS Statistics*, version 20, *MacroSAS for multivariate normality*, and *EduStat* 4.05.

3.1. Variables, sources of data, and temporal-spatial scope

Using the current territorial organization, defined by the nomenclature of statistical territorial units (NUTS) in the Republic of Serbia (RS), secondary data for selected five indicators of economic development (*Table 2*) were collected for 24 administrative districts and Belgrade area (NUTS-3 level). Data were obtained from electronic databases of complex publications of the Statistical Office of the Republic of Serbia (SORS, 2014), and *Report on Small and Medium Enterprises and Entrepreneurship* (ME & NARD, 2014). All collected data refer to the year 2013. In addition, since 1999 SORS provides no information for districts within the Autonomous Province of Kosovo and Metohija, so they are not included in the research.

Table 2: List of selected indicators of economic development of districts

Symbols	*Variables*	*Units of measurement*
X_1	Number of SMEs per 1000 inhabitants	number of SMEs
X_2	Gross value added (GVA) per capita	in 000 RSD
X_3	Employment rate	in %
X_4	Unemployment per 1000 inhabitants	number of unemployed
X_5	Average wage per employee	in RSD
Values of variable X_1 are calculated as the ratio of total number of SMEs (ME & NARD, 2014) and total number of inhabitants in particular district (SORS, 2014), multiplied by 1000;		
Values of variable X_2 are calculated as the ratio of total gross value added for particular district (ME & NARD, 2014) and total number of inhabitants in that district (SORS, 2014);		
Employment rate values are calculated as percentage participation of the number of employed persons in the working-age population of particular district (SORS, 2014); and		
Values of variables X_4 and X_5 were taken from SORS (2014).		

Source: Authors' representations

To obtain the values of selected economic indicators, the authors used not absolute numbers (i.e. number of SMEs, employed/unemployed persons, and total amount of GVA) but their values per capita, per 1000 inhabitants, or percentage participation. This was done in order to neutralize and/or mitigate the impact of total demographic mass of individual districts on the composite indicator values, and thus the resulting classification of districts.

Why were these five specific indicators of economic development selected? The number of indicators that will constitute the composite index is the trade-off between the risk of overlapping information and risk of information loss (Mazziotta & Pareto, 2013). It is also determined by the ratio of number of observations and number of variables (minimum acceptable ratio is 5:1), important for the implementation of factor analysis (Hair et al., 2010). Starting from the fact that SMEs sector represents the most efficient segment of the Serbian economy (ME & NARD, 2014; Jakopin, 2015), and given its positive contribution to employment, gross value added and total turnover, it is clear that indicator of density of SMEs territorial distribution (i.e. number of SMEs per 1000 inhabitants), unequivocally provides useful information on degree of economic development of districts. The rest of the selected variables are the most commonly used regional economic development indicators in the "official" studies analyzing regional development of territories within Serbia (Republic Development Bureau (RDB), 2009; NARD, 2013; Province of Vojvodina Secretariat for Interregional Cooperation and Local Government, 2014; Jakopin, 2014), from the univariate point of view. In addition, regardless of the included dimensions of regional development, the common characteristic of the multivariate studies presented in *Table 1* is that they incorporate at least one or a group of indicators of economic development in the analysis of regional inequalities. Scientific papers focused exclusively on the economic character of regional disparities are relatively rare. In addition to that the lack of a consensus regarding the used economic indicators has caused a pronounced variation in their selection. Accordingly, it is important to emphasize that for the purposes of this research, a specific combination of economic indicators has been created by selecting those indicators that are most often used as indicators of the economic dimension of regional development in the presented studies (*Table 1*). More precisely, depending on the availability of data, the subjective choice of authors and the objectives of the analysis, in more than 55% of analyzed papers, the employment rate (or employment per 1000 inhabitants), the unemployment rate (or unemployment per 1000 inhabitants) or GDP (or GVA) per capita were used as indicators of the economic dimension of regional development. The average wage per employee, as an economic indicator, is present in approximately 40% of the studies in *Table 1*. Approximately 50% of presented papers simultaneously use employment rates and GDP (or GVA) per capita. One third of the studies include indicators of employment and unemployment (e.g. Cziraky et al.,

2003; del Campo et al., 2008; Kurnoga-Živadinović, 2007; Melecky, 2014; Pintilescu, 2011; Rašić-Bakarić, 2006; Lepojević et al., 2015), while quarter of them simultaneously uses these two indicators in combination with GDP (or GVA) per capita (e.g. del Campo et al., 2008; Kurnoga-Živadinović, 2007; Melecky, 2014; Pintilescu, 2011; Rašić-Bakarić, 2006; etc.). Although the simultaneous use of employment and unemployment indicators may initially be debatable and questionable, their incorporation into the structure of composite indicator is fully justified. Namely, due to changes in the manner of recording and treatment of unemployment (SORS, 2014)[3], employment and unemployment are not exactly the two sides of the same coin, since unemployment is not determined as a simple difference between the total working age population and the number of employees. In addition, relatively low correlation coefficient value (r = 0.475) between these indicators (*Table 5*) and the absence of multicolinearity problem, which will be explained in detail in the following section, clearly confirm this.

4. Empirical analysis and results

4.1. Data preparation and preprocessing

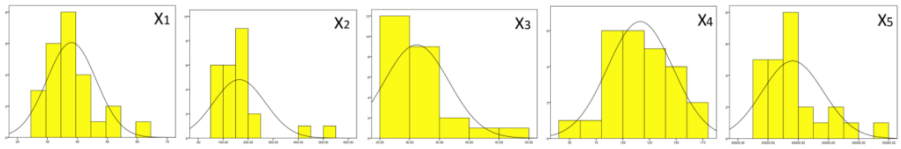
Although certain methods of multivariate analysis are less sensitive to non-fulfillment of certain statistical assumptions, before the application of selected multivariate methods, an examination of whether, and to what extent, the available sample of univariate and multivariate observations meets the defined statistical requirements is performed. Accordingly, normality testing of univariate and multivariate distribution of selected indicators, as a basic prerequisite of a valid implementation of factor analysis, is a necessary step in understanding the nature of the used variables. Presented graphical displays (*Figure 2*), indicate that in the case of most of the variables, a certain (smaller or larger) deviation of the distribution shape from the normal distribution is present, except in the case of variable X_4, for which histogram of frequencies follows and reflects the contour of a normal curve. The remaining four variables are characterized by a positively skewed distribution, and kurtosis smaller than height characteristic for normal distribution.

The results of univariate normality testing (*Table 3*, 2nd and 3rd column), at the level of significance α=0.05, confirm previously stated, visually determined conclusions. More precisely, in contrast to the variable X_4 for which there was not enough evidence to reject the null hypothesis, for variables X_2, X_3 and X_5, the results of *Anderson-Darling* and *Shapiro-Wilk* normality tests suggest the

[3] "Unemployed are considered as persons aged 15 and over who did not work on the reference week, but were actively looking for a job and were ready to start working if they were offered a job within two weeks in relation to the "critical moment of recording." (SORS, 2014: 325)

rejection of null hypothesis, which states that the observed sample of univariate observations is normally distributed, since the corresponding *p*-values are lower than predefined level of test significance *α*.

Figure 2: *Histograms of frequencies with normal curve for original variables*



Source: Authors' representations

Table 3: *The results of statistical tests for univariate normality*

Variables	Normality testing – original variables				Normality testing after transformation			
	Anderson-Darling normality test		*Shapiro-Wilk normality test*		*Anderson-Darling normality test*		*Shapiro-Wilk normality test*	
	statistic	*p-val.*	*statistic*	*p-val.*	*statistic*	*p-val.*	*statistic*	*p-val.*
X_1	0.560	0.148	0.928	0.079	0.128	0.985	0.986	0.974
X_2	2.000	0.000	0.753	0.000	0.426	0.314	0.957	0.359
X_3	1.031	0.010	0.847	0.002	0.257	0.723	0.975	0.761
X_4	0.240	0.776	0.978	0.841	-	-	-	-
X_5	1.528	0.000	0.850	0.002	0.069	0.999	0.993	1.000

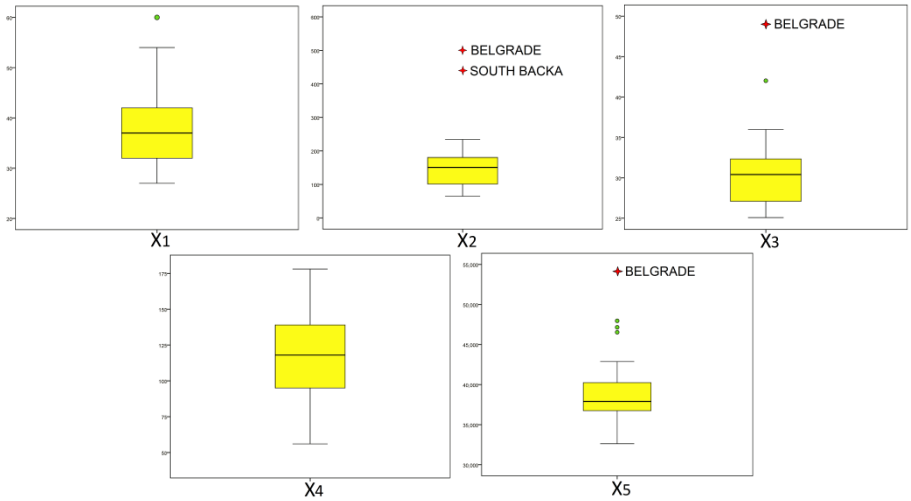
Source: Authors' calculations (*IBM SPSS Statistics 20.0 and EduStat 4.05*)

Based on the graphical representations of univariate distributions (isolated intervals of values at the right tail of distributions), as well as the results of conducted statistical tests, it can be assumed that the identified deviations from normal distribution are, most probably, a consequence of the presence of outliers. *Box-plots* (*Figure 3*) confirm this assumption in the case of variables X_2, X_3, and X_5. For example, the variable X_2, as one with the most pronounced deviation from normal distribution, is characterized by the presence of two outliers (marked with stars). For variable X_4, as expected, the presence of outliers is not recorded, unlike the variable X_1 for which Belgrade area is identified as a potential outlier (marked with circle) and most likely, as a cause of positively asymmetric distribution.

In order to ensure the fulfillment of normality assumptions, and to mitigate the impact of identified non-standard observations, the *Box-Cox's transformation* (*Expression 5*) of the original values of variables X_1, X_2, X_3 and X_5 was performed. The optimal values of the transformation parameter λ for each

variable are –1.267 (for X_1), –0.356 (X_2), –3.052 (X_3), and –3.049 (X_5). In order to verify the effects of the applied transformation method, previously used univariate normality tests were implemented on the transformed values, at the level of significance α=0.05, and obtained results are presented, for the purpose of comparing, in *Table 3* (4^{th} *and* 5^{th} *column*). Presented results confirm the efficiency of the *Box-Cox's* transformation, since, in the case of all four transformed variables, they suggest that there is not enough evidence to reject the corresponding null hypothesis. Testing the assumption on multivariate normality (MVN) was conducted on transformed variables using *Mardia's skewness/kurtosis* and *Henze-Zikler's* MVN test. The results of implemented statistical MVN tests (*Table 4*) suggest that there is not enough evidence to reject the assumption regarding the normality of multivariate distribution, since the obtained *p*-values for all three MVN tests are significantly higher than significance level α = 0.05.

Figure 3: *Box plots for original variables*



Source: Authors' representations

In addition, by comparing the calculated values of Mahalanobis distance for each district (ranging from 1.235 to 10.021), as a measure that approximately follows a chi-square distribution with *p* degrees of freedom, and the value of 97.5 percentile of chi-square distribution ($\chi^2_{5;\ 0,975}$=12.8325), as cut-off value, the presence of multivariate outliers has not been identified in the sample of multivariate observations, since there is no Mahalanobis distance value that is higher than the cut-off value.

Table 4: *The results of statistical tests for multivariate normality*

MVN tests	*test statistic*	*p-value*	*decision*
Mardia's skewness test	24.424	0.910	Ho
Mardia's kurtosis test	-1.737	0.082	Ho
Henze-Zirkler's test	0.786	0.980	Ho

Source: Authors' calculations (*MacroSAS for multivariate normality*)

For the verification of assumption regarding the absence of multicolinearity problem, the analysis of degree and statistical significance of correlation between the observed variables is performed. Based on the analysis of calculated simple linear correlation coefficient values (*r*) for all pairs of variables, presented in the form of correlation matrix type 5x5 (*Table 5*), the following conclusions were made: 1) there are no correlation coefficient values that are lower than 0.30, or higher than 0.80, since they all range from r_{min}=0.439 to r_{max}=0.769, (indicating no weak linearity or multicolinearity problem between the selected indicators), 2) the presence of direct correlation between all pairs of variables is dominant, except in the case of variable X_4, which is negatively correlated with other indicators, and 3) based on the procedure of testing statistical hypothesis regarding the population correlation coefficient (ρ), the statistical significance of linear correlation between all pairs of variables was confirmed.

Table 5: Correlation matrix

Variables	X_1'	X_2'	X_3'	X_4'	X_5'
X_1'	1.000	0.586**	0.439*	0.471*	0.554**
X_2'	0.586**	1.000	0.769**	0.483*	0.764**
X_3'	0.439*	0.769**	1.000	0.475*	0.654**
X_4'	0.471*	0.483*	0.475*	1.000	0.748**
X_5'	0.554**	0.764**	0.654**	0.748**	1.000
Note: (**) and (*) denote statistical significance of correlation at 0.01 and 0.05 level of significance, respectively.					

Source: Authors' calculation (*IBM SPSS Statistics 20.0*)

Finally, having in mind that the selected variables are measured in different units, within the last step of data preparation process, the procedure of normalization of their values is carried out, using the method of *min-max transformation* (*Expression 6, 7*) for converting original and/or transformed values of variables (X_1, X_2, X_3, X_4, X_5) into normalized values (X_1', X_2', X_3', X_4', X_5') ranging from 1 to 10. An inverse coding (*Expression 6*) has been performed only for the variable X_4, since higher unemployment implies a lower level of economic development, and vice versa.

4.2. Construction of Index of Economic Development

Starting from the confirmed linearity, univariate and multivariate normality, as well as the absence of multicolinearity, univariate and multivariate outliers, in order to check the adequacy of selected indicators for the application of factor analysis, an additional examination of strength and statistical significance of linear interrelationship between variables, based on the interpretation of *Kaiser-Meyer-Olkin* (*KMO*) *measure of sampling adequacy*, and the results of *Barttlet's sphericity test*, was carried out. The overall KMO value (0.754) for the set of variables is greater than the specified (Tobachnick & Fidell, 2006), minimally acceptable value (0.50), indicating the existence of a strong linear interdependency between the observed variables. Due to the Bartlett's test statistic value (χ^2=68.244) with 10 degrees of freedom, and corresponding *p*-value (0.000) at significance level 0.05, the null hypothesis is rejected, and the alternative, stating that correlation matrix (at the population level) differs from an identity matrix (i.e. there is a statistically significant linear interrelationship between the variables), is accepted. *KMO* values for individual variables range from 0.699 (moderate interdependence) to 0.881 (very strong interdependence), and thus confirm the statistical significance of correlation of individual variables with other variables.

Table 6: Factors extraction results

Factor	*Initial eigenvalues*			*Sum of communalities after extraction*		
	Total	*% of explained total variance*	*Cumulative %*	*Total*	*% of explained total variance*	*Cumulative %*
1	**3.399**	**67.976**	**67.976**	**3.399**	**67.976**	**67.976**
2	0.640	12.805	80.782	-	-	-
3	0.575	11.503	92.284	-	-	-
4	0.254	5.088	97.373	-	-	-
5	0.131	2.627	100.00	-	-	-

Source: Authors' calculation (*IBM SPSS Statistics 20.0*)

Based on the confirmed sampling adequacy, a valid implementation of the *R-type exploratory factor analysis*, without a priori determination of a number of common factors, was carried out (*Table 6*). In order to identify the best linear combination(s) of variables (i.e. factors) accounting for most of the variance in the data as a whole, following criteria were used: *Kaiser-Guttman rule*, *percentage of variance explained criterion*, and *Cattell's scree test.* After the interpretation of these criteria, a single-factor solution was extracted as an optimal, since of the total of 5 identified latent variables, only the first is characterized by an eigenvalue (3.399) greater than 1.00 (*Kaiser-Guttman rule*), accounting for approximately 68% of the total initial variance of the

observed set of variables (more than 60% is defined threshold by *percentage of variance explained criterion*). This was confirmed by Cattell's scree test.

Communalities for individual variables (*Table 7*, *3rd column*) also confirm the validity of extracted single-factor solution, which explains 67,976% of the total variance of the initial sets of variables. In fact, more than 50% of the variance of each of the observed variables is explained by a common factor (*Table 7, 4th column*), whereby variables X_2' and X_5' are characterized with highest percentages of explained initial variance (78.3% and 83.1%), followed by variable X_3' (66.8%), X_4' (58.8%), and finally X_1' (52.9%).

Table 7: Communalities, factor loadings and weights for each variable

Variables	*Initial Variability*	*Communalities*	*% of total variance explained by common factor*	*Factor loadings*	*Weights*
X1'	1.000	0.529	52.9	0.727	0.1556
X2'	1.000	0.783	78.3	0.885	0.2304
X3'	1.000	0.668	66.8	0.817	0.1965
X4'	1.000	0.588	58.8	0.767	0.1730
X5'	1.000	0.831	83.1	0.912	0.2445
Total	= 5.000	= 3.399	≈ 67.976	/	= 1.000
In fifth column, the unrotated factor loadings are presented, since only one factor was extracted, and therefore, there were no empirical grounds, nor the need for the implementation of the factor rotation method.					

Source: Authors' calculation (*IBM SPSS Statistics 20.0*)

In addition, based on insight into factor loadings of individual variables (ranging from 0.727 to 0.912), as indicators of strength of correlation between each of the variables and extracted factor (*Table 7, 5th column*), it can be concluded that practical significance of the obtained single-factor solution is satisfied, since all variables are characterized by factor loadings which are significantly greater than +0.30, as a minimum satisfactory threshold value. Since each of the five variables is used for measuring and describing particular, specific aspect of a common phenomenon (i.e. economic development level), the extracted factor is named *Index of Economic Development* (*IED*). Finally, on the basis of presented and interpreted values of communalities and factor loadings, determination of weights (ω_j) of initial economic indicators in the structure of a composite indicator *IED* (extracted factor) is performed by calculating the relative share of communalities in the identified eigenvalue of extracted factor (*Table 7, 6th column*), using *Expression* (1). After weighting of the normalized values of individual indicators, the *IED* values are calculated, using the *Expression* (2). More specifically, the structure of the formed composite indicator (*IED*) can be presented as follows:

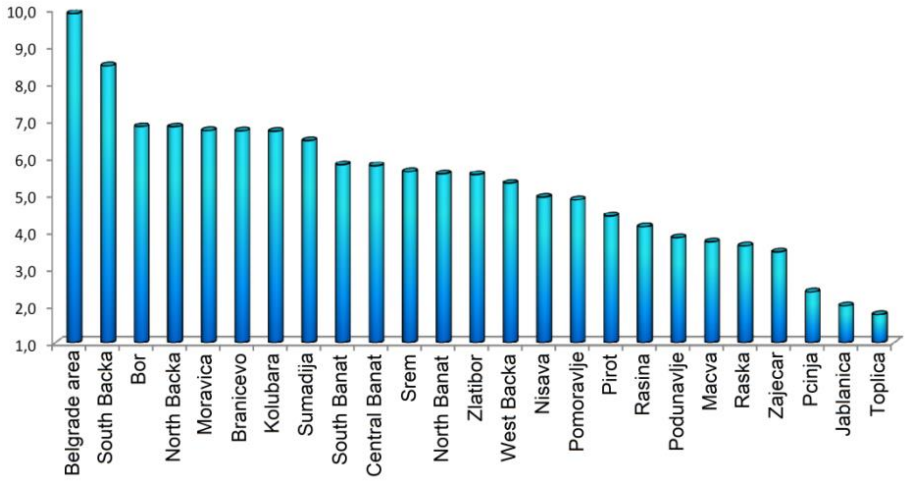
$$\boldsymbol{IED_i} = 0.16 * \boldsymbol{X_{i1}'} + 0.23 * \boldsymbol{X_{i2}'} + 0.20 * \boldsymbol{X_{i3}'} + 0.17 * \boldsymbol{X_{i4}'} + 0.24 * \boldsymbol{X_{i5}'},$$

where X_{ij}' are normalized values of individual indicators (for i = 1, 2,…, 25, and j =1, 2,…, 5). The *IED* values range from 1 to 10, where districts with a values closer to 10 are characterized by a higher degree of economic development, and vice versa. The maximum value of the index of economic development of districts in Serbia is 9.8851 (Belgrade area), and minimum value is 1.7513 (Toplica district), which is within defined range of values.

4.3. Classification of districts

Ranking of districts according to their *IED* values (*Figure 4*) clearly indicates the presence of significant regional disparities, in terms of the level of economic development.

Figure 4: *Ranking of districts in Serbia based on their IED values*

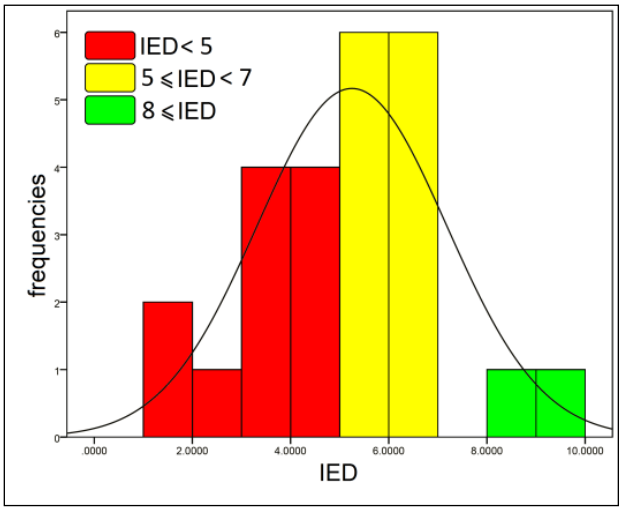


Source: Authors' representation

Belgrade area (9.8851) and South Backa district (8.4732) are characterized by the highest *IED* values (*IED* ≥ 8), while lowest values are inherent to the districts in the Region of Eastern and Southern Serbia, i.e.: Pcinja (2.3642), Jablanica (1.9876), and Toplica district (1.7513). With the exception of Bor and Braničevo, all remaining districts whitin this region are among the last ranked districts, occupying 7 of the last 11 positions, with values *IED* < 5. On the other hand, the greatest number of observed districts (12 out of 25, approx. 48%) is characterized by *IED* values in the range 7 > *IED* ≥ 5, which comprises all seven districts in Vojvodina region, and by two of districts from

the Region of Sumadija and Western Serbia and Region of Eastern and Southern Serbia. This modal interval also includes medial (5.54) and average *IED* value (5.25). Described intervals of *IED* values are shown in *Figure 5*.

Figure 5: *Histogram of frequencies for IED interval values*



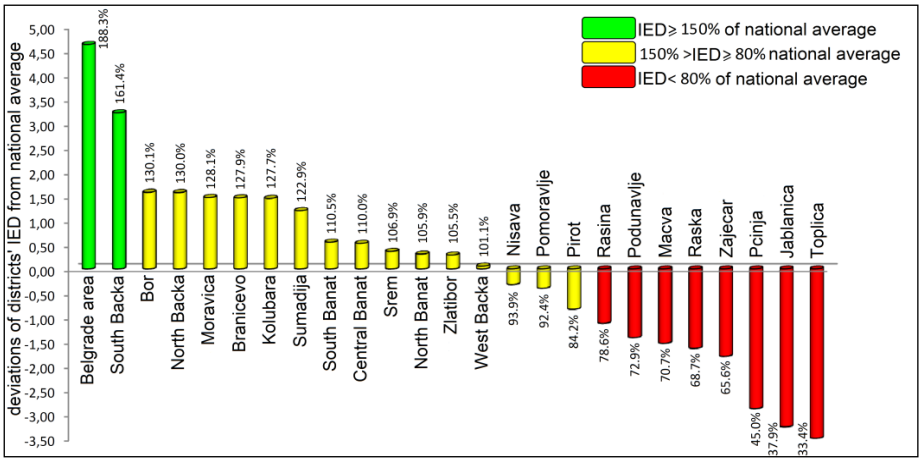
Source: Authors' representation

Starting from the presented ranking of districts and official methodology (Government of the Republic of Serbia, 2013) for classification of regions and local government units according to their development level, a classification of districts based on *IED* values is performed. The average *IED* value (25.5) was determined, and relative to it, as the national average, a comparison of *IED* values for all the observed districts was done, in order to define their classification according to economic development. Deviations of *IED* values for individual districts, with the respective percentages of achieved economic development in relation to the national average, are presented in *Figure 6*.

Belgrade area and South Backa district are territories characterized by the *high level of economic development* (Group 1), since their *IED* values are over 150% of the national average. On the other hand, eight districts (mainly within the Region of Eastern and Southern Serbia) recorded *IED* values under 80% of the national average, due to which the same can be classified within the group of districts with *low level of economic development* (Group 3). The largest number of districts (precisely 15, i.e. approx. 60% of the total number of observed districts) is characterized by the *IED* values moderately above or slightly below the determined average value, i.e., between 80% and 130% (<150%) of the national average, which is why they can be classified within

the group of district with *moderate (average) level of achieved economic development* (Group 2).

Figure 6: *Individual deviations of districts from the average IED value*



Source: Authors' representation

For evaluation of identified structure of the three groups of districts, a *non-hierarchical cluster analysis* was implemented on normalized values of selected economic indicators, using *k-means method*. More precisely, the obtained results are used for statistical evaluation of quality and validity of constructed composite indicator (*IED*), and therefore, the justifiability of its application for classification of districts in Serbia according to the achieved degree of economic development. By comparing the classification solutions based on *IED* values and non-hierarchical clustering, a minimum deviation has been identified. In fact, identical allocation was obtained for all districts regarding their membership to certain groups, except in the case of Pirot district, which was, according to the results of applied *k-means* method, excluded from the group of districts with moderate economic development and assigned to the group of low economic development. The final decision regarding the position of Pirot district was made based on the comparison of silhouette coefficient values calculated for both classifications (*Table 8*).

Table 8: Silhouette coefficient values for defined classification alternatives

Classification alternatives	*Overall Silhouette coefficient*	*Silhouette coefficient for Pirot district*
Solution based on IED values	0.50732	+ 0,1203
Nonhierarchical clustering solution	0.49545	– 0,1203

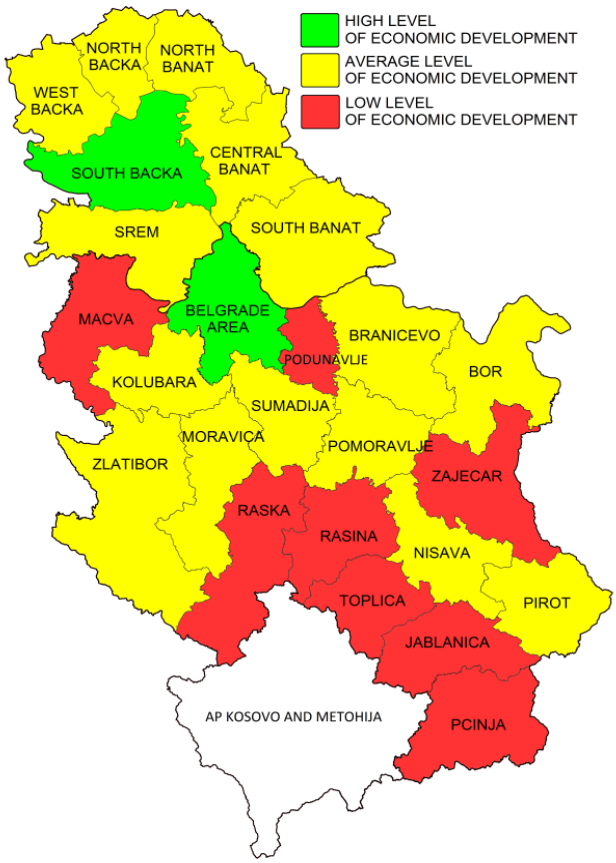
Source: Authors' calculations

The higher value of the overall silhouette coefficient, and the negative value of individual silhouette coefficient for Pirot district in the non-hierarchical solution, suggest the retention of the classification alternative based on *IED* values, as more favorable ("more optimal").

4.4. Interpretation of classification of districts

The formed classification of districts in Serbia, based on their economic development, is presented visually, through a cartographic representation (*Figure 7*), and Chernoff faces (*Figure 8*), along with their interpretation.

Figure 7: Cartographic representation of districts in Serbia by the level of economic development



Source: Authors' representation

Cluster 3 (*low level of economic development*) consists of 8 districts, i.e. ≈30% of total number of observed districts. Approximately $^{3}/_{5}$ (or ≈ 60%) of the districts within this group belong to the Region of Eastern and Southern Serbia. The average *IED* value in this cluster is 3.10, which is at the level of ≈60% of the national average (5.25). The average and even maximum values of all individually observed economic indicators are here significantly below the corresponding national average (*Table 9*). More precisely, twice lower value of GVA per capita, ≈10% lower amount of average wage, 4.5% lower employment rate, 19% more unemployed, and 13% smaller number of SMEs per 1000 inhabitants, in average, compared to the corresponding national averages, unambiguously point to the general economic vulnerability and catastrophic consequences of mismatches of natural, production and material resources, as constraints from earlier periods, amplified by the wave of economic devastation that hit these districts during the transition process. By comparing with average values in Cluster 1, the presence of extremely uneven regional development and alarming regional economic disparities between the districts has been confirmed. The ratio of degree of development of these two clusters varies from 1.43:1, regarding the amount of average wages, to drastic 5.57:1, in terms of GVA per capita, in favor of Cluster 1. The seriousness of the observed situation from the aspect of economic development of districts in Cluster 3 is perhaps best illustrated by the "sad" expressions of Chernoff faces on the verge of tears, with lowered eyebrows, narrow and small nose, and small and downward-facing mouth, which is illustrated later, in *Figure 8*.

Table 9: Average values of economic indicators by groups of districts

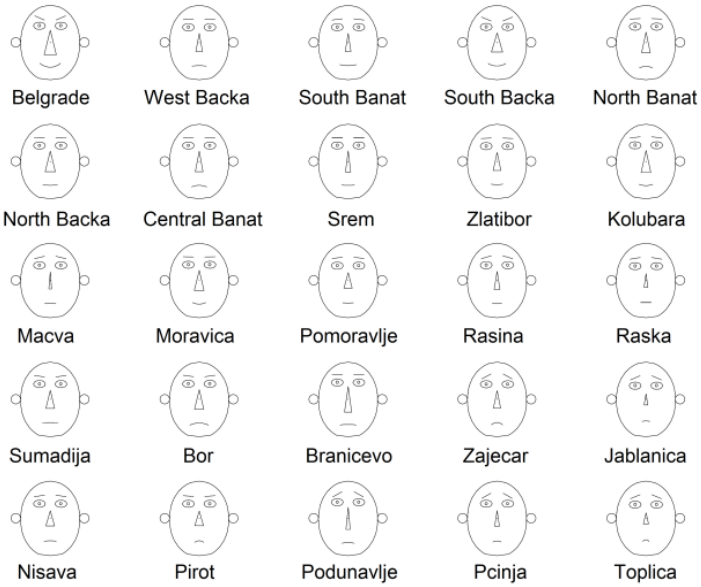
Variables	National Average	Average values		
		Group 1	Group 2	Group 3
Number of SMEs per 1000 inhabitants	38	57	38	33
Gross value added per capita	165	469.7	167.4	84.3
Employment rate	31.2	45.5	31.6	26.8
Unemployment per 1000 inhabitants	116	87	108	138
Average wage per employee	39083	50628	39548	35325

Source: Authors' calculation (*IBM SPSS Statistics 20.0*)

Cluster 2 (*average level of economic development*) accounts for 60% of the total number of observed districts. Their average *IED* value is 5.87, which is about 12% above the national average. Unlike the previous group, characterized by the dominant presence of the districts in the Region of Eastern and Southern Serbia, within this group, the presence of districts in the Region of Vojvodina is dominant, since they account for 40% of its structure. Almost identical average values at the group level with corresponding republic averages for individual economic indicators (*Table 9*) additionally confirm the

quality of the performed classification from the aspect of the number and composition of the clusters. Compared to the individual indicators' average values in Group 3, double GVA per capita, 15% more SMEs, and approximately 22% fewer unemployed per 1000 inhabitants, with nearly 5% higher employment rate and 12% higher amount of average wages, clearly point to the more favorable economic position of the districts in Cluster 2. On the other hand, expected, although less pronounced than in the case of Group 3, lagging behind in economic development in relation to Group 1 varies from 1:1.24 (measured by the number of unemployed persons) to 2.81:1, from the aspect of GVA per capita. The "indifferent" expressions of Chernoff faces representing districts in Group 2, with horizontal position of the eyebrows, mostly normal (average) width and length of nose, and flat or slightly curved, upward or downward, mouth, confirm the above stated facts.

Figure 8: Chernoff faces of districts by the level of economic development



Source: Authors' representation

Group 1 (high level of economic development) includes only South Backa district and Belgrade area. This classification outcome is partially expected, bearing in mind that these same districts were, in the pre-processing phase, identified as outliers in almost all variables (*Figure 3*). Their allocation within Group 1 is conditioned by the fact that in their territory are the cities of Belgrade and Novi Sad, which represent the leading economic,

administrative, cultural and university centers in the RS and Vojvodina region, with exceptionally favorable geographical position and infrastructure preconditions (road and railroad nodes, Danube corridor, airport) for intensive economic development and attraction of domestic and foreign investments. The dominant position of these two areas in terms of economic development is confirmed by their average *IED* value of 9.18, which is ≈75% above the national average. Approximately 30% higher average wages and lower number of unemployed, 14.3% higher employment rate and 50% more SMEs per 1000 inhabitants explain almost three times higher amount of GVA per capita in these two areas, compared to the corresponding national averages per individual economic indicators. The "prepotent" expressions of their Chernoff faces, high-edged eyebrows, wide / big nose and cynical smile, best explain their position from the aspect of achieved economic development in relation to the rest of the RS territory.

Although there are numerous studies dealing with the analysis of regional disparities within the territory of Serbia (e.g. Republic Development Bureau, 2009; NARD, 2013; Province of Vojvodina Secretariat for Interregional Cooperation and Local Government, 2014; Molnar, 2013; Jakopin, 2014), regardless of similarity of the objectives, the results are not comparable with ours, due to theoretical and methodological differences and temporal scope of data. More precisely, these studies mainly provide univariate analysis based on the separated classifications of different territorial units (regions, districts, or municipalities) for one particular or several individual regional economic development indicators. In other words, each of the used economic indicators has its own specific classification of territorial units, and classifications based on different indicators mainly differ. The results of the multivariate classification of the districts presented in this paper, which is based on a simultaneous analysis of the five selected economic indicators, are not comparable with the corresponding five univariate, individual classifications. The methodological approach applied here, enables sublimation, with minimal loss, of information contained in separated, individual classifications, and their presentation in a single common classification, providing more contextualized, less fragmented insight into the phenomenon at hand.

Nevertheless, the obtained results confirm the well-known fact that the Belgrade area and the South Bačka district represent the engine of economic development in Serbia, in contrast to the districts in the eastern and southern parts of the territory. However, the proposed methodological framework based on the construction and calculation of values of the innovative composite indicator provides quantification of the achieved level of economic development of each Individual territorial unit, allowing their comparison and precise view of their position, in terms of the values of all five used economic indicators, in relation to the leading districts, those least developed, or any other district within the same or some other cluster. These are precisely the

key advantages, in relation to the multivariate classifications provided by the application of cluster analysis in the research carried out by Popović & Maletić (2008), Lepojević et al. (2015), and Janković-Milić, Marković & Igić (2013). In addition to this comparison of methodological approaches, due to the differences in the spatial-temporal coverage of data, as well as the difference in terms of used indicators and dimensions of regional development, a concrete comparison of empirical results would not be valid.

In comparison to Igić (2014), which uses data for the year 2011 and covers the same territorial units (districts) in Serbia, our research provides a more rigorous multivariate statistical analysis used for construction of a proposed composite indicator, and fulfillment of statistical assumptions on which the valid application of presented MVA methods is based. Methodological differences are reflected in the application of univariate and multivariate outlier analysis, testing the normality of univariate and multivariate distribution of indicators, and the use of non-hierarchical clustering procedure and corresponding optimality criteria for verification of validity of the performed classification of districts based on the composite indicator values. In addition, the use of indicators of demographic, technological and environmental dimensions of regional development, in addition to the economic one, results, as a rule, in different weights in the composite index structure preventing direct comparison of the obtained classification results.

5. Conclusions

In this paper, an alternative multivariate statistical approach for the classification of districts in Serbia according to economic development is proposed. It is based on simultaneous observation and measurement of values of five representative economic indicators, and their aggregation in the form of composite indicator *IED*. Developed as intensive and statistically justified use of selected MVA methods, the proposed approach in the analysis of regional economic disparities is characterized by the following practical and methodological advantages in relation to the majority of previously implemented research of similar character.

- Multivariate compression of five representative indicators to the level of univariate composite indicator provides a more practical yet equally informative classification of the observed territories according to the degree of economic development. In comparison to the approach based on a separate interpretation of values of individual economic indicators, the classification obtained by this study is more illustrative, comprehensible and informative when used to formulate and efficiently implement concrete measures in the domain of regional development strategy.

- When it comes to opportunities for future research, the classification of districts according to the values of the proposed composite indicator (*IED*) can serve as a suitable basis for further and deeper analysis of the situation and tendencies within their territories in relation to other important dimensions of regional development, and formulation of adequate conclusions regarding their interdependence and causality.

- The proposed multivariate statistical approach provides a tool for the estimation of the development level of each territorial unit that allows comparison of the evolution of development over time, identification of disparities and growth patterns and, therefore, can be used as a policy tool for planning and resource allocation.

- The key advantage of the proposed methodological approach is the thorough verification of fulfillment of statistical assumptions on which the valid application of presented MVA methods is based, as an indispensable and initial, though frequently neglected and omitted, step in securing the scientific foundation of the results and drawn conclusions.

- The statistical validity of the *IED* (i.e. weight coefficients of individual indicators in its composition), and thus derived classification of districts, was confirmed by the results of non-hierarchical clustering analysis.

- On the other hand, the obtained classification, complemented by detailed interpretation, cartographic representation and Chernoff faces, as a specific, rarely used, but very informative multivariate graphical representation, confirms pronounced regional economic inequalities among observed territories, in Serbia in 2013. By analyzing the *IED* values, three groups of districts have been identified according to the level of economic development, and they are as follows: group of districts with high (group 1), average (group 2), and low (group 3) economic development. The ratios of the average IED values of the identified groups vary from 1.56:1 (group 1 vs. group 2) to 1.89:1 (group 2 vs. group 3) and 2.96:1 (group 1 vs. group 3). The ratio of the best and worst-ranked districts (Belgrade and Toplica) of 5.64:1, best describes the degree and seriousness of the present disparities among the districts in Serbia. Generally, the performed classification confirms the well-known division between the developed northern and undeveloped southern part of Serbia, since all districts of the Vojvodina region (except for South Bačka) are positioned within the group of average economic development, while the largest part of the Region of Eastern and Southern Serbia is characterized by low economic development.

To sum up, the applied multivariate approach enables a clear, objective, transparent and statistically valid analysis of the degree of economic development of territorial units in Serbia, thus providing a reliable basis for redefining and efficient implementation of appropriate measures within the

framework of the regional development strategy aimed at mitigating the evidently present and expressed regional economic asymmetries.

References

Aumayr, C. (2006). European region types: A cluster analysis of European NUTS 3 regions. Paper presented at the *46th Congress of the European Regional Science Association* in Volos, August 30th–September 3rd, 2006. Retrieved from: *http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa06/papers/836.pdf*

Avram, M. & Postoiu, C. (2016). Territorial patterns of development in the European Union. *Theoretical and Applied Economics*, XXIII(1): 77-88.

Brauksa, I. (2013). Use of cluster analysis in exploring economic indicator differences among regions: The case of Latvia. *Journal of Economics, Business and Management,* 1(1): 42-45.

Cziraky, D., Puljiz, J., Jurlin, K., Maleković, S. & Polić, M. (2003). An econometric model for development level assessment with an application to municipality development classification. In: *Proceedings of the 5th international conference Enterprise in Transition,* University of Split, Split, pp. 173-184. Retrieved from: *http://www.policy.hu/cziraky/SplitRD.pdf*

Cziraky, D., Sambt, J., Rovan, J. & Puljiz, J. (2006). Regional development assessment: A structural equation approach. *European Journal of Operational Research*, 174(1): 427-442.

del Campo, C., Monteiro, C.M.F. & Soares, J.O. (2008). The European regional policy and the socio-economic diversity of European regions: a multivariate analysis. *European Journal of Operational Research*,187(2): 600-612.

Everitt, B.S. (2010). *Multivariable modeling and multivariate analysis for the behavioral science*. Boca Raton: CRC Press (Taylor & Francis Group).

Goletsis, Y. & Chletsos, M. (2011). Measurement of development and regional disparities in Greek periphery: A multivariate approach. *Socio-Economic Planning Sciences*, 45(4): 174-183.

Government of the Republic of Croatia (2010). *Regulation on the Development Index*. Zagreb: Narodne novine, N°: NN 63/10. Retrieved from: *http://narodne-novine.nn.hr/clanci/sluzbeni/full/2010_05_63_1999.html*

Government of the Republic of Serbia (2007). *Regional Development Strategy of Serbia, 2007–2012*. Retrieved from: *http://www.gs.gov.rs/lat/strategije-vs.html*

Government of the Republic of Serbia (2013). *Regulation on the establishment of a single list of development of regions and local government units in Serbia for 2013*. Službeni glasnik Republike Srbije, N°: 62/2013. Retrieved from: *http://narr.gov.rs/index.php/Dokumenta/Pravni-kutak/Uredbe*

Hair, J.F.Jr., Black, W., Babin, B. & Anderson, R. (2010). *Multivariate data analysis, 7th edition*. USA: Pearson Prentice Hall.

Igić, V. (2014). *The multivariate analysis application in creating the composite index of district development in the Republic of Serbia*. [PhD Dissertation, Faculty of Economics, University of Niš, Serbia]. Retrieved from: *https://phaidrani.ni.ac.rs/detail_object/o:952?tab=0#mda*

Jakopin, E. (2014). Regional inequalities and transition: The case of Serbia. *SAE Journal of Business, Economics and Management* [*Ekonomika preduzeća*], LXII(1-2): 117-133.

Jakopin, E. (2015). Regional drivers of economic growth. *SAE Journal of Business, Economics and Management* [*Ekonomika preduzeća*], LXIII(1-2): 99-114.

Janković-Milić, V., Marković, I. & Igić, V. (2013). Cluster analysis of the districts in Serbia according to social development indicators. In: *Proceedings of the 2nd International scientific conference – Post Crisis Recovery*, Belgrade Banking Academy & Institute of Economic Sciences, Belgrade, Serbia: 676-693.

Johnson, R.A. & Wichern, D.W. (2007). *Applied Multivariate Statstical Analysis, 6th edition*. NY: Pearson Education Inc.

Korkmaz, S., Goksuluk, D. & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2):151-162.

Kovačić, Z. (1994). Multivariate analysis [*Multivarijaciona analiza*]. Belgrade, Serbia: Faculty of Economics University of Belgrade.

Kurnoga-Živadinović, N. (2007). Multivariate classification of Croatian Counties [Multivarijatna klasifikacija županija Hrvatske]. *Proceedings of the Faculty of Economics and Business in Zagreb*, 5(1): 1-15.

Kvičalova, J., Mazalova, V. & Široky, J. (2014). Identification of the differences between the regions of the Czech Republic based on the economic characteristics. *Procedia Economics and Finance,* 12: 343–352.

Lepojević, V., Bošković, G. & Janković-Milić, V. (2015). Cluster analysis of the indicators of the development of municipalities in the Region of Southern and Eastern Serbia. In: *Proceedings of the XX International Scientific Conference-Regional development and demographic trends of the countries of South-Eastern Europe,* Faculty of Economics University of Niš, pp. 151-162.

Maletić, R. & Bucalo-Jelić, D. (2016). Definition of Homogeneous and Narrower Areas of the Republic of Serbia. *Agrieconomica* [*Agroekonomika*], 45(69): 13-24.

Mazziotta, M. & Pareto, A. (2013). Methods for constructing composite indices: one for all or all for one? *Rivista Italiana di Economia, Demografia e Statistica, 67*(2): 67-80.

Melecky, L. (2014). Spatial analysis of NUTS 2 Regions based on competitiveness factors and their regional variability. In: *Proceedings of the 5th Central European Conference in Regional Science* (*CERS 2014*): 581-591. Retrieved from: *http://www3.ekf.tuke.sk/cers/files/zbornik2014/PDF/Melecky.pdf*

Michaelides, P.G., Economakis, G. & Lagos, D. (2006). Clustering Analysis Methodology for employment and Regional Planning in Greece. *MPRA Paper No. 74468*. Retrieved from: *https://mpra.ub.uni-muenchen.de/74468/*

Ministry of Economy & National Agency for Regional Development (ME & NARD), (2014). Report on Small and Medium Enterprises and Entrepreneurship for 2013. Belgrade, Serbia: Birograf. Retrieved from: *http://www.privreda.gov.rs/wp-content/uploads/2015/05/Izvestaj-o-MSP-za-2013.pdf*

Molnar, D. (2013). Facts about Regional Differences in Serbia. *Quarterly monitor of economic trends and policies in Serbia*, 32(jan.-mar.): 66-72.

National Agency for Regional Development (NARD), (2013). *Report on Regional Development in Serbia for 2012*. Belgrade, Serbia: Author.

OECD (2008). *Handbook on constructing composite indicators: methodology and user guide*. Retrieved from: *https://www.oecd.org/std/42495745.pdf*

Osborne, J.W. (2010). Improving your data transformations: applying the Box-Cox transformation. *Practical assessment, research and evaluation*, electronic journal, 15(12): 1-9.

Özaslan, M., Dincer, B. & Özgür, H. (2006). Regional disparities and territorial indicators in Turkey: socio-economic development index (SEDI). In: *Proceedings of the 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean"*, August 30th– September 3rd, 2006, Volos, Greece. Retrieved from: *http://hdl.handle.net/10419/118537*

Perišić, A. (2014). Multivariate Classification of Local and Regional Government Units According to Socio-Economic Development. *Društvena istraživanja: Journal for general social issues*, 23(2): 211-231.

Perišić, A. & Wagner, V. (2015). Development index: analysis of the basic instrument of Croatian regional policy. *Financial theory and practice*, 39(2): 205-236.

Pintilescu, C. (2011). Regional economic disparities in Romania. In: *Recent Researches in Applied Economics - Proceedings of the 3rd World Multiconference on Applied Economics, Business and Development* (*AEBD'11*), Iasi, Romania, July 1-3, 2011: 43-47. Retrieved from: *http://www.wseas.us/e-library/conferences/2011/Iasi/AEBD/AEBD-06.pdf*

Popović, B. & Maletić, R. (2008). Grouping of Municipalities of the Republic of Serbia based on Development of Small and Medium Agrobusinesses. *Economics of Agriculture* [*Ekonomika poljoprivrede*], 55(2): 151-161.

Province of Vojvodina Secretariat for Interregional Cooperation and Local Government, (2014). *AP Vojvodina Development Programme, 2014-2020*. Retrived from: *http://www.region.vojvodina.gov.rs/Program_razvoja_APV*

Ramzan, S., Zahid, F.M. & Ramzan, S. (2013). Evaluating Multivariate Normality: A Graphical Approach. *Middle-East Journal of Scientific Research*, 13(2): 254-263.

Rašić-Bakarić, I. (2006). Methods of multivariate analysis to uncover socio-economic differences among spatial-economic entities. In: *Proceedings of the 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean"*, August 30th-September 3rd, 2006, Volos, Greece. Retrieved from: *http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa06/papers/56.pdf*

Republic Development Bureau (RDB), (2009). *The regional development of Serbia in 2009*. Belgrade, Serbia: Author. Retrieved from: *http://www.regionalnirazvoj.gov.rs/Pages/GetFile.aspx?lid=dict&fid=13*

Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of the computational and applied mathematics*, 20(Nov.): 53-65.

Rovan, J. & Sambt, J. (2003). Socio-economic Differences Among Slovenian Municipalities: A Cluster Analysis Approach. In: Ferligoj, A. & Mrvar, A. (editors) *Developments in Applied Statistics, Metodološki zvezki*, 19: 265-278. Retrieved from: *http://ams.sisplet.org/uploadi/editor/rovan.pdf*

Salzman, J. (2003). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Ottawa, Canada: Center for the Study of Living Standards. Retrieved from: *http://www.csls.ca/events/cea2003/salzman-typol-cea2003.pdf*

Soares, J.O., Marques, M.M.L. & Monteiro, C.M.F. (2003). A multivariate methodology to uncover regional disparities: A contribution to improve European Union and

governmental decisions. *European Journal of Operational Research*, 145(1): 121–135.

Statistical Office of the Republic of Serbia (SORS), (2014). Municipalities and Regions in the Republic of Serbia, 2013. Belgrade, Serbia: Author (SORS). Retrieved from: *http://pod2.stat.gov.rs/ObjavljenePublikacije/G2014/pdf/G20142014.pdf*

Tan, P.-N., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. USA: Pearson Education Limited.

Tobachnick, B.G. & Fidell, L.S. (2006). *Using multivariate statistics, 5th edition*. USA: Allyn & Bacon, Inc.

Varmuza, K. & Filzmoser,P. (2009). *Introduction to multivariate statistical analysis in chemometrics*. Boca Raton: CRC Press (Taylor & Francis Group).

Vukmirović, J. (2013). Regionalni razvoj kao preduslov za izlazak iz krize. *Makroekonomske analize i trendovi*, (219): 39-43. Retrieved from: *http://www.ecinst.org.rs/sites/default/files/mat-kratki/temabroja219.pdf*

Vydrova, H.V. & Novotna, Z. (2012). Evaluation of disparities in living standards of regions of the Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, XL(4): 407-414.

Xiaoxin, Q. (2012). Multivariate statistical methods: the level of our comprehensive analysis of all regions. *Physics Procedia*, 24(part C): 1663–1668.