Inzinerine Ekonomika-Engineering Economics, **2013, 24(5), 437-446**

Migration Trends Among Regional clusters in Slovakia

Martin Janotka[1], Vladimir Gazda[2], Denis Horvath[3]

[1]*University of Economics*
Tajovskeho 13, 04001, Kosice, Slovak republic
e-mail: matojano@gmail.com

[2,3] *Technical University of Kosice*
Nemcovej 32, 04001, Košice, Slovak Republic
e-mail: vladimir.gazda@gmail.com, denis.horvath@tuke.sk

crossref http://dx.doi.org/10.5755/j01.ee.24.5.3377

The presented study comprises the view on migration in depending on the development level of the region. The development is characterized not only by economic dimension, but expertise also includes social and demographic aspects to gain an overall perspective on the internal regional conditions. With the help of principal component analysis and cluster analysis, we identified homogenous clusters (sets of nearly equally developed districts). Using Kruskall-Wallis test, we have uncovered statistically significant differences among migration flows within given clusters. When considering a migration, two clusters exhibiting apparent differences from others have been detected: the exceptional cluster consisting of large cities districts and cluster of the districts surrounding the capital of Slovakia. Furthermore, for clusters identified in such a way and corresponding to them migration parameters, the cluster-based model has been proposed. Using this, we identified population dynamic trends for the next thirty years. We end by investigating the class of optimally selected migration enforcements (regional migration policies), which may prevent from the depopulation of less favored clusters of districts.

Keywords: *Migration flows, Principal components, Clusters, Analysis of variance, Migration prognosis, Migration enforcements.*

Introduction

Migration processes influence many aspects of economic, social and cultural environment, and thus, understanding of migration trends helps to plan and predict future actions on a regional level. Investigation of spatial distribution and movement of labour forces serves not only for the purpose of regional institutions and agencies, but can also help investors make suitable strategic decisions. In the last years, we can observe two migration trends associated with urbanization and suburbanisation processes that were originally introduced by (Berg *et al.*, 1982). The first one means movement of inhabitants to well-developed urban areas (core) usually with economic motives. In contrast, the suburbanization processes describe people moving from the core to the periphery. Their motives and preferences are usually in the categories of more peaceful and cheaper living under the uninterrupted employment linkages.

Interregional migration becomes a focal point of a variety of economic and social studies. According to (Farwick, 2009), the research framework in this field should cover essential problems like who is moving, where and from what region. (Etzo, 2008), on the other hand, stressed two sides of empirical migration studies depending on whether their aim is to find the determinants of migration or to study the migration consequences. Most of the studies examined the migration by looking on the economic factors of the regions. Even in the first migration empirical study performed by (Ravenstein, 1885), the importance of economic aspects has been emphasized. According to neoclassical theories, wages and unemployment belong to most often used determinants pushing or pulling migrants from/to region. In one of the first studies, (Harris & Todaro, 1970) identified these two variables as the ones, stimulating the migration flows. Their contribution served as a starting base for many research papers. As an example the work of (Fidrmuc, 2004), investigating the influence of wages and unemployment on the migration in the post-communists countries of the Central and Eastern Europe, may serve. The findings confirm that these two factors are significant in relation to the migration flows. In a later study carried out together with (Huber, 2007), Fidrmuc offered an alternative view on the migration activity from the perspective of migrants in the Czech Republic. In this case, the higher income was not proved as a decisive factor for moving of the middle-income group of inhabitants. On the other hand, the ownership of family house in the combination with the high age, are the factors reducing incentives to move. Income and unemployment, as relevant factors for migration, are the topics of studies e.g. of (Hazans 2003; Bloze, 2009; Maza & Villaverde, 2004; Karam & Decaluwe, 2008; Rabe & Taylor, 2010; Andrienko & Guriev, 2003; Cattaneo, 2008). Similarly, (Ciarniene & Kumpikaite, 2011) examined the role of various factors using a structured questionnaire survey focussed on Lithuanian students. In the result, 85 % of them present the economic reasons as crucial for moving to more developed areas. (Eltzo, 2008) studied gross migration flows between regions in Italy with the conclusion that among the investigated economic determinants, the GDP per capita represents the main factor influencing the migration activity. The unemployment and infrastructure endowment level were also recognized as motives, but only when they are associated with the sending region. Another view on the migration is supported by the case studies devoted to spatial aspects of migration. The valuable and useful framework of this research is provided by the well-known gravity models. (Cuching & Poot, 2008) in their survey confirmed the importance of the population sizes and distance between sending and receiving regions. (Kumo, 2006) used the gravity model on population migration in Russia with interregional in- and out-migration flow matrix. He similarly presented densely populated regions experiencing a relatively high volume of inward migration. The study of (Aldashev & Dietz, 2012) strengthens the previous findings regarding the migration data of Kazakhstan. The introduced overview of the research studies offers a large variety of the methodological tools used by identification of the factors influencing migration activities.

The main aim of the paper is to investigate the present and to predict future migration flows among clusters made of grouped Slovak districts with the similar development level. We intend to analyze three main issues. Firstly, if the present migration activity is the highest in the most developed clusters? Secondly, if the presence of suburbanisation and urbanization processes can be observed? Finally, we would like to investigate the clusters, which constitute an object of future depopulation. As we consider migration to be a complex social and economic phenomenon, we propose the methodology enabling a research of multidimensional data to avoid an error caused by the insufficient specifications. Thus the presented study comprises application of three empirical methods: principal component analysis (PCA), cluster analysis and analysis of the variance (ANOVA). The application of the cluster analysis enables to categorize equally developed regions into the clusters. Movement of inhabitants among the clusters is studied and further examined by ANOVA method. The outputs of these purely empirical methods are linked with a theoretical population model of the clusters, which has some predictive potential. The structure of the paper is as follows. In the next section we describe the application of PCA. In the second section, we categorize the districts into the clusters and identify their migration specifics by using ANOVA. Model of the population dynamics with the migration between clusters is presented in the third section. Finally the conclusions are presented.

Principal components analysis

Dataset consists of the Slovak district annual data covering a period of the years 2005 - 2010. Each district is characterized by 54 indicators from economic, social, demographic, infrastructure and regional-development fields. To avoid the non-systematic outlier misspecification, we aggregated the annual data and further used their 6-year period averages. To reduce high dimensionality of the data, we decided to use the Principal Component Analysis (PCA). Before its application, we normalized data towards zero mean and unity standard deviation. Basic statistics of the first fourteen principal components are listed in Table 1.

Table 1

Descriptive statistics of the first 14 components

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	4,55560	2,32970	2,11216	1,87052	1,78271	1,67969	1,31798
Proportion of Variance	0,38430	0,10050	0,08262	0,06479	0,05885	0,05225	0,03217
Cumulative Proportion	0,38430	0,48480	0,56745	0,63225	0,69110	0,74334	0,77551
	PC8	**PC9**	**PC10**	**PC11**	**PC12**	**PC13**	**PC14**
Standard deviation	1,21024	1,17543	1,02675	0,9636	0,92513	0,86492	0,8157
Proportion of Variance	0,02712	0,02559	0,01952	0,0172	0,01585	0,01385	0,01232
Cumulative Proportion	0,80264	0,82822	0,84775	0,8649	0,88079	0,89464	0,90696

Each component represents a specific combination of observed indicators (see later Table 2 and 3) and has a different meaning. The proportion of variance determines the coverage of dataset, e.g. first component covers 38 % of dataset variation and logically features the highest standard deviation. Every further component brings lower proportion and standard deviation. The more principal components we take into deliberation, the bigger cumulative proportion of variance we have. Ideally, the most effective way for data analysis by using the PCA is to achieve high cumulative proportion with only couple of components. Therefore, several methods have been developed to provide an eligible number of principal components, which will serve for further analysis. One of the most frequently used is the method, which exploits the histogram of variance for each component. According to Figure 1, the kink in a decrease of variance separates first two most relevant components. However, they reproduce only 48 % of the variance, which seems to be insufficient to describe the whole data set. Another method, sometimes called eigenvalue-one-criterion (Also known as Kaiser´s criterion) drops out components with eigenvalues less than one. Because the standard deviation of a component is given as a square root of the corresponding eigenvalue of the dataset variance-covariance matrix, using the rule and the data listed in Table 1, we identified first ten components intended for further analysis. This relatively large number of components is still difficult to interpret, and, at the same time, also to identify some differences among them. Thus the final reduction is realized by a criterion demanding to explain at least 70 % (For natural sciences, it is recommended to cover 95 % of variance, for social sciences just 65 % (Hair *et al.,* 2005)) of the variance of all components. This criterion led to the choice of the first five components. In Table 2, we introduce loadings of 10 variables with the highest contribution to each underlined component.

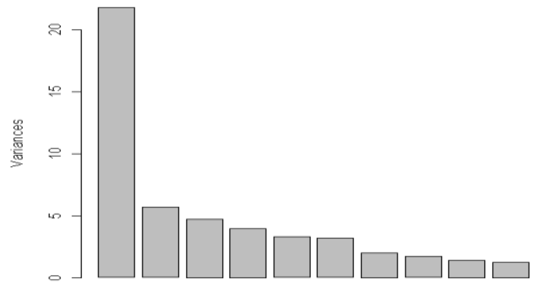


Figure 1. Histogram of the principal components variance

Table 2

Loadings of the most significant variables for components 1-5

Variable	**PC1**	**Variable**	**PC2**	**Variable**	**PC3**
CommercialCompanies/ TotalOrganizations	-0,194	PeopleInProductiveAge/ Population	0,326	BudgetedCompanies/ TotalOrganizations	-0,375
realWage/avgSR	-0,194	LegislatorsAndManagers/EA	-0,234	1st Class roads in km	-0,327
BankBranches/Total	-0,179	YoungApplicants/EA	-0,226	AcommodationFacilities/ Total	-0,303
Tech.Med.Ped.Emp/EA	-0,179	Tech.Med.Ped.Emp /EA	-0,225	NumberOfTowns/Total	-0,280
NumberOfMen/ Population	0,179	LowLevelAdminstration/EA	-0,221	NumberOfVillages/Total	-0,277
NumberOfWomen/ Population	-0,179	Service.Employees.EA	-0,215	StayOvers/Total	-0,266
ScientistsAndSpecialists/EA	-0,178	Long-termApplicants/EA	-0,215	BeginnedApartments/Total	-0,249
IndexOfGrowingOld/AvgSR	-0,177	ScientistsAndSpecialists/EA	-0,211	NumberOfTourist/Total	-0,247
Service.Employees/EA	-0,174	Number of Roma/Population	-0,208	FinishedApartments/Total	-0,237
FDI/Total	-0,174	unemp.avg/SR	-0,204	2nd Class roadsin km	-0,233
Variable	**PC4**	**Variable**	**PC5**		
PeopleInPostProductiveAge/ Population	0,310	Long-termApplicants/EA	0,278		
Businesses50.249/Total	0,282	unemp.avgSR	0,274		
IndexOfGrowingOld(EU)/AvgSR	0,269	EAMen/EATotal	-0,219		
Businesses0.19/Total	-0,263	EAWomen/EATotal	0,219		
PeopleLivingInVillages/ Population	-0,238	NumberOfVillages/Total	0,211		
PeopleLivingInTowns/Population	0,237	NumberOfMen/Population	-0,204		
2nd Class roads	0,236	NumberOfWomen/Population	0,204		
AverageAgeOfRegion/AvgSR	0,228	AverageAgeOfRegion/AvgSR	0,196		
Businesses20.49/Total	0,216	Motorways in km	-0,195		
IndexOfGrowingOld/AvgSR	0,215	AcommodationFacilities/Total	-0,191		

EA = Economically active population

The first component represents a mix of economic and labour market infrastructure and demographic variables. Generally it defines overall status of the region and describes business environment with partial view on every investigated area. Economic dimension include real wages level, share of commercial companies on total organizations, share of bank branches on total and FDI share. The composition of the economic activity represents the employment in the science, health and educational sector. Finally, index of growing old and number of men and women complement entire composition of the component from demographical aspect. When considering the first component, its high values mean the overall lagging of the region.

The second component has very clear structure – labour market and demographic characteristics. It is described by the share of population in productive and pre-productive age, young and long-term applicants for a job and the unemployment rate. Moreover, it includes share of Roma people, who are often marginalized and socially excluded. According to study of (Popper *et al.*, 2009), only 16,6 % of adult Roma population work on permanent contract. Surprisingly, the economic activity in the science, health and educational sectors as well as share of the legislators and managers is here also included and moreover, with the same sign of loadings as the negative characteristics of the labour market (unemployment, job applicants, etc.). This fact can be interpreted by the need to administrate the lagging region, where the free market job positions absent and the positions administratively enforced prevail. In the case of the second component, high values of the component indicate poor state of the labour market and its demographic characteristics.

The role of the infrastructure and tourism is comprised in *the third principal component.* Road infrastructure and number of built apartments enter into the economic attractiveness of a region. On the other hand, it is associated with a social infrastructure reflected in number of organizations budgeted from the state. From another point of view, the region's attractiveness is also determined by a triplet of tourism indicators (number of tourists, accommodation facilities and overnights staying). High values of this component represent low attractiveness of the region.

The structure of the *fourth component* is introduced primarily as a combination of demographic variables and different types of businesses categorized by a corresponding number of employees. The demographic factors are associated here with the indicators of the ageing population. According to (Megyesiova & Hajduova, 2012), the ageing is considered to be a crucial factor of the future economic development. This component informs about a lack of labour forces (more in the view of quantity than quality) within the region. It means that positive values of the component represent regions, which are getting old, with urban rather than village type population. The structure of businesses is oriented on the middle and large companies.

The *fifth component* contains again indicators with an emphasis on demographic structure and unemployment status. Similarly, high levels of these components explain lagging countryside regions with high unemployment, old population and disrupted infrastructure.

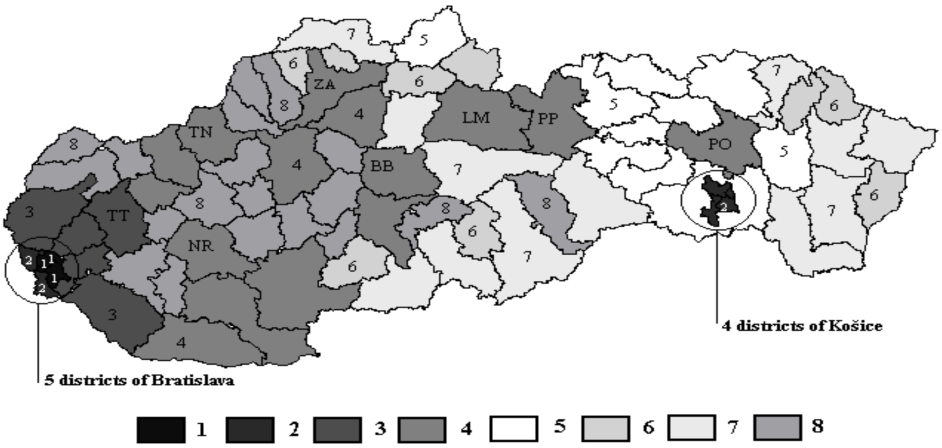
Table 3

Significance of variables in first 5 components

		PC1	PC2	PC3	PC4	PC5
Wages level	Real wages	**HIGH**	LOW	LOW	LOW	LOW
Infrastructure level	Road	LOW	LOW	**HIGH**	**HIGH**	LOW
	Financial	**HIGH**	LOW	LOW	LOW	LOW
	New-built flats	LOW	LOW	**HIGH**	LOW	LOW
Businesses level	Commercial Companies	**HIGH**	LOW	LOW	LOW	LOW
	Small BU	LOW	LOW	LOW	**HIGH**	LOW
	Large BU	LOW	LOW	LOW	**HIGH**	LOW
Unemployment level	Rate	LOW	**HIGH**	LOW	LOW	**HIGH**
	Candidates	**HIGH**	**HIGH**	LOW	LOW	**HIGH**
Age level	Preprod	LOW	**HIGH**	LOW	LOW	LOW
	Prod	LOW	**HIGH**	LOW	LOW	LOW
	Post prod.	LOW	LOW	LOW	**HIGH**	LOW
Tourism level	Tourism	LOW	LOW	**HIGH**	LOW	LOW
Demographic level	Variables	**HIGH**	LOW	LOW	**HIGH**	**HIGH**

Cluster analysis and ANOVA

The regions are now defined by characteristics of the above mentioned five components. Using them, the Ward´s method as a clustering algorithm is applied. The eight-cluster categorisation is displayed in the map of Slovakia (Figure 2). For deeper understanding, we interpreted the principal component values in the sense of Table 3 followed by a detailed analysis of the original dataset. Here, according to the considered criteria, the clusters 1, 2 and 3 appear to be extraordinary highly developed with respect to the rest of the clusters. On contrary, specification of the set of the "underdeveloped" clusters is rather fuzzy and depending on the selected criteria. However, clusters 5, 6 and 7 rank the lagging position according to the most criteria. It is obvious that most developed districts with best business environment are clustered separately. It can be recognised by observing districts of Bratislava (Bratislava I, II and III), which are grouped together. It results primarily from an economic character of the regions characterized by a high level of economic development, wages and low unemployment. The infrastructure attains very high level with well developed road and rail network and the international airport. Bratislava itself became a centre of trade because of its central location in the financial and other business areas (Bratislava is the capital city and the biggest city in Slovakia with direct connection to Austria (Vienna), Hungary and the Czech Republic). As a site with high wages and opportunities for a suitable job application is attractive for migrants coming from the poorer regions. Furthermore, other Bratislava districts (IV and V) and Kosice districts (second largest industrial centre in Slovakia) establish the second cluster. Again, this cluster incorporates highly developed regions. It should be noted that Kosice city is located in the eastern part of Slovakia.



Poprad (PP) - Liptovsky Mikulas (LM) - Presov (PO) - Nitra (NR) - Zilina (ZA) - Banska Bystrica (BB) - Trencin (TN) - Trnava (TT)

Figure 2 Map of Slovakia according to Ward´s clustering method

Clusters 1 and 2 contain a small number of districts (nine districts of Bratislava and Kosice). For proper use of ANOVA, we merged them into one denoted as 1+2. As indicated above, the cluster contains nine most developed cities districts. Economic conditions for enterprises, infrastructure, foreign direct investments, skilled labour and qualified human resources, low unemployment and high real wages create possibilities for economic growth and sustainable development. Moreover, good business environment and attractive labour market encourage people to move there. As shown in Figure 3, migration activity in these areas is high. Both immigration and emigration attain the maximum average value if compared to other clusters.

Cluster 3 encompasses the districts surrounding Bratislava city. These regions are also developed as they are taking an advantage of being close to the biggest center in Slovakia. The district of Trnava (TT) was also connected to well developed regions because of foreign direct investments and direct connection to Bratislava (highway, railway). The average of finished and built apartments is the second highest in Slovakia. The attractiveness of these districts is confirmed also in Figure 3, where the median immigration is the second highest.

Cluster 4 contains developed centres; either tourist resorts (districts of Poprad (PP) and Liptovsky Mikulas (LM), which are located near The High and Low Tatras or agricultural districts in the south-west Slovakia. Moreover, there are included 5 of 8 centres of NUTS III regions – Presov (PO), Nitra (NR), Zilina (ZA), Banska Bystrica (BB) and Trencin (TN). The remaining NUTS III centers are already included in previous clusters – Bratislava, Košice and Trnava (TT). According to real wages and unemployment, this cluster took position just after the previous ones. The value of foreign direct investments is the third highest. Furthermore, we see the highest density of the road network and highest number of miles of highways, which clearly contribute to levels of development.

Poorer regions of Eastern Slovakia are separated in *cluster 5*. They are characterized by low real wages, extremely high unemployment and a significant proportion of the Roma minority. The proportion of pre-productive age inhabitants is the highest of all clusters. The regions are not highly developed. Foreign direct investments are the lowest within all clusters. For possible reasons, the lack of the qualified labour forces, poor infrastructure and weak innovation activity are frequently counted.

Regarding to *cluster 6*, it comprises not developed regions, because of the lowest average population, the lowest road network density and mainly forest character of districts. The situation is confirmed also by the average number of new apartments, which is the lowest among all clusters. The real wages attain the second lowest and the unemployment the third highest level that clusters attribute to the poor regions. Cluster 6 consists of the considerably small districts regarding its population and area. The rural character of the 6-th cluster seemingly becomes a target for suburbanization movement from the surrounding regions. Analyzing map in Figure 2, we come to the conclusion that the districts of this cluster are neighbouring by larger districts of clusters 4 and 7 with comparably better industrial potential. Inspecting the box plots in Figure 3 and migration data introduced in Tables 5 and 6 we get the argument supporting this hypothesis.

In contrast to cluster 6, *cluster 7* includes also agricultural districts. Although the area and population of hese districts is much higher than in cluster 6, the remaining characteristics are the same. The highest unemployment, which is determined by long-term unemployed people, low real wages and smallest number of motorways, predetermine weakness of these regions. On the other hand, the number of businesses with 0-19 employees represents the highest value, which can be explained by the dominance of the small businesses due to the lack of job vacancies.

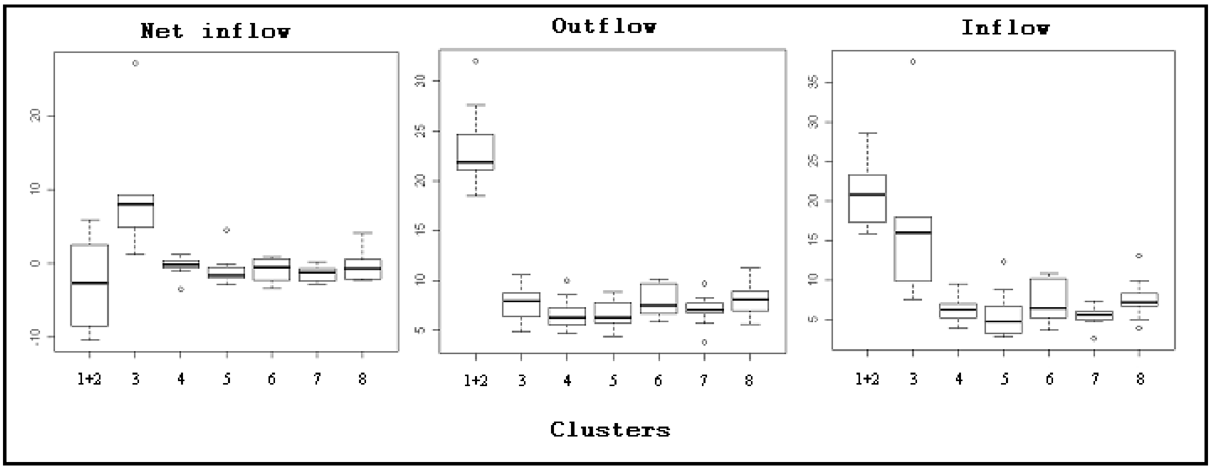


Figure 3. Box plots of the Net Inflow, Outflow, and Inflow (as per-mille of population in region).

Cluster 8 includes regions in the West Slovakia. Regarding to developmental level, they are located between well and weakly developed cluster. They could belong to developed clusters, because of advantage of developed infrastructure (road network density is the third highest) and good connectivity to Bratislava. However, the wage level is still below the level of first three clusters.

In Table 4, we presented some migration and demographic indicators of given clusters. First, the number of live births is almost the same as in all clusters themselves apart cluster 5. This was set up as an outlier, due to high natality especially of Roma. Since the relative number of the deceased is low, cluster has the highest population natural growth. Cluster 1 is characterised by the highest share of deaths and consequently by the highest population drop. The international migration is also included in the table, although our analysis covers exceptionally interregional Slovak migration. In case of cluster 1, immigration from abroad reaches the highest level. This can be explained by the attractiveness of the region. Otherwise, international migration plays a minor role if compared to internal migration. The reason is that migration data are based on the change of residence address. For international migration, young people, as the most active migrants move abroad without changing permanent address, so they are officially still living in Slovakia, however, actually living and working abroad. Next, we can observe that clusters 1+2 and 3 absorb a large portion of migrants. On the other side, clusters 1+2 loose many migrants moving into the peripheral areas. To determine, whether there is statistically the same level of the migration activities in all of the considered clusters, we decided to use ANOVA test.

Table 4

Migration and Demographic statistics for Clusters as per-mille of population in region (indicators counted cumulatively for each cluster over 6 years)

Cluster	1	2	3	4	5	6	7	8
Number of live-born children	10,58	10,69	10,54	9,49	14,32	10,22	10,71	9,12
Number of the deceased	12,62	7,90	9,29	10,05	8,48	10,88	10,68	10,73
Inflows from Abroad	4,73	1,32	1,59	1,34	0,52	0,67	0,74	0,98
Outflows to Abroad	0,68	0,43	0,28	0,36	0,31	0,20	0,28	0,25
Inflows from Slovakia	25,52	22,52	17,85	6,26	5,66	7,28	5,51	7,50
Outflows to Slovakia	25,85	18,55	7,72	6,53	6,55	8,05	6,97	7,94

The null and alternative hypotheses are:

$$H_0: \eta_{1+2} = \eta_3 = \dots = \eta_8 \quad (1)$$

H_1: at least one cluster has different mean η

where η_i is mean value of the *i*-th cluster. Clusters investigation is threefold: by net migration, immigration and emigration. The underlying assumption of the parametric ANOVA is the clusters variance homogeneity of the respective migration indicators. The null hypothesis (of equal variances) has been rejected according to the Bartlett test (p-values < 0.01 in case of all 3 variables). Thus, we decided further to use the Kruskall-Wallis test, which is robust against the ANOVA assumptions violation and provides test (1) on median values. Using it, we accepted the alternative hypothesis at 1 percent significance level in all three cases considered.

To achieve a clear view which clusters are the same and which are different, it is necessary to use one of the post-hoc tests. As discussed above, the migration data exhibit the heterogeneity of variance. The Tukey's honestly significant difference test seem to be inapplicable, thus we decided to use the heteroscedasticity robust Games-Howell test. For the particular migrations, we found out that cities districts (1+2) are characterized by a much stronger migration activity that sets them apart from the others. In case of inflows both the clusters (1+2) and 3 are revealed as statistically equal ($p=0,72$), but each of them differs significantly from others (in all considered pairs at 1 percent level). The immigration activity is high for both, albeit for different reasons. Increased immigration in cluster 1+2 is by urbanisation, which is quite obvious in immigration from all considered clusters except clusters 6 and 8 (see Table 5). On the other hand, the immigration to the cluster 3 is intensified by the suburbanisation process attracting people from the Bratislava districts to the neighbouring rural districts (see Table 5). In their research, Gajdos and Moravanska (2008) found good connections to towns, better options for living and more healthy environment as main reasons for moving to peripheral areas. If we look onto outflow, only cluster 1+2 appears as statistically significantly different from the other clusters ($p < 0,01$ in all considered pairs). Again, this can be explained by the suburbanization trends. For other clusters, we can conclude that they do not statistically significantly differ. In case of net migration, the Games-Howell test shows that cluster 3 is different from the others ($p < 0,01$). This is explained by the imbalance of migration flows where the inflow markedly prevails over outflow. At the same time, it is to mention that in case of the net migration the assumption of mutual independence of the observations in different clusters is violated. Thus, the results of the Kruskall-Wallis and Games-Howell are presented here just for the demonstration purposes.

Table 5

Six years average of migration flows between districts pre 1000 inhabitants

From\to	Cluster 1+2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Cluster 1+2	1,7119	1,0672	0,0668	0,1625	0,0502	0,0676	0,0956
Cluster 3	0,2579	2,0244	0,0793	0,0132	0,0250	0,0204	0,1405
Cluster 4	0,1667	0,1696	0,7030	0,0964	0,1356	0,0837	0,1923
Cluster 5	0,1912	0,0612	0,0612	0,7520	0,0617	0,0844	0,0209
Cluster 6	0,0431	0,0359	0,0374	0,0255	0,8021	0,0695	0,0269
Cluster 7	0,1187	0,0725	0,0656	0,0580	0,2268	0,8018	0,0357
Cluster 8	0,0681	0,1185	0,0769	0,0085	0,0368	0,0176	0,5217

The above mentioned empirical findings supported existing researches (Kuma, 2006; Cuching & Poot, 2008; Aldashev & Dietz, 2012, and others). The developed regions in Slovakia experienced bigger migration activity than lagging regions. The results confirmed also the presence of urbanization (migration flows to developed centers) and suburbanisation trends (migration flows from development centers in nearby sub-areas). The performing of this analysis might further serve for the purpose of regional policy decision-making in order to help less developed regions to converge. The public effort for co-creation of suitable living condition may prevent the region from depopulation and thus losing the potential labour forces. In order to find the most critical regions and their perspectives from the standpoint of depopulation, in the next section we construct and examine the model of the population dynamics.

Population dynamics of clusters

This section concerns the proposal of the empirically supported model based on the principles of the spatially distributed social-dynamic population models (see e.g. Bijak (2004)). Since the clusters have been selected by previous empirical methodologies as relatively homogeneous elements of the original heterogeneous system, in order to follow their evolution, we suppose it is sufficient to limit ourselves to the population model designed for the homogeneous subunits - clusters. As the information on the asymptotic saturation of the population is not available for the observation, we restrict to a linear formulation. If $N_i^{\,t}$ is the number of individuals at year *t*, the population size in the next year is

$$N_i^{\,t+1} = 1 + r^G \; N_i^{\,t} + T_i^{(t)}, \quad (2)$$

where; $T_i^{(t)}$ is the total number of inter-cluster transitions from the cluster i minus impact of the inter-cluster transitions out of the i-th cluster (all transitions within Slovakia). The total population growth rate r_i^G consists of the contributions.

$$r_i^G = r_i^B - r_i^D + r_i^I - r_i^E, \quad (3)$$

where r_i^B, r_i^D stands for the population born and death rate, respectively. The model assumes openness of Slovakia to migration. The phenomena are described by the rates of international immigration r_i^I and emigration r_i^E. Thus, now for some explanation, the number of emigrations is described by $r_i^E N_i^{\,t}$. The number of transitions from the ith cluster and out to the cluster (within Slovakia) is given by (Weidlich (2000)).

$$T_i^{\,t} = {}_{j\in Cset\setminus\, i}\; N_j^{(t)} r_{j,i}^T - {}_{j\in Cset\setminus\, i}\; N_i^{(t)} r_{i,j}^T, \quad (4)$$

where $r_{j,i}^T$ describes the number of transitions (migrations) per person per year. The growth and transition rates have been estimated from the Table 4 as the ratios of the six year average of the migrations divided by the actual number of residents of a given cluster. In our study the time-course of the population of very different clusters we decided to define the relative quantity.

$$R_i^{(0,t)} = \frac{N_i^{(t)} - N_i^{(0)}}{N_i^{(0)}}, \quad (5)$$

in which the superscript (0) represents the last issue of the database (year 2010).The results of the calculation are depicted in Figure 4a and 4b.

The relative importance of the transfer for the population dynamics has been characterized by the ratio $\frac{T_i^{\,t}}{N_i^{\,t}}$.Its relatively large values have been observed in the cluster (1+2) (positive value) and also 3 (negative value). Figure 4a implies that the depopulation of the clusters 6, 8 may be regarded as a most critical. Given this, consider what happens if certain links will be suddenly strengthened in the future by the regional conceptual decisions. Another words, we claim to model the consequences of the effect of making clusters 6, 8 more attractive.

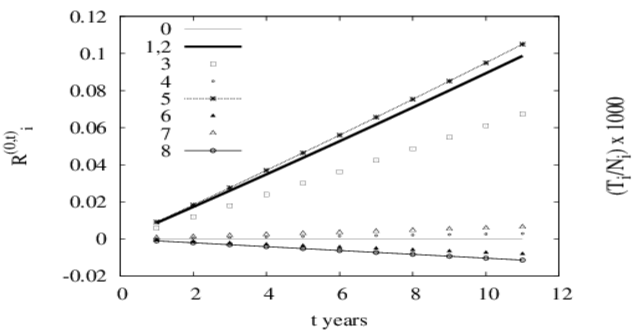


Figure 4a.The model-estimated relative population changes monitored over the next thirty years. Nearly linear shape is observed only due to shortness of time. It has been verified that much longer runs yield exponential time dependencies.

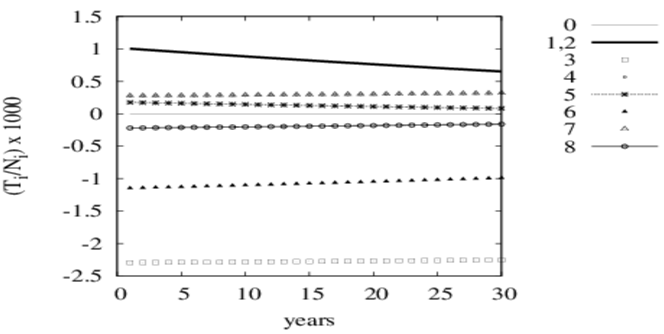


Figure 4b. Figure showing the relative importance of the transfer term $T_i^{\,t}$ with respect to the given cluster size. The results are nearly constant over the 30 years. The most pronounced decrease in proportion is seen in cluster 1, largely due to a high denominator effect of $N_i^{(t)}$.

We suppose that this aspect can be achieved by the reparametrization of the links directed in clusters 6 and 8 by using single parameter $c^T > 1$. Formally it means the performing series of the particular multiplicative replacements.

$$r^T_{j,i} \to r^T_{j,i} c^T, \qquad i \in 6,8 \;, \quad j \in Cset\setminus\{i\} \qquad (6)$$

where $r^T_{j,i}$ denotes the original value of the matrix element and $r^T_{j,i}c^T$ is the value achieved after the particular link strengthening with help of c^T. Note that the remaining links $i \notin 6,8$ are left untouched. It should be also noted that due to the system complexity, the above manipulations are only of the hypothetical kind, because the consequences of such manipulations might not be trivial and quite different from what we really want to achieve.

Our multiplicative modifying of links simply claims to avoid from the unwanted population drop (after 30 years of evolution), which may be expected for less attractive clusters 6 and 8. To investigate this we performed a series of simulation runs for gradually changing c^T by starting from the identical initial real-world conditions. For each run, which covers thirty years of the hypothetical system evolution toward the future, the transition matrix elements $r^T_{j,i}$ and $r^T_{j,i}c^T$, are kept constant. The final values $R_i^{(0,t=30\,years)}$have been recorded and depicted in Figure 5. We see that larger artificial preferences of 6, 8, as we have expected, affect as well a situation in other clusters (the most remarkable influence is observed in the case of the cluster 1+2). What is quite positive finding, the expected population supply of 6 and 8 has been achieved without drastic drops in other clusters. Rather unrealistic feature of our finding is that redirection of the population flows requires action of relatively high c^T. The zero growth compensation point values $R_i^{(0,t=30\,years)} \simeq 0$ for cluster 6 is attained byc^T =3.2. Even more demanding, should be halting of depopulation of the cluster 8 by means of c^T =4.3. Thus further optimization analysis and strategies are needed to improve the above inverse methodology directed toward preventing great future depopulation of certain regions.

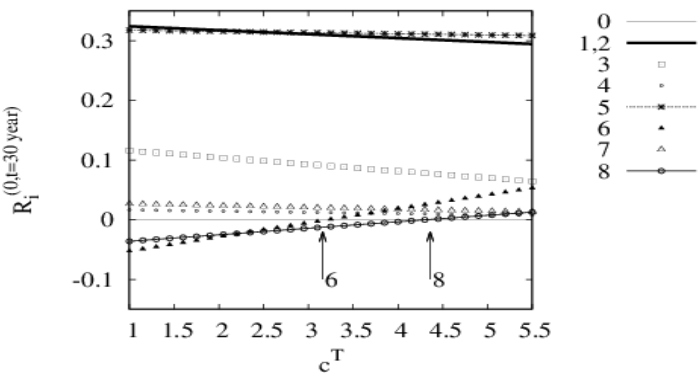


Figure 5. The hypothetical influence of the transfer enforcement of c^T by looking at the results of the 30 year evolution according to the model with a modified transition matrix. The compensation points where population drop changes turns to the increase for the most problematic clusters 6 and 8 are shown by the arrows.

Conclusions

In this paper we aim to show how internal migration behaves in the regions with different levels of development. In consistency with gravity models and already realised empirics, we observed the dominating migration activity occurring in the most developed centers in Slovakia – city districts of Bratislava and Kosice. The level of development of these districts compared to other regions is very high, which is reflected in their attractiveness and migration. Moreover, the results in this paper confirmed the presence of urbanization and suburbanisation processes. Firstly, urbanization was exhibited in inward migration of cities districts. These districts are labelled as the outliers with respect to the inflows. Secondly, suburbanisation process is revealed in the significantly different outward migration of cities districts. In addition, districts surrounding Bratislava city become the outliers for inflows. These facts emphasize the importance of the level of development of regions for migration flows.

On the other side, the facts might help in regional policy for decision-making. The regions with high emigration (and low immigration) lose the potential of labour forces and their chances for possible convergence become lower. Therefore, it seems to be important for state, municipalities and public organizations to encompass the migration into the process of decision-making aimed to manage support.

To localize critical regions highly susceptible to depopulation we decided to perform forecasts of the population evolution trends, which could be stimulating for regional policy. The relative homogeneity of numerically specified cluster has brought an important and innovative aspect of our approach with respect to the modelling of the population dynamics. Focusing on regions with expected future depopulation, we simulated prognosis with artificially modified transfer enforcement to localize a parametric turning point, in which the future population decline turns into population growth. By this way the population increase may be achieved in all clusters. Less feasible in practical policy terms it seems adjusting of the original migration by relatively high transfer enforcement.

References

Aldashev, A., & Dietz, B. (2012). Economic and Spatial Determinants of Interregional Migration in Kazakhstan. IZA Discussion Papers, No. 6289. Available from internet: http://ftp.iza.org/dp6289.pdf

Andrienko, Y., & Guriev, S. (2003). Determinants of Interregional Mobility in Russia: Evidence from Panel Data. William Davidson Institute Working Papers Series No. 551. Available from internet: http://www.wdi.umich.edu/files/ Publications/WorkingPapers/wp551.pdf

Berg, L. van den, Drwet, R., Klaasen, L. H., Rossi, A., & Vijverberg, C. H. T. (1982). Urban Europe: A Study of Growth And Decline, Oxford: Pergamon Press.

Bijak, J. (2004). Forecasting International Migration: Selected Theories, Models, and Methods, Central European Forum For Migration Research in Warsaw, CEFMR Working Paper 4/2006.

Bloze, G. (2009). Interregional Migration and Housing Structure in an East European Transition Country: A View of Lithuania 2001-2008. *Baltic Journal of Economics*, 9(2), 47-66.

Cattaneo, C. (2008). The Determinants of Actual Migration and the Role of Wages and Unemployment in Albania: an Empirical Analysis. *The European Journal of Comparative Economics*, 5(1), 3-32.

Ciarniene, R., & Kumpikaite, V. (2011). International Labour Migration: Students Viewpoint. *Inzinerine Ekonomika-Engineering Economics*, 22(5), 527-533.

Cushing, B., & Poot, J. (2008). Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling. *Papers in Regional Science*, 83(1), 317-338. http://dx.doi.org/10.1007/s10110-003-0188-5

Etzo, I. (2008). Determinants of Interregional Migration in Italy: A Panel Data Analysis. MRPA Paper No. 8367. Available from internet: http://mpra.ub.uni-muenchen.de/5307/

Etzo, I. (2008). Internal Migration: A Review of the literature. MRPA Paper No. 8783. Available from internet: http://mpra.ub.uni-muenchen.de/8783/1/MPRA_paper_8783.pdf

Farwick, A. (2009). Internal Migration. Challenges and Perspectives For The Research Infrastructure. Working Paper Series of the German Council for Social and Economic Data No. 97. Available from internet: http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_97.pdf

Fidrmuc, J. (2004). Migration and Regional Adjustment to Asymmetric Shocks in Transition Economies. *Journal of Comparative Economics,* 32 (2), 230-247. http://dx.doi.org/10.1016/j.jce.2004.02.011

Hair, J. F., Black, B., Babin, B., Anderson, R. E., & Tatham, R. L. (2005). The Multivariate Data Analysis. Prentice Hall, 6th Edition.

Harris, J., & Todaro, M. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *American Economic Review*, 60, 126-142.

Hazans, M. (2003). Determinants of Inter-regional Migration in the Baltic Countries. Zentrum für Europaische Integrationforschung (ZEI) - Research papers, B 17-2003. Available from internet: http://econstor.eu/bitstream/ 10419/39479/1/370774256.pdf

Huber, P., & Fidrmuc, J. (2007). The Willingness to Migrate in the CEECs Evidence from the Czech Republic. *Empirica,* 34(4), 351-369. http://dx.doi.org/10.1007/s10663-006-9028-6

Gajdos, P., & Moravanska, K. (2008). Suburbanization in the Slovakia. Available from internet: http:// www.sociologia.sav.sk/cms/uploaded/1242_attach_suburbanizacia.pdf

Karam, F., & Decaluwe, B. (2008). Migration Impact on Moroccan Unemployment: a Static Computable General Equilibrium analysis. HAL, Post-Print No. halshs-00331322.

Kumo, K. (2006). Interregional Population Migration in Russia: Using an Origin-to-Destination Matrix. Institute of Economic Research, Hitotsubashi University, Discussion Paper No. 483. Available from internet: http://hermes-ir.lib.hit-u.ac.jp/rs/bitstream/10086/13463/1/DP483.pdf

Maza, A., & Villaverde, J. (2004). Interregional Migration in Spain: A Semiparametric Analysis. *Review of Regional Studies,* 34(2), 156-171.

Megyesiova, S., & Hajduova, Z. (2012). Demographic Challenges Across the European Union Member States. *Workplace Health & Safety*, 60(7), 321-326. http://dx.doi.org/10.3928/21650799-20120625-05

Popper, M., Szeghy, P., & Sarkozy, S. (2009). Roma Population a Healthy: An Analysis of Situation in Slovakia. Partners for Democratic Change Slovakia. Available from internet: http://www.gitanos.org/upload/13/60/Eslovaquia-corrected.pdf

Rabe, B., & Taylor, M. P. (2010). Difference in Opportunities? Wage, Unemployment and House-price Effects on migration. Institute for Social and Economic Research, Working Paper Series, No. 2010-05. Available from internet: https://www.iser.essex.ac.uk/publications/working-papers/iser/2010-05.pdf

Ravenstein, E. G. (1972). Die Gesetze der Wanderung I. Regionale Mobilitat. Nymphenburger Verlagshandlung, 41-58.

Statistical Office of the Slovak Republic, Regional Database (REGDAT). Available from internet: http://px-web.statistics.sk/PXWebSlovak/index_en.htm

Weidlich, W. (2000). Sociodynamics - A Systematic Approach to Mathematical Modelling in the Social Sciences. Amsterdam: Harwood Academic Publishers.

Martin Janotka, Vladimír Gazda, Denis Horváth

Migracijos kryptys Slovakijos regioniniuose klasteriuose

Santrumpa

Šiame tyrime panaudoti trys empiriniai metodai: pagrindinių komponentų analizė (plg. angl. *principal component analysis* (PCA)), klasterio analizė ir pokyčių analizė (*plg. angl.* analysis of variance (ANOVA)). Tyrimo tikslas – išanalizuoti tarpregioninės migracijos tarp homogeniškų klasterių srautus. Šiuos empirinius metodus papildo teoriniu gyventojų skaičiaus klasteriu pagrįstas modelis, kuriuo mes bandėme nustatyti galimą kritinį klasterio gyventojų skaičiaus mažėjimą 30-ies metų laikotarpiu gyventojų ir galimą migracijos spaudimą, kuris galėtų padėti pagerinti situaciją klasteriuose, kuriuose sumažėjęs gyventojų skaičius. Nagrinėtus duomenis sudaro Slovakijos rajonų metiniai duomenys (54 rodikliai), apimantys 2005 – 2010 metų laikotarpį. Mūsų naudotas išankstinis duomenų apdorojimas yra pagrįstas šešerių metų kasmetinių duomenų vidurkiu. Prieš taikant PCA, duomenys buvo sustandartizuoti iki nulinės reikšmės ir vieneto standartinio nukrypimo.

Norėdami sumažinti duomenų matuojamumą, mes pritaikėme PCA metodiką. Buvo sukurti keli metodai reikiamam pagrindinių komponentų skaičiui suskaičiuoti. Po išankstinių skaičiavimų ir įvertinimų, mes nusprendėme naudoti kriterijų, kuriam reikia mažiausiai 70 % viso kitimo. Tai mums naudinga pasirenkant pirmuosius penkis komponentus. Pirmasis komponentas parodo ekonomikos, darbo rinkos, infrastruktūros ir demografinių kintamųjų mišinį. Dažniausiai jis apibrėžia bendrą regiono statusą ir apibūdina verslo aplinką bei dalinį požiūrį į kiekvieną nagrinėjamą sritį. Antrasis komponentas rodo labai aiškią struktūrą – darbo rinką ir demografines savybes. Jį apibūdina produktyvaus ir iki-produktyvaus amžiaus gyventojų skaičiaus dalis, jaunų ir ilgalaikių darbo ieškotojų dalis ir nedarbo lygis. Antrojo komponento atveju, aukšti komponento lygiai rodo prastą darbo rinkos būseną ir jos demografines savybes. Infrastruktūra ir turizmas sudaro trečią pagrindinį komponentą. Aukštos šio komponento reikšmės rodo menką regiono patrauklumą. Ketvirtojo komponento struktūra yra parodoma pirmiausia kaip demografinių kintamųjų, skirtingų verslo tipų, kuriuos apibūdina atitinkamas darbuotojų skaičius, derinys. Demografiniai veiksniai čia yra siejami su gyventojų populiacijos senėjimo rodikliais. Teigiamos komponento reikšmės išskyrė regionus, kuriuose yra senstanti miesto tipo gyventojų populiacija. Penktasis komponentas apima rodiklius, kurie nedarbo statusu pabrėžia regiono demografinę struktūrą. Aukšti šių komponentų lygiai parodė kaimiškus regionus, kuriuose yra aukštas nedarbo lygis, vyresnio amžiaus gyventojai ir skurdi infrastruktūra.

Šie pirmieji penki nurodyti komponentai vėliau buvo apdoroti panaudojant klasterio analizę. Regionai buvo nustatyti pagal minėtų komponentų savybes. Mes pritaikėme *Ward* hierarchinę klasterio analizę ir interpretavome aštuonių grupių (klasterių) sprendinį. Išsivystę rajonai, su verslui palankiausia aplinka, yra sugrupuoti atskirai. Tai patvirtino Bratislavos rajonų (Bratislava I, II ir III) sugrupavimas į vieną klasterį. Be to, kiti Bratislavos rajonai (IV ir V) bei Košice (kaip antrojo didžiausio pramoninio centro) rajonai sudaro antrąjį klasterį. Paminėtina tai, kad šis klasteris apima gerai išsivysčiusius regionus. Taip pat atkreipiame dėmesį į tai, kad Košice miestas yra rytinėje Slovakijos dalyje. Mes nusprendėme sujungti šiuos du klasterius į vieną (pažymėdami 1+2), kad vėliau galėtume panaudoti ANOVA metodą. Tokiu būdu, sukurtas klasteris apėmė devynis, geriausiai išsivysčiusių miestų, rajonų, turinčių palankiausias sąlygas įmonėms, infrastruktūrą, tiesiogines investicijas iš užsienio, kvalifikuotą darbo jėgą ir kvalifikuotus žmogiškuosius resursus, mažą nedarbą ir aukštus, realius darbo užmokesčius. Klasteris 3 apėmė rajonus, supančius Bratislavos miestą, kurie naudojosi tuo privalumu, kad yra arti didžiausio Slovakijos centro. Klasteris 4 apėmė išsivysčiusius centrus, turistinius kurortus (Poprad - PP ir Liptovský Mikuláš - LM), žemės ūkio rajonus Slovakijos pietvakariuose. Be to, jį sudarė ir penki iš aštuonių NUTS III regionų administracinių centrų. Tiesioginių investicijų iš užsienio vertė yra trečia pagal reikšmę. Be to, mes matome didžiausią kelių tinklo su ilgomis magistralėmis tankumą. Tai turėjo įtaką išsivystymo lygiui. Vargingesni Rytų Slovakijos regionai, kuriuos apibūdina mažas realusis darbo užmokestis ir labai didelis nedarbo lygis, formuoja penktąjį klasterį. Čia nedarbas yra labai susijęs su romų tautine mažuma (remiantis ankstesniais tyrimais, tik 16,6 % jų dirba pagal ilgalaikes sutartis). Tiesioginės investicijos iš užsienio visuose klasteriuose buvo žemiausios. Klasteris 6 apima neišsivysčiusius regionus dėl mažiausio gyventojų skaičiaus, mažiausio kelių tinklo tankumo ir didžiausio miškingumo. Situaciją taip pat patvirtina žemiausias tarp visų klasterių (lyginant su kitais) naujų gyvenamųjų būstų skaičius. Priešingai klasteriui 6, klasteris 7 taip pat apima žemės ūkio rajonus. Šių rajonų plotas ir gyventojų skaičius yra daug didesni negu klasteryje 6, tačiau kiti bruožai yra panašūs. Didžiausias nedarbas, kurį apibūdina ilgalaikių bedarbių skaičius, mažas realusis darbo užmokestis ir mažiausias greitkelių skaičius iš anksto šiuos rajonus priskyrė prie „silpnųjų". Antra vertus, šiuose rajonuose verslų, kuriuose dirba iki 19 darbuotojų, skaičius yra didžiausias. Tai galima paaiškinti tuo, kad dėl įdarbinimo galimybių trūkumo vyrauja smulkusis verslas.

Apibūdinus kiekvieną klasterį, toliau buvo atliekamas tyrimas pritaikant ANOVA metodą bendrai migracijai, imigracijos ir emigracijos srautams. Nulinė hipotezė (vienodi pokyčiai – kurie yra pagrindinė šio parametrinio metodo prielaida), buvo atmesta, remiantis Bartlett testu (p-reikšmės < 0.01 visų 3 kintamųjų atveju). Todėl mes pirmiausiai nusprendėme panaudoti Kruskall-Wallis testą, kuris yra atsparus ANOVA prielaidos pažeidimui. Naudodami jį, mes priėmėme alternatyvią hipotezę su 1 procento reikšmingumo lygiu visuose trijuose nagrinėtuose atvejuose. Tačiau, Kruskall-Wallis testas parodė bendrą statistinį reikšmingumą. Norėdami sužinoti migracijos tarp klasterių skirtumus, mes tęsėme testus po to, kai atlikome ANOVA. Mes atsisakėme naudoti Tukey, tikrai svarbų skirtumų testą, ir nusprendėme naudoti mažiausių kvadratų metodui atsparų Games-Howell testą. Analizuojant migracijos atvejus mes sužinojome, kad klasterio (1+2) miestų rajonai pasižymi daug stipresne migracine veikla. Tai juos išskiria iš kitų rajonų. Imigracijos atveju, abu klasteriai (1+2) ir 3 buvo nustatyti kaip statistiškai lygūs (*p=0, 72),* bet kiekvienas iš jų labai skiriasi nuo kitų (visose nagrinėtose porose 1-0 procento lygiu). Imigracinė veikla yra didelė abiejuose, nors ir dėl skirtingų priežasčių. Padidėjusią imigraciją klasteryje 1+2 sukėlė urbanizacijos procesai, kurie yra pakankamai aiškūs imigracijoje iš visų nagrinėtų klasterių, išskyrus klasterius 6 ir 8. Iš kitos pusės, imigraciją į klasterį 3 apibūdina suburbanizacijos procesas, kuris pritraukia iš Bratislavos rajonų į klasterio 3 kaimo vietoves. Emigracijos atveju, tik klasteris 1+2 statistiškai labai skiriasi nuo kitų klasterių *(*$p < 0.01$ visose nagrinėtose porose). Kitų klasterių atveju mes galime daryti išvadą, kad vieno procento lygiu tarpusavyje jie statistiškai labai nesiskiria. Šie empiriniai rezultatai vėliau buvo panaudoti mūsų pasiūlytame modelyje, pagrįstame klasterio gyventojų skaičiumi. Šis modelis, numatantis didelį indėlį į migracijos požiūrį, buvo pritaikytas norint sumodeliuoti gyventojų skaičiaus dinamiką būsimiems 30-iai metų. Panaudojus šį modelį, santykinai didelė migracijos spaudimo, apibūdinančio migraciją, reikšmė buvo pastebėta klasteryje (1+2) (teigiama reikšmė), taip pat ir klasteryje 3 (neigiama reikšmė). Modeliavimo rezultatai reiškia, kad būsimasis gyventojų skaičiaus mažėjimas klasteriuose 6, 8 gali būti laikomas kritiškiausiu. Mes svarstėme ar įmanoma pakeisti šį nepageidaujamą ateities įvykį. Mes apsvarstėme hipotetinį modelio sprendimą, kur buvo sustiprinti tam tikri pasirinkti ryšiai (aišku, modelyje), ateityje priimant tinkamus, konceptualius, regioninius sprendimus. Tiksliau kalbant, mes sustiprinome klasterių 6 ir 8 patrauklumą. Klasterių 6 ir 8 patrauklumas, kaip mes tikėjomės, taip pat daro įtaką situacijai kituose klasteriuose (didžiausia įtaka buvo pastebėta 1 klasterio atveju). Labai teigiamas rezultatas buvo tas, kad tikėtinas klasterių 6 ir 8 gyventojų skaičiaus padidėjimas buvo pasiektas be drastiško gyventojų skaičiaus sumažėjimo kituose klasteriuose. Gana nereali mūsų rezultatų ypatybė yra ta, kad numatytas populiacijos srautų nukreipimas reikalauja santykinai aukštos, perkėlimo dauginimo parametro c^T (atstovaujančio spaudimą) reikšmės. Klasterio 6 atveju, gyventojų skaičiaus praradimas yra kompensuojamas per c^T =3.2. Dar daugiau, (c^T =4.3, turėtų sustabdyti gyventojų skaičiaus mažėjimą klasteryje 8. Tolesnės optimizacijos reikės, norint pagerinti šį perspektyvų modelį ir siekiant apsaugoti problemiškus regionus nuo gyventojų skaičiaus mažėjimo.

Raktažodžiai: *migracijos srautai, pagrindiniai komponentai, klasteriai, pokyčių analizė, migracijos prognozė, migracijos spaudimas.*

The article has been reviewed.

Received in January, 2013; accepted in December, 2013.