

***Grouping Economies Based on Digitalisation Performance
via Candecom/Parafac Model and Cluster Analysis***

Hrachya Zakaryan

DOI: <https://doi.org/10.58726/27382915-2023.1-360>

Key words: *digital economy, Republic of Armenia, three-dimensional data set, extraction of latent components, agglomerative hierarchical clustering, k-means clustering*

Introduction

Digital transformation and interconnected processes cause many urgent issues and obstacles, making it crucial to reorganise public policy implementation for overcoming such challenges. Under modern developments, national economies become more interrelated, which means that it is necessary to inspect possible connections and realise the main drivers of current realities. These results cannot be achieved until the policymakers do not find to what extent the economy they are responsible for is similar (or dissimilar) to other economies in regard of digital processes.

A popular scientific method used for inspection of those circumstances is cluster analysis of countries based on digital economy indicators, different digitalisation variables, or their substitutes. Different researches are available, in which cluster analysis and other methods (depending on the targets and possible results for which those researches are held) are simultaneously used, and countries that are grouped in that works into different clusters are either European Union (EU) member states [1; 2; 8], or developed and developing economies within and outside the EU [12; 15; 18]. In two papers, the Republic of Armenia is also included. One of the articles divides 90 countries into four clusters to find the place of the Armenian IT sector in the global economy, using indicators that describe the development of the sector, the importance for the government, the tax climate, and further development preconditions [20]. Another one considers indicators relating to e-government and digital processes, using cluster analysis methods for the years 2014, 2016, and 2018, to reveal changes of cluster belongingness for 19 economies [19].

In this research, cluster analysis for 25 countries (economies from Central and South-Eastern Europe and the Black Sea Region) is conducted using 6 digital economy indicators, based on their percentage change values for 10 years (2011-2020). The aims of the article are to find out how those economies are grouped, what can be the drivers of such a division, and what are the countries in the same cluster with Armenia. The scientific novelty is the usage of the Candecom/Parafac

model for extraction of components (with equal quantity for every mode), as a three-dimensional data set is used. After these components are extracted, clustering procedures (agglomerative hierarchical and the k-means algorithms) are used, based on the values for the country mode, to divide the states into homogeneous groups.

Research Methodology and Data Description

Extraction of components. As a three-dimensional data set (countries by variables by years) is considered, the Candecomp/Parafac (CP) model is used to reduce three-way array \underline{X} of order $I \times J \times K$ by extracting the same number of components (R) for every mode. In scalar notation, the CP model can be written as:

$$x_{ijk} = \hat{x}_{ijk} + e_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk},$$

$$i = \overline{1, I}, j = \overline{1, J}, k = \overline{1, K}, r = \overline{1, R},$$

where x_{ijk} is the generic element of \underline{X} , a_{ir} are the component scores for the country mode, b_{jr} are the scores for the variable mode, c_{kr} are the scores for the year mode, and e_{ijk} is a residual term [7, 1, 4; 14, 61; 3, 152].

In terms of the frontal slices of \underline{X} , the equation can be written as:

$$X_k = AD_k B^T + E_k,$$

where X_k is the k th frontal slice ($I \times J$) of the three-way array \underline{X} , A is an $I \times R$ matrix of the component scores a_{ir} for countries, B is a $J \times R$ matrix of scores b_{jr} for variables, D_k is an $R \times R$ diagonal matrix with the k th row of C ($K \times R$, the generic element is c_{kr}) on its diagonal, and E_k is an $I \times J$ matrix of residuals [13, 59; 14, 62].

For the extraction of components, the R package *ThreeWay* (version 1.1.3) is used [5; 7]. As a *data.frame* object of order $I \times JK$ containing the matricized array (frontal slices) is used, the matrix formulation of the CP model is considered:

$$X = AH(C' \otimes B) + E,$$

where $X = [X_1 \ X_2 \ \dots \ X_k \ \dots \ X_K]$ is an $I \times JK$ matrix with X_k defined as above, and E is defined similarly. A , B , and C are the matrices introduced above, and H ($R \times R^2$) is the matricized version of the three-way ($R \times R \times R$) identity array \underline{I} (thus, there are the only ones on its diagonal, while other elements of H are equal to zero). The symbol \otimes denotes the Kronecker product [7, 2-4, 12; 14, 16, 84].

The optimal solution can be obtained by alternating least squares algorithm, which minimizes the residual sum of squares over A , B , and C :

$$\min_{A, B, C} \|X - AH(C \otimes B)\|^2,$$

where $\|\cdot\|$ is the Frobenius norm.

The algorithm fixes B and C to solve for A , then fixes A and C (for B), and then fixes A and B (for C), continuing to repeat the whole procedure, unless the convergence criterion is satisfied [11, 457, 471; 14, 111, 114].

In order to choose the appropriate number of components, a numerical convex

hull based model selection procedure is considered. The procedure numerically assesses the higher boundary (and the location of the elbow in this boundary) of the convex hull of a plot of goodness of fit g versus number of free parameters fp (which in the CP model equals to $(I + J + K)R - 2R$). The goodness-of-fit value g of the obtained solution is calculated as:

$$g = \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \hat{x}_{ijk}^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K x_{ijk}^2}.$$

For the selection of the solution l with the best balance of g and fp (after determining the subset of solutions that are on the higher boundary of the convex hull of the plot), the following expression should be maximized:

$$st_l = \frac{g_l - g_{l-1}}{fp_l - fp_{l-1}} \bigg/ \frac{g_{l+1} - g_l}{fp_{l+1} - fp_l}.$$

A relatively large st -value (scree test value) indicates that allowing for fp_l free parameters (instead of fp_{l-1} parameters) increases the fit of the model considerably, whereas allowing for more than fp_l parameters hardly increases it. Thus, that's the solution after which the increase in fit levels should be chosen [4, 137, 139, 141].

It is important to check whether solutions are degenerate or not. A typical sign of a degeneracy is that two of the components become almost identical, but with opposite sign or contribution to the model. An indication of the problem can be obtained by monitoring the correlation between all pairs of components. The triple cosine measure called Tucker's congruence coefficient (also known as uncorrected correlation coefficient) is calculated for that purpose:

$$T_{m,n} = \cos(a_m, a_n) \cos(b_m, b_n) \cos(c_m, c_n) = \frac{a'_m a_n}{\|a_m\| \|a_n\|} \frac{b'_m b_n}{\|b_m\| \|b_n\|} \frac{c'_m c_n}{\|c_m\| \|c_n\|},$$

with m and n indicating the m th and n th components, respectively. A coefficient value close to -1 is an indication of a degenerate solution [14, 107-108; 3, 160-161].

Clustering. After the components are extracted, clustering of the countries based on these components is implemented. Two methods are considered: agglomerative hierarchical clustering and the k-means clustering. As in the k-means clustering it is necessary to specify the number of clusters in advance, a common approach is applying agglomerative hierarchical clustering first, in order to determine the number of clusters.

Agglomerative hierarchical clustering method creates a hierarchical decomposition of the given set of objects. In the beginning, each object is considered as a separate cluster, then these singletons are merged into larger and larger clusters, until all of the objects are in a single cluster, or until certain termination conditions are satisfied. A tree structure called dendrogram is used to show how objects are grouped together after each step.

Let $x_i = (x_{i1}, x_{i2}, \dots, x_{iF})$ and $x_j = (x_{j1}, x_{j2}, \dots, x_{jF})$ be two objects that are described by F numeric attributes (in our case, these features are the extracted components). The Euclidean distance between the objects is defined as:

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{iF} - x_{jF})^2}.$$

However, if those features have different measurement scales, attributes with large values and variances will tend to dominate over others. This problem can be solved by normalizing the data to make each feature contribute equally to the distance. A commonly used method is data standardization, also known as z-score normalization:

$$x_{if}^s = \frac{x_{if} - \bar{x}_f}{\sigma_f},$$

$$i = 1, n, f = 1, F,$$

where x_{if}^s is the generic form of standardized values, x_{if} represents the raw data, \bar{x}_f is the sample mean, and σ_f is the sample standard deviation.

The merger of a pair of clusters or the formation of a new cluster is dependent on the definition of the distance function between two clusters. A large number of such linkage methods exists. In this research, Ward's method is used: the distance function between a cluster C_i and a new cluster C_{ij} , formed by the merger of two clusters (C_i and C_j), is calculated as follows:

$$d(C_i, (C_i, C_j)) =$$

$$= \frac{n_i + n_l}{n_i + n_j + n_l} d(C_i, C_i) + \frac{n_j + n_l}{n_i + n_j + n_l} d(C_i, C_j) - \frac{n_l}{n_i + n_j + n_l} d(C_i, C_j),$$

where $d(\cdot)$ is the distance function between clusters, and n_i, n_j, n_l are the numbers of data points, belonging to the clusters C_i, C_j, C_l , respectively [9, 114, 449, 459-460; 17, 22-23, 33-34, 36].

As a centroid-based partitioning technique, the *k-means clustering algorithm* takes the input parameter, k ($k \leq n$), and partitions a data set of n objects, D , into k clusters C_1, C_2, \dots, C_k , such that $C_i \subset D$, and $C_i \cap C_j = \emptyset$ for $1 \leq i \leq k, 1 \leq j \leq k$. It is important to get clusters with high intracluster, but low intercluster similarity. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid (centre point). The method creates an initial set of k partitions – k objects are selected (randomly, or based on some prior knowledge) as an initial cluster centres. The remaining objects are assigned to the cluster to which they are the most similar, based on the Euclidean distance between each object $p \in C_i$ and the cluster mean c_i . The next step is an iterative relocation that attempts to improve the partitioning by moving objects from one cluster to

another. The new mean for each cluster is then computed, and the process continues until the criterion function converges. The sum-of-squared-error criterion is used:

$$E = \sum_{i=1}^k \sum_{p \in C_i} [d(p, c_i)]^2,$$

where E is the sum of the squared error for all objects in the data set, and $d(p, c_i)$ is the Euclidean distance between each object $p \in C_i$ and the cluster centroid c_i . Thus, for each object in each cluster, the distance from the object to its cluster centre is squared, and the distances are summed. The purpose of this objective function is to make the resulting k clusters as compact and as separate as possible [9, 451-452; 17, 63-64, 68].

Clustering procedures are implemented via IBM SPSS Statistics 23 software [6, 271-274, 280; 10, 109-113].

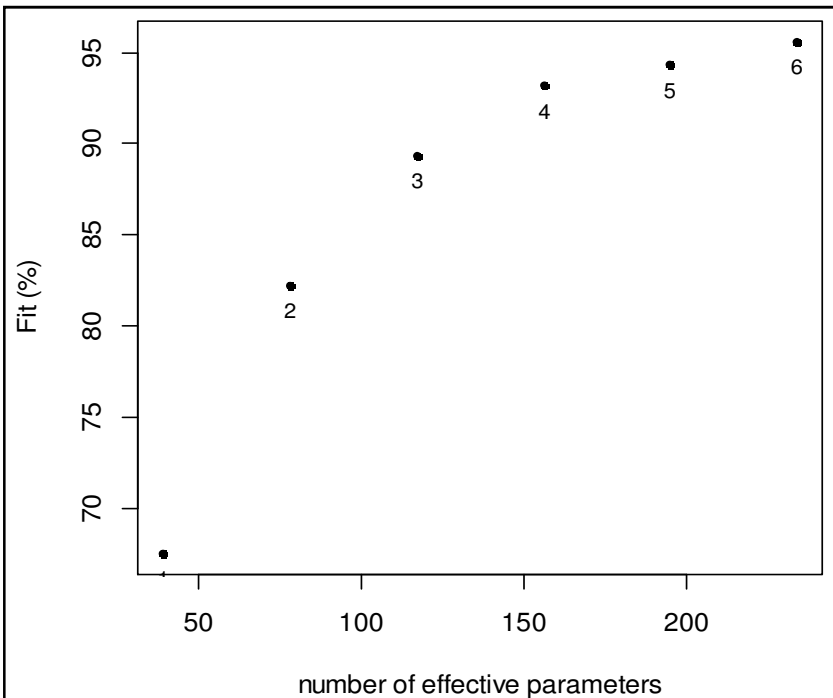
Data description. The indicators used are the number of mobile cellular subscriptions per 100 people, the number of fixed broadband subscriptions per 100 people, the number of secure internet servers per 1 million people, the percentage share of high-technology exports in manufactured exports, the number of automated teller machines (ATMs) per 100000 adults, and the number of debit cards per 1000 adults. The indicators are taken from the “World Development Indicators” database [16]. 25 countries are observed, and the values of indicators are for 2010-2020. Values of the number of fixed broadband subscriptions per 100 people and the number of mobile cellular subscriptions per 100 people for Kyrgyz Republic for the year 2020, the number of debit cards per 1000 adults for Bulgaria and Kazakhstan for the year 2010, and the percentage share of high-technology exports in manufactured exports for Ukraine for the year 2010 were not available and were generated in Microsoft Excel by the function FORECAST, based on values of 2010-2019 for the first two cases, and on the values of 2011-2020 – for the last cases.

In order to eliminate possible problems caused by differences between scales when extracting components, the percentage growth values of the indicators are used (for the number of ATMs per 100000 adults – the percentage decrease, as the growth in their number may be explained by the growing demand in the cash money, while the decrease may reflect the fact that more transactions are done via digital solutions).

Considered 25 countries are Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Estonia, Georgia, Hungary, Kazakhstan, the Kyrgyz Republic, Latvia, Lithuania, Moldova, Montenegro, North Macedonia, Poland, Romania, the Russian Federation, the Slovak Republic, Slovenia, Turkey, and Ukraine.

Results of Analysis

CP model solution. Graph 1 presents goodness of fit versus number of effective parameters plot, and Table 1 shows the model complexity selection results based on the convex hull heuristic approach. The method indicates that 4 components should be extracted. It is also shown that the analysis based on 4 components gave a fit of 93.15%. Based on the matrix of triple congruences (Table 2), it can be seen that there is no degeneracy problem. Table 3 presents the solutions for the country mode. In Table 4, component scores for the variable mode are listed in the descending order, and Table 5 is for the solutions for the year mode. In those tables, values are scaled – normalized to unit sum of squares: for example, in Table 3, the first four columns indicate that the variable and year modes' values are scaled, and this scaling is compensated in the country mode values, while the last four columns present the solutions when the country mode and any other mode values are scaled, with the compensation in the remaining one. All component scores are rounded (with two numbers after decimal point).



Graph 1. Goodness of fit versus number of effective parameters

Table 1

Model complexity selection results (based on Convex Hull procedure)

Number of components	Goodness-of-fit values (%)	Scree test values
1	67.5101	0
2	82.2316	2.39
3	89.3365	2.36
4	93.1549	4.29
5	94.3259	1.24
6	95.5773	-

Table 2

Degeneracy check results

	Component 1	Component 2	Component 3	Component 4
Component 1	1.0000	0.0005	0.0246	-0.0253
Component 2	0.0005	1.0000	-0.2059	0.0016
Component 3	0.0246	-0.2059	1.0000	0.0114
Component 4	-0.0253	0.0016	0.0114	1.0000

Table 3

Component values for the country mode (when scaling is compensated in the country mode, and when the country mode is scaled)

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.1	Comp.2	Comp.3	Comp.4
Albania	213.58	567.06	409.35	113.87	0.09	0.69	0.53	0.10
Armenia	162.45	103.58	105.69	-31.82	0.07	0.13	0.14	-0.03
Azerbaijan	261.45	-195.31	304.25	33.00	0.11	-0.24	0.39	0.03
Belarus	519.51	24.46	92.23	230.14	0.21	0.03	0.12	0.19
Bosnia and Herzegovina	503.67	13.35	-12.89	560.57	0.21	0.02	-0.02	0.47
Bulgaria	814.53	9.09	-6.73	393.49	0.34	0.01	-0.01	0.33
Croatia	560.30	-16.26	-10.75	582.07	0.23	-0.02	-0.01	0.49
Czech Republic	486.81	8.35	21.71	-126.05	0.20	0.01	0.03	-0.11
Estonia	318.91	12.82	29.72	38.72	0.13	0.02	0.04	0.03
Georgia	483.73	62.66	153.49	312.21	0.20	0.08	0.20	0.26
Hungary	432.17	16.75	3.38	45.18	0.18	0.02	0.00	0.04
Kazakhstan	598.97	10.22	61.64	116.72	0.25	0.01	0.08	0.10
Kyrgyz Republic	172.03	537.37	513.53	75.68	0.07	0.66	0.66	0.06
Latvia	285.09	-4.77	41.32	215.85	0.12	-0.01	0.05	0.18
Lithuania	354.09	20.62	44.93	51.21	0.15	0.03	0.06	0.04
Moldova	619.08	-10.97	-37.07	211.20	0.25	-0.01	-0.05	0.18
Montenegro	92.72	-0.52	89.74	33.29	0.04	0.00	0.12	0.03
North Macedonia	278.20	13.96	54.64	-131.75	0.11	0.02	0.07	-0.11
Poland	247.34	7.06	52.65	71.57	0.10	0.01	0.07	0.06
Romania	868.64	23.39	-20.14	-166.17	0.36	0.03	-0.03	-0.14
Russian Federation	365.01	33.24	88.84	60.29	0.15	0.04	0.11	0.05
Slovak Republic	257.04	0.08	65.50	9.40	0.11	0.00	0.08	0.01
Slovenia	287.37	18.65	46.07	149.47	0.12	0.02	0.06	0.13
Turkey	316.28	9.58	40.75	21.26	0.13	0.01	0.05	0.02
Ukraine	1116.65	-14.97	5.92	-491.34	0.46	-0.02	0.01	-0.41

Table 4

Component values for the variable mode in descending order (when scaling is compensated in the variable mode, and when the variable mode is scaled)

Component 1		
Secure Internet servers per 1 million people (% growth)	2429.92	1.00
Fixed broadband subscriptions per 100 people (% growth)	50.38	0.02
Number of debit cards per 1000 adults (% growth)	45.24	0.02
Mobile cellular subscriptions per 100 people (% growth)	0.34	0.00
Number of ATMs per 100000 adults (% decrease)	-16.16	-0.01
High-technology exports: % share in manufactured exports (% growth)	-16.26	-0.01
Component 2		
High-technology exports: % share in manufactured exports (% growth)	813.68	1.00
Fixed broadband subscriptions per 100 people (% growth)	21.75	0.03
Number of debit cards per 1000 adults (% growth)	10.78	0.01
Mobile cellular subscriptions per 100 people (% growth)	-3.59	0.00
Number of ATMs per 100000 adults (% decrease)	-9.93	-0.01
Secure Internet servers per 1 million people (% growth)	-72.24	-0.09
Component 3		
High-technology exports: % share in manufactured exports (% growth)	623.78	0.80
Secure Internet servers per 1 million people (% growth)	430.81	0.55
Fixed broadband subscriptions per 100 people (% growth)	149.82	0.19
Number of debit cards per 1000 adults (% growth)	83.04	0.11
Mobile cellular subscriptions per 100 people (% growth)	16.79	0.02
Number of ATMs per 100000 adults (% decrease)	-59.82	-0.08
Component 4		
Secure Internet servers per 1 million people (% growth)	1193.59	1.00
Number of debit cards per 1000 adults (% growth)	14.38	0.01
High-technology exports: % share in manufactured exports (% growth)	13.15	0.01
Fixed broadband subscriptions per 100 people (% growth)	9.23	0.01
Number of ATMs per 100000 adults (% decrease)	-0.67	0.00
Mobile cellular subscriptions per 100 people (% growth)	-0.94	0.00

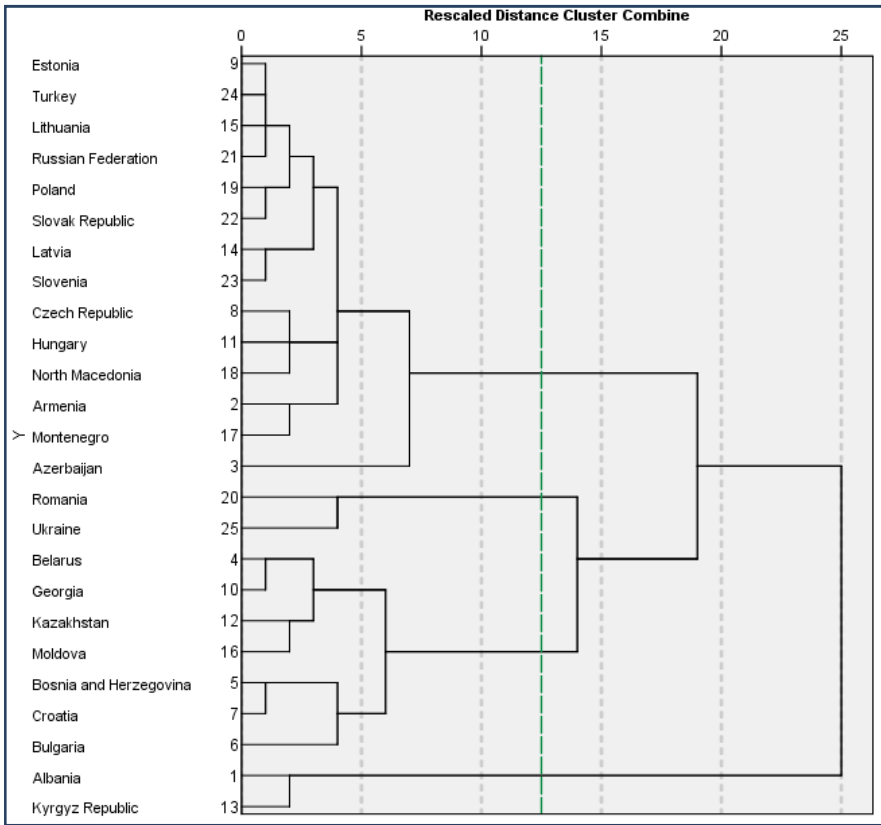
Table 5

Component values for the year mode (when scaling is compensated in the year mode, and when the year mode is scaled)

	Component values (when scaling is compensated)			
	1	2	3	4
2011	213.94	28.52	164.03	30.58
2012	351.60	-503.71	688.85	81.44
2013	123.16	-92.24	158.15	6.69
2014	164.88	-111.61	50.74	42.25
2015	203.63	603.15	215.93	-11.94
2016	2098.30	40.27	-47.74	-657.47
2017	1097.78	-75.08	9.47	991.53
2018	151.27	-121.02	72.01	-0.08
2019	139.51	45.14	123.20	13.22
2020	77.85	68.92	97.64	2.55
	Component values (when scaled)			
	1	2	3	4
2011	0.09	0.03	0.21	0.03
2012	0.14	-0.62	0.88	0.07
2013	0.05	-0.11	0.20	0.01
2014	0.07	-0.14	0.07	0.04
2015	0.08	0.74	0.28	-0.01
2016	0.86	0.05	-0.06	-0.55
2017	0.45	-0.09	0.01	0.83
2018	0.06	-0.15	0.09	0.00
2019	0.06	0.06	0.16	0.01
2020	0.03	0.08	0.13	0.00

Based on the variable mode scores, the following names for the components are proposed: *“Digital financial services”, “High-technology industry (production, exports, application in different spheres)”, “FinTech solutions”, and “Platforms, online marketplaces, and other interconnected fields”.*

Clustering results. Based on the z-scores of the country mode values of the extracted components, clustering procedures mentioned in the previous section are implemented. In the Graph 2, the dendrogram for agglomerative hierarchical clustering (using the Euclidean distance and Ward linkage) is shown. The number of clusters is proposed to be equal to 4.



Graph 2. The agglomerative hierarchical clustering dendrogram

After choosing the number of clusters, the k-means clustering procedure is implemented. The following tables (from Table 6 to Table 10) include information that is related to the k-means clustering. Cluster membership results are shown in Table 11.

Table 6

Initial cluster centres

	Cluster			
	1	2	3	4
Component 1	-1.04512	2.86328	-0.67514	1.61325
Component 2	3.06838	-0.40889	-1.54423	-0.25742
Component 3	3.21608	-0.59804	1.64358	-0.69309
Component 4	-0.08699	-2.62393	-0.27795	1.33495

Table 7

Iteration	Change in Cluster Centres			
	1	2	3	4
1	0.420	0.904	2.071	1.360
2	0.000	0.000	0.156	0.463
3	0.000	0.000	0.000	0.000

Table 8

	Cluster			
	1	2	3	4
Component 1	-0.95916	2.35020	-0.53192	0.66639
Component 2	3.16183	-0.28814	-0.29482	-0.23141
Component 3	2.82469	-0.69594	-0.11161	-0.38498
Component 4	-0.00155	-1.89650	-0.28510	1.11250

Table 9

Cluster	1	2	3	4
1		6.232	4.564	5.069
2	6.232		3.353	3.463
3	4.564	3.353		1.862
4	5.069	3.463	1.862	

Table 10

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Component 1	6.652	3	0.193	21	34.548	0.000
Component 2	7.251	3	0.107	21	67.739	0.000
Component 3	6.046	3	0.279	21	21.660	0.000
Component 4	5.665	3	0.334	21	16.983	0.000

Table 11 reveals that 2 clusters each contain 2 countries (Albania and the Kyrgyz Republic are in the 1st cluster, Romania and Ukraine – in the 2nd cluster). The 3rd cluster has 14 members (Armenia, Azerbaijan, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Montenegro, North Macedonia, Poland, the Russian Federation, the Slovak Republic, Slovenia, and Turkey), and the remaining 7 states

are in the 4th cluster (Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Georgia, Kazakhstan, and Moldova). It is interesting that in both hierarchical clustering and the k-means clustering procedures countries are grouped into the same clusters (results are shown in Graph 2 and Table 11, respectively), which means that results are stable.

Table 11

Cluster membership results after the k-means clustering procedure

Country	Nº	Distance	Country	Nº	Distance
Albania	1	0.420	Kyrgyz Republic	1	0.420
Armenia	3	0.924	Latvia	3	0.857
Azerbaijan	3	2.159	Lithuania	3	0.339
Belarus	4	0.727	Moldova	4	0.826
Bosnia and Herzegovina	4	1.087	Montenegro	3	0.854
Bulgaria	4	1.021	North Macedonia	3	0.747
Croatia	4	1.139	Poland	3	0.303
Czech Republic	3	1.120	Romania	2	0.904
Estonia	3	0.329	Russian Federation	3	0.391
Georgia	4	1.047	Slovak Republic	3	0.194
Hungary	3	0.764	Slovenia	3	0.569
Kazakhstan	4	1.038	Turkey	3	0.247
			Ukraine	2	0.904

Conclusions

Clustering results reveal that the Republic of Armenia is in the 3rd cluster, which is the biggest cluster and consists of 14 economies. 13 of them have a common – post-socialist nature (this “subgroup” consists of 6 post-soviet countries and 7 other post-socialist economies). In the context of regional trade blocs, this is the same cluster for 2 members of the Eurasian Economic Union (Armenia and the Russian Federation), 8 states that are members of the EU (the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, the Slovak Republic, Slovenia), and 3 candidate countries for membership of the EU (Montenegro, North Macedonia, and Turkey). Thus, in order to implement public policy that will develop digital economy in Armenia with the necessary outcomes, it is crucial to see how the countries included in this cluster make it homogeneous, what problems are similar in these economies, and what are the best practices of solving them. Further research might shed light on the processes that have their influence on the phenomena, considered as latent indicators of digital transformation (*digital financial services, high-technology industry, FinTech solutions, and platforms, online marketplaces, etc.*).

References

1. Bánhidi Z., Dobos I., Nemeslaki A., What the overall Digital Economy and Society Index reveals: A statistical analysis of the DESI EU28 dimensions // *Regional Statistics*. 2020, Volume 10, No. 2, pp. 42-62.
2. Bilozubenko V., Yatchuk O., Wolanin E., Serediuk T., Korneyev M., Comparison of the digital economy development parameters in the EU countries in the context of bridging the digital divide // *Problems and Perspectives in Management*. 2020, 18(2), pp. 206-218.
3. Bro R., PARAFAC. Tutorial and applications // *Chemometrics and Intelligent Laboratory Systems*. 1997, Volume 38, Issue 2, pp. 149-171.
4. Ceulemans E., Kiers H.A.L., Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method // *British Journal of Mathematical and Statistical Psychology*. 2006 (59), pp. 133–150.
5. Del Ferraro M.A., Kiers H.A.L., Giordani P., Package ‘ThreeWay’. October 12, 2022 (first published: September 7, 2015), 81 p., // URL: <https://cran.r-project.org/web/packages/ThreeWay/ThreeWay.pdf> (accessed: 03.04.2023).
6. George D., Mallery P., *IBM SPSS Statistics 23 Step by Step: A Simple Guide and Reference* (14th edition). Taylor & Francis, 2016, 382 p.
7. Giordani P., Kiers H.A.L., Del Ferraro M.A., Three-Way Component Analysis Using the R Package ThreeWay // *Journal of Statistical Software*. April 2014, Volume 57, Issue 7, pp. 1-23, // URL: <https://www.jstatsoft.org/article/view/v057i07> (accessed: 03.04.2023).
8. Greco L., Maresca P., Caja J., Big Data and Advanced Analytics in Industry 4.0: A comparative analysis across the European Union // *Procedia Manufacturing*. 41 (8th Manufacturing Engineering Society International Conference), Madrid, Spain, 2019, 1156 p., pp. 383-390.
9. Han J., Kamber M., Pei J., *Data Mining: Concepts and Techniques* (3rd edition). Elsevier, USA, 2012, 703 p.
10. *IBM SPSS Statistics Base 23 (Manual)*. IBM, USA, 202 p.
11. Kolda T.G., Bader B.W., Tensor Decompositions and Applications // *SIAM Review*. 2009, Volume 51, Issue 3, pp. 455-500.
12. Konstantinova L.A., Kramarenko I.V., Formation of Profiles of Digital Development of the Modern World Countries // *Proceedings of the International Scientific Conference “Smart Nations: Global Trends in the Digital Economy”*. Springer, Cham, Switzerland, 2022, Volume 1, 557 p., pp. 25-32.

13. Kroonenberg P.M., Applied Multiway Data Analysis. John Wiley & Sons, Hoboken, NJ, USA, 2008, 608 p.
14. Smilde A., Bro R., Geladi P., Multi-way Analysis with Applications in the Chemical Sciences. John Wiley & Sons, Chichester, United Kingdom, 2004, 396 p.
15. Tokmergenova M., Bánhidi Z., Dobos I., Analysis of I-DESI dimensions of the digital economy development of the Russian Federation and EU-28 using multivariate statistics // St Petersburg University Journal of Economic Studies. 2021, Volume 37, Issue 2, pp. 189-204.
16. World Development Indicators: The World Bank DataBank, // URL: <https://databank.worldbank.org/source/world-development-indicators> (accessed: 03.04.2023).
17. Xu R., Wunsch D.C., Clustering. John Wiley & Sons, Hoboken, NJ, USA, 2009, 358 p.
18. Zoroja J., Klopotan I., Stjepić A.-M., Quality of e-commerce practices in European enterprises: Cluster analysis approach // Interdisciplinary Description of Complex Systems. 2020, 18(2-B), pp. 312-326.
19. Խուրդաթյան Վ., Հայաստանի Հանրապետության էլեկտրոնային կառավարման համակարգի համեմատական բնութագիրը կլաստերային վերլուծության մեթոդով // Այլընտրանք. 2020 (հուլիս-սեպտեմբեր), էջ 257-266:
20. Սարգսյան Ա., ՀՀ ՏՏ ոլորտի համեմատական նկարագիրը քլաստերային վերլուծության միջոցով // ՀՀ տնտեսության զարգացման առաջնահերթությունները. ներքին և արտաքին մարտահրավերներ (ՀՊՏՀ 28-րդ գիտաժողովի նյութեր). Երևան, ՀՊՏՀ «Տնտեսագետ» հրատարակչություն, 2021, 760 էջ, էջ 610-619:

**Տնտեսությունների խմբավորում թվայնացման կատարողականի հիման վրա
Candecomp/Parafac մոդելի և կլաստերային վերլուծության միջոցով**

Հրայա Ջաթարյան

Ամփոփում

Հանգուցային բառեր. թվային տնտեսություն, Հայաստանի Հանրապետություն, եռաչափ տվյալների հավաքածու, լատենտ բաղադրիչների դուրսբերում, ազլոմերատիվ հիերարխիկ կլաստերացում, k-միջիններով կլաստերացում

Թվային փոխակերպումը և փոխկապակցված գործընթացներն առաջացնում են բազմաթիվ հրատապ խնդիրներ ու խոչընդոտներ՝ կարևոր դարձնելով հանրային քաղաքականության իրականացման վերակազմակերպումը նման մարտահրավերների հաղթահարման համար: Ազգային տնտեսությունները դառնում են առավել փոխկապակցված, ինչը նշանակում է, որ անհրաժեշտ է ստուգել տարբեր տնտեսությունների միջև հնարավոր կապերը և գիտակցել ներկայիս իրողությունների հիմնական շարժիչ ուժերը: Վերակազմակերպման արդյունքներին հնարավոր չէ հասնել, քանի դեռ քաղաքականություն մշակողները չեն պարզել, թե որքանով է տնտեսությունը, որի համար իրենք պատասխանատու են, նման այլ տնտեսություններին թվային գործընթացների առումով, քանի որ թվայնացման կոնտեքստում նման արդյունքներն առավել քան արժեքավոր են:

Այդ հանգամանքների ստուգման համար լայնորեն օգտագործվող գիտական մեթոդ է երկրների կլաստերային վերլուծությունը՝ հիմնված թվային տնտեսության ցուցանիշների, թվայնացման տարբեր փոփոխականների կամ դրանց փոխարինող ցուցիչների վրա: Տարբեր հետազոտություններում միաժամանակ օգտագործվում են կլաստերային վերլուծությունը և (նպատակներից և հնարավոր արդյունքներից կախված) այլ մեթոդներ, և երկրները, որոնք այդ աշխատանքներում խմբավորված են տարբեր կլաստերների ներքո. կա՛մ Եվրոպական միության անդամ պետություններ են, կա՛մ զարգացած և զարգացող տնտեսություններ (Եվրոպական միության անդամ և ոչ անդամ երկրներ): Հայաստանի Հանրապետությունը դիտարկող աշխատանքները սակավաթիվ են և նպատակ են ունեցել պարզելու համաշխարհային տնտեսության մեջ Հայաստանի տեղեկատվական տեխնոլոգիաների ոլորտի տեղի, նաև էլեկտրոնային կառավարման և թվային գործընթացների համար կլաստերային պատկանելությունը և վերջինիս փոփոխությունները:

Սույն հետազոտությունում իրականացված է 25 երկրների (ներառյալ Հայաստանի Հանրապետությունը) կլաստերային վերլուծություն՝ օգտագործելով թվային տնտեսության 6 ցուցանիշներ՝ հիմնվելով 2011-2020 թվականների համար դրանց տոկոսային փոփոխության արժեքների վրա: Նպատակն է պարզելը, թե ինչպես են խմբավորվում այդ տնտեսությունները, ինչպես

նաև նման բաժանման շարժիչ ուժերը: Գիտական նորույթը Candecomp/Parafac մոդելի օգտագործումն է եռաչափ տվյալների հավաքածուի յուրաքանչյուր չափման համար հավասար քանակությամբ լատենտ բաղադրիչներ ստանալու նպատակով: Վերջիններիս դուրսբերումից հետո օգտագործվում են ազդուերատիվ հիերարխիկ և k-միջիններով կլաստերացման ալգորիթմներ՝ հիմնված երկրների չափման արժեքների վրա:

Հայաստանում թվային տնտեսության զարգացմանը նպաստող և անհրաժեշտ արդյունքներն ապահովող քաղաքականություն կիրարկելու համար առանցքային է ուսումնասիրել, թե ինչպես են Հայաստանի հետ նույն կլաստերում ընդգրկված երկրները դարձնում այն համասեռ, ինչպիսի ընդհանուր խնդիրներ են բնորոշ այս տնտեսություններին, և որն է դրանց լուծմանը միտված լավագույն փորձը:

Группирование экономик на основе результатов цифровизации с помощью модели Candecomp/Parafac и кластерного анализа

Грачи́я Закаря́н

Резюме

***Ключевые слова:** цифровая экономика, Республика Армения, трехмерный набор данных, извлечение латентных компонентов, агрегативная иерархическая кластеризация, кластеризация методом k-средних*

Цифровое преобразование и взаимосвязанные процессы вызывают множество неотложных проблем и препятствий, что делает крайне важной реорганизацию реализации государственной политики для преодоления таких вызовов. Национальные экономики становятся более взаимосвязанными, а это значит, что необходимо проверить возможные связи между различными экономиками и осознать основные движущие силы современных реалий. Результаты реорганизации не могут быть достигнуты до тех пор, пока политики не узнают, насколько экономика, за которую они ответственны, похожа на другие экономики с точки зрения цифровых процессов, поскольку в контексте цифровизации такие результаты еще более ценны.

Популярным научным методом изучения этих обстоятельств является кластерный анализ стран на основе показателей цифровой экономики, различных переменных цифровизации или их заменителей. В различных исследованиях одновременно используются кластерный анализ и (в зависимости от целей и возможных результатов) другие методы, а страны, сгруппированные в этих работах в разные кластеры, являются либо государствами-членами Европейского союза, либо развитыми и развивающимися странами внутри и за пределами ЕС. Работы, рассматривающие Республику Армения, немногочис-

ленны и были направлены на выявление кластерной принадлежности и изменений последней для места сектора информационных технологий Армении в мировой экономике, а также для электронного правительства и цифровых процессов.

В данном исследовании проведен кластерный анализ для 25 стран, включая Республику Армения, с использованием 6 показателей цифровой экономики, исходя из значений их процентного изменения за 2011-2020 годы. Цель состоит в том, чтобы выяснить, как эти экономики группируются, а также движущие силы такого разделения. Научная новизна заключается в использовании модели Candecom/Parafac для извлечения латентных компонентов с одинаковым количеством для каждого измерения трехмерного набора данных. После извлечения, на основе значений для измерения стран, используются агломеративный иерархический алгоритм и алгоритм кластеризации методом k-средних.

Для реализации политики, способствующей развитию цифровой экономики в Армении и обеспечивающей необходимые результаты, важно изучить, как страны, входящие в один кластер с Арменией, делают этот кластер однородным, какие общие проблемы характерны для этих экономик, и какова наилучшая для их решения практика.

Ներկայացվել է 06.04.2023 թ.
Գրախոսվել է 29.04.2023 թ.
Ընդունվել է տպագրության 25.05.2023 թ.