sciendo

South East European Journal of Economics and Business
Volume 18 (1) 2023, 15-26
DOI: 10.2478/jeb-2023-0002

REGIONALIZATION BASED ON SOCIO-ECONOMIC DEVELOPMENT INDICATORS IN BOSNIA AND HERZEGOVINA – EXAMPLE OF FEDERATION OF BOSNIA AND HERZEGOVINA

Adela Delalić, Mirza Šikalo, Selma Numanović

Abstract

This study classifies local self-government units (local communities) in the Federation of Bosnia and Herzegovina (FBIH) according to socio-economic characteristics by applying the following multivariate methods: principal component analysis (PCA), regression and cluster analysis. The selection of variables was based on literature and adjusted by FBIH specifics, covering the four hypothetical dimensions of regional differentiation: macroeconomic, demographic, infrastructural and socio-cultural indicators. PCA has identified five components: the productivity component, demographic component, component of economic activity potential, spatial component and employment component. Further analysis showed that all identified factors are significant predictors of local communities' development, measured by the development index. The cluster analysis resulted with four clusters in the FBIH with significant differences in development level. Considering that FBIH municipalities are administrative units of local government and that the classification is based on socio-economic dimensions, identified clusters correspond to the NUTS principles.

Keywords: *development index, socio-economic factors, NUTS classification, multivariate analysis.*

JEL classification: *R12, R58, C10*

1. Introduction

The development of Bosnia and Herzegovina (BIH) is limited by major regional differences in the development of local communities. The significance of an even regional development has been pointed out by the European Union (EU) through their financing of regional projects encouraging regionalization with the main idea of an even, sustainable and competitive socio-economic development of the country. When analyzing Bosnia and Herzegovina, it can be concluded that a complex, asymmetric and multiple-tier government structure has led to an inefficient local government structure (Mujakić 2010). Economic, demographic and territorial differences between regions are the main obstacle for a balanced and harmonious development

Adela Delalić, PhD
Associate Professor
School of Economics and Business
University of Sarajevo
Bosnia and Herzegovina
E-mail: adela.delalic@efsa.unsa.ba
ORCID: 0000-0003-2556-0461

Mirza Šikalo, MA (corresponding author)
Senior Teaching Assistant
School of Economics and Business
University of Sarajevo
Humska 438, Sarajevo 71000
Bosnia and Herzegovina
E-mail: mirza.sikalo@efsa.unsa.ba
ORCID: 0000-0001-8709-3745

Selma Numanović, MA
Senior expert associate for business with large clients
Bosna Bank International d.d.
Bosnia and Herzegovina
E-mail: selma_sa94@hotmail.com
ORCID: 0000-0002-1441-4293

at regional and country level. Since major geographical units lack homogeneity, i.e. the geographical proximity does not automatically entail socio-economic proximity, such differences among territorial units of a single country are a factor of significant interest to economists and policy makers. A thorough analysis must consider smaller geographical units with political authority and a wide spectrum of socio-economic indicators (Soares et al. 2003).

1.1. Mechanisms of regionalization

There is not a single system of measuring disparity in the EU. On the contrary, EU countries use different approaches and methods and apply selected relevant indicators and methods. According to Lipshitz and Raveh (2010), research on regional disparities and policies for their reduction do not focus enough on socio-economic differences within peripheral and key regions, highlighting the inequality among regions.

The current regionalization in BIH can be classified into three categories: administrative regionalization regulated by the Constitution of FBIH, economic regionalization founded in the operating area of regional development agencies, and regionalization according to the vision of the European Commission-the NUTS classification (Osmanković et al. 2009). As part of the EU joining process, BIH is obliged to establish a statistical regionalization following the criteria of the Eurostat NUTS classification, which is based on the adoption of spatial units from political and governing systems such as areas of local communities. The aim of NUTS is to provide the framework for the division of the economic territory of EU into territorial units for collecting and publishing standardized statistical data and to develop and harmonize regional statistics suitable for analysis and to direct the political interventions on the regional level (Zarić and Vuković 2010). For example, regions selected to receive special aid from the EU structure funds will not be chosen on an ad hoc basis, but rather match certain regions in the NUTS hierarchy. Most importantly, regionalization provides policy makers with an opportunity for a better supervision of municipality development by using the most relevant predictors.

1.2. The regionalization based on socio-economic factors

Multivariate methods of analysis have been used in multiple studies with an aim of grouping territorial units based on different and socio-economic variables idiosyncratic to a particular area. This study utilized factor analysis to summarize and identify the main socio-economic factors, cluster analysis to classify local communities in FBIH in homogeneous groups and regression analysis to examine the impact of socio-economic factors on local communities' development.

Numerous studies have utilized cluster analysis in research and classification of local communities in relation to their development (Soares et al. 2003 - Portuguese NUTS 2 regions, Rovan and Sambt 2003 – municipalities in Slovenia, Lovrinčević et al. 2005 – Croatian NUTS 2 regions). Similar to this paper, some of the previous studies conducted their research of regional inequalities and classification of local communities based on combination of factor and cluster analysis (Del Campo et al. 2008 - 241 regions in 25 EU countries, Palevičienėa and Dumčiuvienėa 2015 - NUTS 2 regions in EU countries, Rašić-Bakarić 2005 – Croation NUTS 2 regions, Kurnoga-Živadinović 2007 – NUTS 1, 2 and 3 in Croatia).

Considering the selection of variables for the study, justification is also found in the inputs of numerous studies in the region, the EU and the world (Del Campo et al. 2008, Palevičienė and Dumčiuvienė 2015, Bartkowska and Riedl 2012, Cruz-Jesus et al. 2012, Melecky 2012, Kurnoga-Živadinović 2007, Rovan and Sambt 2003, Rašić-Bakarić 2005, Aragon el al. 2003, Gonzalez and Morini 2000, Peschel 1998, Pettersson 2001, Kronthaler 2003, Ru'a Vieytes et al. 2003, Stimson et al. 2001 – socio-economic indicators, Soares et al. 2003 - demographic, economic, health, educational and cultural characteristics, Lovrinčević et al. 2005, Pejanović and Kordej De-Villa 2015 - economic, demographic and social characteristics, Botrić et al. 2006 - development of road structure construction, Botrić et al. 2003 - wages, employment, unemployment, labor mobility, Bloom et al. 2001, Feyrer 2002, Gómez and Hernández de Cos 2006, Hartmann 2010, Balan 2015, Kelley 2001, De Haas 2008 - demographic indicators etc.).

1.3. Objectives and Research Questions

The topicality of the research arising from obligations of BIH, following the Stabilization and Association Agreement is to determine statistical regional units in accordance with the regulated Eurostat standards. The aim is to define an optimal region for statistics, organize regional policy and provide easier access to funds, programs and support projects for regional economic development. It is vital to observe local community development through its multidimensional nature and conduct multivariate analyses based on

available data to identify the relevant factors of economic development.

The main aims of this study are to investigate the efficiency of different multivariate analysis methods conducted to analyze the effect of socio-economic indicators on the development of local communities in the FBIH and utilize the obtained results to identify a way of their efficient classification with a goal of encouraging a harmonized economic development.

Following the aforementioned aims, the following research questions are posed:

1. Is it possible to identify the socio-economic factors in the characteristics of local communities in the FBIH?
2. Is it possible to determine the direction and intensity of the impact of socio-economic factors on the development of local communities in the FBIH?
3. Is it possible to efficiently classify FBIH municipalities into homogenous groups? What are the defining properties of these groups?
4. Does an eventual classification of local communities create a better surrounding for profiling the policies of regional development?

The aforementioned research questions will be tested based on the latest data available during the period of the research.

2. Methodology

2.1. Data and variables of interest

The successful classification of local self-government units largely depends on the adequate selection of manifest variables. The original intention was to conduct the research on the entire territory of Bosnia and Herzegovina. However, the statistical system in BIH is still under the jurisdiction of two entity institutions for statistics: Institute for Statistics of the FBIH and Institute for Statistics of Republika Srpska, whose reports are not harmonized. Therefore, at the time of the research, it was impossible to provide comparable data at the local community level for both entities in BIH. The survey was conducted in all 79 municipalities in the FBIH. The initial set, after the selection of variables in accordance with the relevant literature, consisted of 21 manifest variables. All data used refer to 2020. The literature review revealed that studies used different indicators as variables of interest, without paying too much attention to their selection and ignoring the fact that successful classification depends, to a large extent, on the adequate selection of variables. It should be taken into account that, in the process of variable selection, some common socio-economic variables were not available at the local

Table 1. Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Unemployment rate	79	13.70	77.40	43.89	12.59
Tax revenue per capita	79	25.00	717.00	156.19	113.19
Average net salary	79	591.00	1381.00	864.16	164.26
Average retirement income	79	252.00	640.00	426.38	62.07
Export-to-import ratio	79	0.00	5386.44	235.53	728.38
Ratio of workers to retirees	79	0.60	3.70	1.46	0.64
Number of business entities per 1000 residents	79	27.00	183.70	63.99	26.44
% of uncultivated land	79	0.00	96.90	44.15	31.67
The size of the local community in km^2	79	9.90	1175.00	330.51	268.81
Population density	79	2.70	6451.60	243.96	787.11
Population migrations in 2020/2013	79	0.00	4.10	0.91	0.48
Vitality index	79	0.00	119.64	58.33	25.61
Share of working age population	79	54.40	76.30	69.02	4.00
Share of unemployed with a university degree	79	0.00	2400.00	312.05	467.87
Number of pupils per 1000 residents	79	4.00	106.00	72.54	20.95
Number of citizens per one doctor	79	48.35	2428.00	1067.93	461.86
Number of citizens per one dentist	79	0.00	20071.00	5852.21	3984.89
Length of main roads in km	79	0.00	115.00	25.42	22.92
Number of registered vehicles per capita	79	0.05	0.65	0.29	0.09
Construction work	79	0.00	96338.00	11113.06	17319.24
Number of TV and radio stations	79	0.00	6.00	1.15	1.20

Source: The authors' calculation

level, such as ethnic and religion structure or variables related to tourism. In addition, some variables, such as hospital bed availability or direct foreign investment, were not appropriate considering that they are irrelevant at local level. Bearing in mind the previously stated facts and the variables used in similar researches in the region and the world, the variables in the paper were initially selected from four categories: macroeconomic indicators, demographic indicators, infrastructure indicators and socio-cultural indicators. Selected variables and basic descriptive indicators are shown in Table 1.

2.2. Selected methods of multivariate statistical analysis

Multivariate analysis methods provide a simultaneous analysis and reduction of a larger number of indicators from one or more sets of indicators, leading to a simplification of input and the reduction of the risk of not fulfilling the premises in the succeeding steps of statistical analysis. This paper will use the multivariate analysis methods to research the interdependence within a set of selected indicators, to analyze the structure and the degree of influence of the extracted factors on economic development and to classify municipalities on the basis of their common features. This study, based on multivariate method features and set goals, relies on the results of factor analysis, i.e. the method of principle component with an aim of researching interdependence, structure and identifying the factors of development. Next, it relies on the results of regression analysis with an aim of determining significance of identified factors on the index of municipalities' development and finally, it relies on the results of the cluster analysis which was used to classify municipalities in the FBIH based on socio-economic factors. Following the intention to include factors of local community development as input variables in the regression model, the principle component analysis (PCA) for factor extraction was chosen as appropriate (Johnson and Wichern 2014).

PCA aims at summarizing most of overall variance of the initial variables to a minimal number of factors needed to make estimates when previous knowledge suggests that specific variance and error variance represent a relatively small portion of the overall variance (Hair et al. 2010). The primary aim of the model is to maximize the explained variability of manifest variables through factor extraction. The essence of principal component analysis is to define a set of non correlated component variables $Y_1, Y_2, \ldots Y_n$, with each component presented in the form of linear combination of manifest variables $X_1, X_2, \ldots X_n$. The first step refers to the determination of the component:

$$Y_{1k} = \sum_{i=1}^{n} a_{1i} X_{ik} \quad (1)$$

for $k = 1,2, \ldots, n$, whose variance $\sigma_{Y_1}^2$ covers the largest possible part of the total variance σ^2 of the manifest variables:

$$\sigma_{Y_1}^2 = \frac{1}{n} \sum_{k=1}^{n} (y_{1k} - \bar{Y}_1)^2 \quad (2)$$

The variance of the second extracted component Y_2 covers the largest possible part of the remaining variance $\left(\sigma^2 - \sigma_{Y_1}^2\right)$. The procedure is repeated until a set of component variables $(Y_1, Y_2, \ldots Y_n)$ is generated. Since $\sigma_{Y_i}^2 = \sum_{i=1}^{n} \sum_{i=1}^{n} a_{li}\, C_{ij} a_{jr}$ the problem is narrowed down to calculating the conditional extreme:

$$\max_{l} \sigma_{Y_i}^2 = \sigma_{Y_1}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{li} C_{ij} a_{jr},$$

$$under\ condition \ \sum_{j=1}^{n} {a_{il}}^2 = 1 \quad (3)$$

Using the method of Lagrange multipliers to determine the conditional extreme in the process of maximizing the variance $\sigma_{y_1}^2$, the following equation is generated:

$$\left|\left[C_{ij}\right] - \lambda l_n\right| = 0 \quad (4)$$

with an n of positive solutions, i.e. characteristic roots/eigenvalues $\lambda_l > 0, l = 1,2 \ldots, n$. It needs to be noted that each characteristic root λ_l is assigned with a characteristic vector $\boldsymbol{a_l}$.

In other words, an n of principal components $Y_l, l = 1,2, \ldots, n$ can be determined for the covariance matrix $C = \left[C_{ij}\right]_{mxn}$ alongside respective eigenvalues $\lambda_l > 0, l = 1,2 \ldots, n$, characteristic vectors $\boldsymbol{a_l} = [a_{il}]_{nx1}$, $\sum_{i=1}^{n} a_{il}^2 = 1,\ l = 1,2, \ldots, n$ and variances $\sigma_{Y_l}^2 > 0,\ l = 1,2, \ldots, n$.

Accordingly, principal component variances $\sigma_{Y_l}^2$ have equal value as eigenvalues of the covariance matrix C, i.e. the following is stated:

$$\sigma_{Y_l}^2 = \lambda_1,\ \ l = 1,2, \ldots, n. \quad (5)$$

The goal of PCA is to determine the component with the maximal variance in each of n phases. This leads to a formation of a descending sequence $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$ of the eigenvalues of the matrix C and a corresponding sequence of principal components: $Y_1, Y_2, \ldots, Y_n$.

If the obtained solution is not clear and interpretable, it is recommended to rotate the factors

with an aim to achieve simple structure in which each variable will be loaded with as few factors as possible, while maximizing the number of high loadings on each variable (Rummel 1970). The paper uses the orthogonal varimax rotation, which is based on minimizing the number of variables that have large loadings on each factor and reducing small loadings from a certain factor to make it even smaller. This rotation technique has proven to be very successful, which is why it is most often used (Kurnoga-Živadinović 2007).

Since the main goal of factor analysis is the condensation of a larger number of initial indicators to a smaller number of latent dimensions, Kaiser criterion is used to reduce the number of principal components. Based on the results of the factor analysis and the obtained factor scores, further analysis is done in multiple linear regression model using the ordinary least squares method (OLS):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + e_i \qquad (6)$$

where the factor scores of socio-economic components are used as independent variables and development index is used as a dependent variable. The development index measures the degree of development of municipalities and it is one of the most important instruments in the regional policy of BIH. It is a composite indicator calculated as a weighted average of multiple basic socio-economic indicators: tax revenue per capita income (X_1), employment rate (X_2), population migration (X_3), share of the elder population (X_4), the educational level of the workforce (X_5). The value of each indicator is normed in a way that the normed value of each indicator is put in ratio with the normed indicator value at the level of the FBIH. Further, the weight for each indicator considered when calculating the development index is determined. The following weights were used in the process of determining the used development index: 25% for X_1, 20% for X_2, 20% for X_3, 15% for X_4 and 20% for X_5 (Regulation on the creation of a development index 2019).

Finally, the non-hierarchical method of cluster analysis based on a previously determined number of clusters was used for the regionalization of municipalities. This method enables moving objects from one cluster to another in different stages of the analysis. The most popular non-hierarchical algorithm is the *K-means,* where, after the selection of the number of *K* clusters, the initialization of *K* clusters and positioning the objects into clusters is conducted randomly. The transfer of objects from one cluster to another is done based on the distance of the centroid of a particular cluster from the objects that are joined to other clusters. If an object is less distant from the centroid of another cluster than its own, it will be transferred to another cluster and the centroids of the clusters will be recalculated. The procedure is repeated until there are no more objects to move, which means that each object is closest to the centroid of the cluster to which it belongs. There are several measures of distance than can be used in cluster analysis. This study uses the most common Euclidean distance.

Formally, a set of objects is given $\{x_1, x_2, ..., x_m\}$, where each one presents an *n* dimensional vector, with coordinates equal to the values of the *n* variables at *m* objects. The K-means method distributes objects in a *K* number of sets (clusters): $C = \{C_1, C_2, ..., C_K\}$ $(K \le m)$ in such a way to minimize the sum:

$$\sum_{k=1}^{K} \sum_{x_i \in C_k} d^2(x_i, \mu_k) \qquad (7)$$

where $\mu_1, \mu_2, ..., \mu_k$ present the cluster centroids $C_1, C_2, ..., C_K$, and *d* is the selected distance measure. The selection of the non-hierarchical method of clusterization was chosen, as the results are less susceptible to the effect of the chosen distance measure and the most distanced data.

3. Results and discussion

The analysis was completed in three main steps: (i) identifying the principal factors of municipality development using PCA, (ii) analyzing the impact of these factors on the municipality development index using the OLS regression model and (iii) classifying the municipalities into clusters.

3.1. Identifying the main factors across the selected data

Factor analysis was repeated several times in the process of identifying factors of economic development. Through the process, the selection of the final set of manifest variables was based on the communalities, anti-image correlations and factor loadings criteria. When analyzing the adequacy of data for the application of factor analysis, the factorability of the input set of variables was examined by using

the Kaiser-Meyer-Olkin measure (KMO = 0.627) and Bartlett's test of sphericity (p = 0.000). After observing the results, it was concluded that the selected data were suitable for factor analysis. The PCA with varimax rotation provided clear factor structure with five main components extracted that explained over 81% of the total variability. Table 2 shows the factor loadings of 14 manifest variables on 5 extracted components.

Table 3 provides an overview of identified components and related variables. The first extracted component is related to: tax revenue per capita, average net salary, average retirement income, population density and share of unemployed with a university degree. Considering that this component is based on total population earnings and corresponding taxes and that it considers the share of the unemployed with a university degree, it is called the *productivity component*. Moreover, population density is fairly expected in this component, bearing in mind that according to Yegorov (2009) and Hummel (2020), population density is related to income, employment rate and total economic growth potential. The second, *demographic*

Table 2. The rotating matrix of factor structure

	Component				
	1	2	3	4	5
Tax revenue per capita	0.715				
Average net salary	0.766				
Average retirement income	0.730				
Population density	0.767				
Share of unemployed with university degree	0.786				
Unemployment rate					-0.837
Ratio of workers to retirees					0.899
Number of business entities per 1000 residents			0.821		
Share of working age population			-0.838		
Number of registered vehicles per capita			0.687		
The size of the local community in km^2				0.888	
Length of main roads in km				0.911	
Vitality index		0.888			
Number of pupils per 1000 residents		0.920			

Source: The authors' calculation

Table 3. Principal components

Number	Component name	Variable
1	Productivity component	Share of unemployed with an university degree Average net salary Average retirement income Tax revenue per capita Population density
2	Demographic component	Vitality index Number of pupils per 1000 residents
3	Economic activity potential component	Number of business entities per 1000 residents Share of working age population Number of registered vehicles per capita
4	Spatial component	Length of main roads in km Size of the local community in km2
5	Employment component	Ratio of workers to retirees Unemployment rate

Source: The authors' creation

component, is related to the vitality index and the number of pupils per 1000 citizens which indicate the population health, fertility and age structure. The variables related to the third component are as follows: the number of business entities per 1000 residents, share of working age population and the number of registered vehicles per capita, which is recognized as *the economic activity potential* of local communities. The fourth, *spatial dimension* of local communities is defined by the size of the local community (km2) and the length of main roads (km). The fifth component is related to the unemployment rate and ratio of workers to retirees, which indicate the state of *employment* in the local community.

Depending on the selected set of manifest variables, different studies that used the PCA method resulted with the identification of different factors. The factors identified in this study are essentially closest to the factors identified by Del Campo et al. (2008): unemployment, economic development, education and two demographic factors. After reviewing other research papers that use factor analysis (Palevičienėa and Dumčiuvienėa 2015; Rašić-Bakarić 2005; Kurnoga-Živadinović 2007), some common characteristics of the identified factors can be observed: the socio-economic status of population, the work status of population, employment, economic activity and education.

3.2. The effect of socio-economic factors on regional development

Factor analysis reduced the number of manifest variables to five socio-economic components. These components represent independent variables in a regression model while the development index is a dependent variable.

The model itself did not meet the assumptions of a regression model on the adequacy of the functional form, homoscedasticity and normality of residuals. An outlier was spotted during the estimation of the linear regression model - the municipality of Centar Sarajevo with a standardized residual of > 4, which could affect the estimate of the regression model. In order to isolate the effect of this outlier on the regression coefficients, we defined a dummy variable and included it in the model.

The estimated model is evaluated by using several tests of linear regression diagnostics. The adequacy of functional form is confirmed by the Ramsey RESET test (p=0.159). The assumption on homoscedasticity was tested using the Breusch-Pagan's test. According to results of Breushch-Pagan's test (p=0.9600), the homoscedasticity assumption cannot be rejected. The assumption on normality of the residuals was tested using the Skewness-Kurtosis test (p=0.603) and the Shapiro-Wilk normality test (p=0,884). Both tests indicate that the normality of residuals assumption cannot be rejected. Considering that the set of exploratory variables contains five uncorrelated factors and one dummy variable, a multicollinearity problem does not exist in the model. Following the Gauss-Markov theorem, the obtained OLS estimator is BLUE (the Best Linear Unbiased Estimator). Table 4 presents the suggested valid model.

Based on the estimated regression model, it can be concluded that 94% of development index variation can be explained by the variations of independent

Table 4. The evaluated regression model

N	F (6.72)	Prob > F	R Squared	Adjusted R Squared
79	217.25	0.0000	0.9477	0.9433

	Coeff.	Stand. Error	t	P > | t |	95% conf. interval	
fac1	0.1831	0.0089	20.40	0.000	0.1652	0.2010
fac2	0.1465	0.0083	17.66	0.000	0.1210	0.1630
fac3	0.1064	0.0087	12.19	0.000	0.0890	0.1238
fac4	-0.0341	0.0083	-4.09	0.000	-0.0508	-0.0175
fac5	0.0874	0.0086	10.16	0.000	0.0703	0.1046
outdummy	0.4309	0.0864	4.99	0.000	0.2586	0.6031
(constant)	0.8471	0.0083	101.97	0.000	0.8305	0.8637

Source: The authors' calculation

variables (adj. R^2=0.9433). All independent variables: productivity component (fac1), demographic component (fac2), economic activity potential component (fac3), spatial component (fac4), employment component (fac5) and the outlier dummy variable are significant on all conventional error levels. All components, except the spatial one, have a positive impact on local communities' development. Following these results, the first two research questions are answered: there are five socio-economic factors among the characteristics of local communities in the FBIH that have significant impact on their development level.

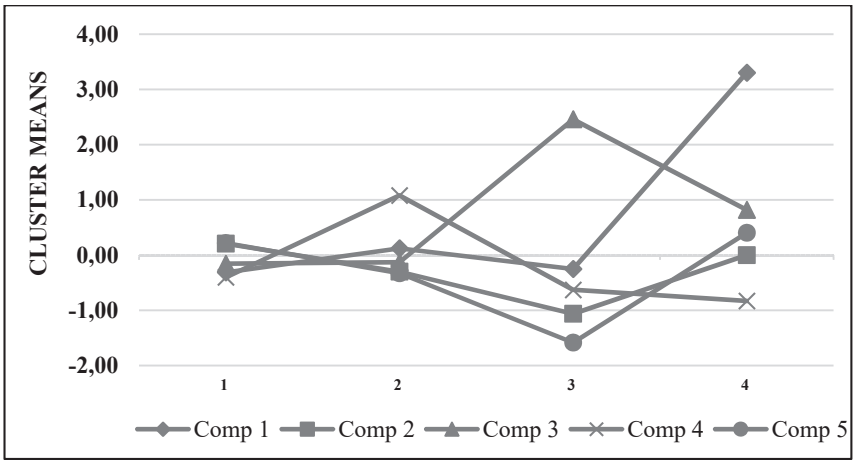
3.3. Regionalization based on socio-economic components

The K-means nonhierarchical cluster method was used for grouping municipalities into clusters. Factor scores obtained by conducting a factor analysis were used as input data. Since factor analysis produced five principal components, five factor scores were calculated for each municipality. The factor scores indicate the relationship between the objects and components, i.e. factor scores would represent the score of each municipality on the underlying latent component. Figure 1 shows the means of factor scores of all five components classified by clusters. The number of clusters (4) was determined based on the results of analysis of variance (ANOVA).

The first cluster contains 49 municipalities. It is characterized with the all five average factor scores close to 0, which indicates that local communities in this cluster have no significant advantages or disadvantages in terms of identified socio-economic factors. The second cluster contains 23 municipalities. This cluster is characterized by the pronounced spatial component, which is expected considering that this cluster contains many spatially large municipalities such as Zenica, Tuzla, Mostar, Bihać, etc. The third factor contains three municipalities: Foča, Pale and Trnovo, characterized by the fact that the pre-war areas of those municipalities were administratively divided between FBIH and entity of Republika Srpska. As expected, the divided i.e. peripheral municipalities, are characterized by a low value of the employment and demographic factor. The economic activity potential component is moderately pronounced in this cluster. In the fourth cluster, the municipalities that make up the administrative area of the city of Sarajevo stood out. As the most developed local communities in FBIH, they are characterized by a very high score of the productivity component and moderately high score of the economic activity potential component. The negative score of the spatial dimension indicates the relatively small area of these municipalities. Grouping of capital cities or regional centers into one cluster is common in most researches based on the regionalization of local communities (Soares et al. 2003, Rovan and Sambt 2003, Lovrinčević et al. 2005).

The cluster analysis, based on factor scores, provided an efficient way to classify FBIH municipalities into homogeneous groups. The values of factor scores inside the clusters enabled the identification of clusters' common characteristics, which served as an answer to the third research question: "Is it possible to efficiently classify FBIH municipalities into homogenous groups? What are the defining properties of these groups?" Geographical distribution of obtained clusters is presented in Figure 2.

Figure 1. The relationship between the clusters and municipality development components



Source: The authors' creation

Figure 2. Municipality classification based on five socio-economic factors

Source: The authors' creation

Testing the differences in the development index among the clusters was conducted by the non-parametric Kruskal–Wallis test since the Kolmogorov-Smirnov test indicated that the development index variable for 79 municipalities does not follow normal distribution. Following the Kruskal–Wallis test, it can be concluded that there are statistically significant differences in the development index per municipality depending on the cluster they belong to. According to the average value of the development index, cluster IV, which consists of municipalities which are a part of the city of Sarajevo, is the most developed one (Table 5).

Identification of common characteristics of the clusters provided a partial answer to the fourth research question: "Does an eventual classification of local communities create a better surrounding for profiling the policies of regional development?" Aiming to provide more useful foundation for planning regional development strategies, clustering is also performed on the basis of each individual socio-economic component.

Table 5. Average development index value per cluster

Cluster		Development index
I	N	49
	Mean	0. 83531
II	N	23
	Mean	0. 75622
III	N	3
	Mean	0.76000
IV	N	4
	Mean	1.68750

Source: The authors' calculation

Clusterization based on the *productivity component* (Figure 3.a.) classified Tuzla, Mostar and four municipalities of the city of Sarajevo into two clusters (cluster II and cluster IV) with the highest values of variables related to the productivity component. The remaining municipalities are classified into two clusters: cluster I with moderate and cluster III with low value of variables corresponding to the productivity component.

According to the *demographic component* (Figure 3.b), the local units are classified into four clusters, with different values of the *vitality index* and the *number of pupils per 1000 residents*. For example, the lowest values of the vitality index (14.03) and the average number of pupils per 1000 residents (15.67) are in the first cluster, while these values are six times higher in the fourth cluster, 82.93 and 90.07 respectively.

According to the component of the *economic activity potential* (Figure 3.c.), no single cluster has a dominantly high value of all variables. The variables *number of business entities per 1000 residents* and *number of registered vehicles per capita* are the highest in cluster III (145.5 and 0.65 respectively), while the *share of working age population* is the highest in cluster IV (71.42).

When the spatial component is taken into account (Figure 3.d), larger geographical municipalities are grouped into clusters 1 and 4, while smaller municipalities are classified into clusters 2 and 3. The average size of the municipality in cluster 4 is 1084.5 km2, so the length of main roads is consequently the longest. On average, these municipalities have over 100 km of main roads. The smallest municipalities are grouped in cluster 3, with an average size of the local community of 145.77 km2 and only 5.57 km of main roads in average.

According to the employment component (Figure 3.e), cluster 1 has the most desirable indicators. It consists of only six municipalities: Tešanj, Centar Sarajevo, Usora, Neum, Kupres, and Ravno. In this cluster, the unemployment rate is the lowest (21.5%) and there are 3 employed persons per one pensioner on average. On the contrary, the second cluster has the highest average unemployment rate of 63.4% and it is the only cluster with a ratio of workers to retirees lower than 1. This cluster includes the following municipalities: Foča, Pale, Sapna, Teočak, Čelić, Kladanj, Trnovo, Zavidovići, Stolac.

Figure 3. Municipality classification based on individual components of development

a. Productivity component

b. Demographic component

c. Component of economic activity potential

d. Spatial component

e. Employment component

Source: The authors' creation

Considering the fourth research question: "Does an eventual classification of local communities create a better surrounding for profiling the policies of regional development?", classification based on individual socio-economic factors enables the identification of local communities that suffer from deficiencies in the context of individual dimensions of regional development. Identified weaknesses at the local level will make it easier for decision-makers to target local communities and create future strategies of harmonized development in the entire FBIH.

4. Conclusions and recommendations

The question of regional inequality, regional development in particular, have taken the spotlight in the literature on economics in the last two decades. Bosnia and Herzegovina, a rather small country in terms of size and population, stands out regarding its big differences at the level of local communities. Since the Stabilization and Association Agreement requires the proposition of a regional division base, this study is aimed to provide a classification of local communities in FBIH, harmonized with the European statistical standards (NUTS). Accordingly, the first task of this research refers to the possibility of identifying socio-economic factors and the evaluation of their impact on the development of local communities. Based on the available data on municipalities in FBIH, five factors were identified: productivity, demographic factor, economic activity potential, spatial factor and employment. It was found that each of them has significant impact on the development of local communities in FBIH. In addition, the task of this study was to find an efficient way of classifying local communities into homogeneous groups and to identify their common characteristics. Based on the identified factors of local communities' development, FBIH municipalities were classified into four homogeneous clusters and their specificities were identified. The resulting classification can serve as the basis for regionalization required by the Stabilization and Association Agreement. Finally, aiming to support the efficient planning of regional development policies, a classification of local communities was carried out on the basis of individual socio-economic factors. Policy makers can utilize these data in the context of planning more efficient strategies and activities, classifications based on productivity, demography, economic activity potential, employment and identifying and evaluating the strengths and weaknesses among FBIH municipalities. The results shown in the paper are a significant support to policy makers in the process of classifying local units and focused enhancement of regional development based on identified features of a given self-government unit and its grouping. The continuation of the research could be focused on expanding the variables considered to be important for research such as indicators of ethnical and religious structure, indicators of housing stock and position, as well as indicators associated with tourism that were not available at municipality level. Further, the limitation of the paper is that the focus is narrowed down to the FBIH alone, hence the analysis should be expanded to encompass the entire area of BIH, i.e. it should be insisted on the harmonization of entity statistical data as data resource.

References

Aragon, Y., Haughton, D., Haughton, J., Eve Leconte, E., Malin, E., Ruiz-Gazen, A., Thomas-Agnan. C. 2003. Explaining the pattern of regional unemployment: The case of the Midi-Pyrenees region. Papers in Regional Science 82: 155-174.

Balan, F. 2015. Testing the Demographic – Economic Paradox for Newly Industrialized Countries: A Panel Data Analysis. Economic Insights – Trends and Challenges. 4: 2-10.

Bartkowska, M., Riedl, A. 2012. Regional convergence clubs in Europe: Identification and conditioning factors. Economic Modelling 29 (1): 22-31.

Bloom, D., Canning, D., Sevilla, J. 2001. Demographic change and economic growth: the importance of age structure. Economic Growth and the Demographic Transition. NBER Working Paper No. 8685.

Botrić V. 2003. Regional differences in unemployment: case of Croatia. Paper presented at the 43rd Congress of the European Regional Science Association, Jyväskylä, Finland, August.

Botrić, V., Šišinački, J., Škuflić, L. 2006. Road infrastructure and regional development: Some evidence from Croatia. In Proccedings from the 46th Congress of the European Regional Science Association" Enlargement, Southern Europe and the Mediterranean, Volos, Greece, August.

Cruz-Jesus, F., Oliveira, T., Bacao, F. 2012. Digital divide across the European Union. Information & Management 49 (6): 278-291.

De Haas, H. 2008. The Myth of Invasion: The Inconvenient Realities of Migration from Africa to the European Union. Third World Quarterly 29 (7): 1305-1322.

Del Campo, C., Monteiro, M. F. C., 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: 600-612.

Feyrer, J. 2002. Demographics and Productivity. Working Paper No. 02-10. Dartmouth College.

Gómez, R., Hernández de Cos, P. 2006. The importance of being mature: the effect of demographic maturation on global per-capita GDP. Journal of Population Economics 21 (3): 589-608.

Gonzalez, J. I., Morini, S. 2000. Socio-economic and business affairs situation in Tenerife Island. Working Paper 2000-07. Universidad de La Laguna.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. 2010. Multivariate data analysis. Upper Saddle River, NJ: Pearson Prentice Hall.

Hartmann, B. 2010. Rethinking climate refugees and climate conflict: Rhetoric, reality and the politics of policy discourse. Journal of International Development. 22 (2): 233-246.

Hummel, D. 2020. The effects of population and housing density in urban areas on income in the United States. Local Economy 35 (1): 27-47.

Johnson, R., Wichern, D. 2014. Applied Multivariate Statistical Analysis. Edinburgh Gate, Pearson Education Limited.

Kelley, A. C., Schmidt, M. R. 2001. Economic and demographic change: A synthesis of models, findings, and perspectives. In Population matters: Demographic change, economic growth, and poverty in the developing world, edited by N. Birdsall, A. C, Kelley, S. W. Sinding, 67-105. Oxford Academic.

Kronthaler, F. 2003. A Study of the Competitiveness of Regions based on a Cluster Analysis: The Example of East Germany. Discussion paper at 43rd Congress of the European Regional Science Association: "Peripheries, Centres, and Spatial Development in the New Europe", Jyväskylä, Finland, August.

Kurnoga-Živadinović, N. 2007. Multivariate classification of Croatian municipalities. Zagreb: Proceedings of the Faculty of Economics in Zagreb 5 (1): 1-15.

Lipshitz, G., Raveh, A. 2010. Socio-economic Differences among Localities: A New Method of Multivariate Analysis. Regional Studies 32 (8): 747-757.

Lovrinčević, Ž., Marić, Z., Rajh, E. 2005. How to optimally regionalize Croatia? Economic review 56 (12): 1009-1160.

Melecký, L. 2012. Evaluation of Cohesion in Visegrad Countries in Comparison with Germany and Austria by Multivariate Methods for Disparities Measurement. International journal of mathematical models and methods in applied sciences 6 (8): 979-989.

Mujakić, M. I. 2010. Local self-government in the Federation of Bosnia and Herzegovina – concept, competences and bodies. Hrvatska Javna Uprava 10 (4): 1045-1058.

Osmankovic, J., Somun-Kapetanovic, R., Resic, E., Halilbasic, M. 2009. The Application of Factorial and Cluster Analysis in Regional Policy. Interdisciplinary Management Research 5: 721-730.

Palevičienėa, A., Dumčiuvienėa, D. 2015. Socio-Economic Diversity of European Regions: Finding The Impact for Regional Performance. Procedia Economics and Finance 23: 1096-1101.

Pejanović, D., Kordej-De Villa, Ž. 2015. Demographic resources as an indicator and factor of disparities in the regional development of Croatia. Social research 24 (3): 321-343.

Peschel, K. 1998. Perspectives of regional development around the Baltic Sea. The Annals of Regional Science 32: 299-320.

Pettersson, O. 2001. Microregional fragmentation in a Swedish county. Papers Regional Science 80: 389–409.

Prelogović, V. 2009. Application of factor analysis in the research of the socio-spatial structure of the city: the example of Zagreb. Zagreb: Hrvatski geografski glasnik 71 (1): 67-85.

Rašić-Bakarić, I. 2005. Application of factor and cluster analysis in the detection of regional inequalities. Privredna kretanja i ekonomska politika 105: 53-76.

Regulation on the creation of development index. http://www.fbihvlada.gov.ba/bosanski/zakoni/2019/Uredbe/4h.htm. (accessed September 13, 2022).

Rovan, J., Sambt, J. 2003. Socio-economic Differences Among Slovenian Municipalities: A Cluster Analysis Approach. Developments in Applied Statistics, Metodološki zvezki 19: 265-278.

Ru'a Vieytes, A., Redondo Palomo, R., Del Campo, C. 2003. Galician socio-economic reality municipal distribution. Revista Galega de Economia 12 (2): 243-262.

Rummel, R. J. 1970. Applied factor analysis. Evanstopn, IL: Northwestern University Press.

Soares, J. O., Marquês M. M. L., Monteiro, M.F. 2003. A multivariate methodology to uncover regional disparities: A contribution to improve European Union and governmental decision. European Journal of Operational Research 145 (1): 121-135.

Stimson, R., Baum, S., Mullins, P., O'Conor, K. 2001. A typology of community opportunity and vulnerability in metropolitan Australia. Papers in Regional Science 80: 45-66.

Yegorov, Y. 2009. Socio-economic influences of population density. Chinese Business Review 8 (7): 1-12.

Zarić, S., Vuković, D. 2010. Regional Competitiveness as a Deciding Factor for Investors. In proceedings of Scientific Meeting „Institutional Changes as a Determinant of Economic Development of Serbia in 2010". Kragujevac, Serbia, February.