ИННОВАЦИИ И ЦИФРОВАЯ ЭКОНОМИКА

JEL: O33, F68, H70

Analysis of I-DESI dimensions of the digital economy development of the Russian Federation and EU-28 using multivariate statistics

M. Tokmergenova, Z. Bánhidi, I. Dobos

Budapest University of Technology and Economics,
2, Magyar Tudósok krt., Budapest, 1117, Hungary

For citation: Tokmergenova M., Bánhidi Z., Dobos I. (2021) 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*, vol. 37, iss. 2, pp. 189–204.
https://doi.org/10.21638/spbu05.2021.201

This paper continues the authors' earlier analysis, in which we used five principle dimensions of the International Digital Economy and Social Index (I-DESI) for the 28 countries of the EU and the Russian Federation to examine how Russia's development relates to that of other EU countries. The aim of this paper is not to establish a ranking, but to determine the relationship between each dimension and the groups into which these 29 countries can be divided by multivariate statistical analysis tools and to analyze the group to which the Russian Federation belongs. We use Principal Component Analysis (PCA) to map our data to a lower-dimensional space (revealing two latent dimensions), and to analyze causal relations between the principal dimensions using partial correlation coefficients, concluding that two of the five main dimensions can be explained by three independent dimensions. Thereafter, we use cluster analysis to group our objects (i. e. the 29 countries) into clusters, and multidimensional scaling (MDS) to visualize the location of these groups and countries on the plane of the two components (from the PCA), focusing on the Russian Federation. According to our results, the Russian Federation can be classified as a moderately developed country in terms of its I-DESI score, but its location on the plane of principle components differs from the group of moderately developed EU-countries, forming a separate "group" on its own, largely owing to the unique characteristics of the country's digital development.

Keywords: digital transformation measurement, DESI index, connectivity, human capital correlation analysis, principal component analysis, partial correlation analysis, cluster analysis, multidimensional scaling, Russia.

Introduction

In an earlier paper in this journal [Bánhidi, Dobos, Nemeslaki, 2019], we compared the digital development of the EU-28 and the Russian Federation. The comparison was performed using three methods:

1) a scoring model with the weights proposed by the EU Commission,
2) the method of multidimensional scaling (to one dimension), and
3) Data Envelopment Analysis (DEA) and its extensions (common weights analysis).

Rankings obtained with DEA and multidimensional scaling were very similar to each other and the ranking using the original DESI weights (as evidenced by the fairly strong correlations between them), indicating their robustness. This may also mean that weights for DESI do not significantly affect the order of countries. In our calculations, Russia was part of the last third of EU countries in digital development, although their ranking showed marked variation.

The former methods are known from decision theory. However, they do not provide information on how each dimension of the International Digital Economy and Social Index (I-DESI) compares statistically to the other dimensions, i. e. what is the relationship between the dimensions as variables. In this paper, we continue to work on a stochastic analysis of these dimensions/variables.

Such a statistical analysis was performed by Bánhidi, Dobos and Nemeslaki [Bánhidi, Dobos, Nemeslaki, 2020], but only with the help of a system of statistical indicators developed for the EU-28 countries.

In addition to the EU-28 DESI data, the European Commission biennially publishes a report built around a system of indicators called the International Digital and Social Index (I-DESI), which extends DESI's geographical coverage using a simplified index to assess the performance of both individual EU countries and the EU as a whole in comparison to 17 other advanced economies (including the Russian Federation). Both composite indices are composed of five principal policy areas, which comprise a variable number of subindices and individual indicators:

1) Connectivity: Fixed Broadband Coverage, Fixed Broadband Take-up, Mobile Broadband Take-up, 4G Coverage, NGA Coverage, Subscriptions to Fast Broadband, Fixed Broadband Price;
2) Human Capital: Internet Users, Regular Internet Users, Employed in knowledge-intensive activities, Tertiary Graduates in ICT;
3) Use of Internet: News, Social Networks, Banking Internet shoppers as a percentage of population, Average number of devices used;
4) Integration of Digital Technology: Availability of latest technologies, Firm-level technology absorption, Social Media, Business to Business Internet use, Secure Internet Servers;
5) Digital Public Services: E-Government Development Index, Online Service Completion, Open Data[1].

The latest edition of I-DESI consists of the same five main dimensions as DESI and 24 individual indicators and utilizes various datasets (aggregating data from the public

[1] International Digital Economy and Society Index 2018. URL: https://ec.europa.eu/digital-single-market/en/news/international-digital-economy-and-society-index-2018 (accessed: 05.06.2019).

databases of the World Economic Forum, OECD, World Bank and ITU etc.) over a four-year time period from 2013 to 2016.

In our earlier paper [Bánhidi, Dobos, Nemeslaki, 2019], we used a dataset comprising scores in these five principal dimensions for the 28 EU countries and the Russian Federation, compiled from the I-DESI report. We use the same dataset for this paper, which is organized as follows. In the second section, we present a short literature review on the DESI and the digital development of Russia. In the third section, we use multivariate statistical analysis to answer five research questions, examining the relationships between the five dimensions and grouping our 29 countries according to the level and characteristics of their digital development. First, we analyze the correlation matrix, then we use Principal Component Analysis (PCA) to perform a mapping of our data to a lower-dimensional space (revealing two latent dimensions), and then map the causal relationships between the five principal dimensions based on the partial correlation coefficients. Thereafter, we use cluster analysis to group our objects (i.e. the 29 countries) into clusters, and multidimensional scaling (MDS) to visualize the location of these groups and countries on the plane of the two components (from the PCA), focusing on the Russian Federation. The last, fourth section of the paper provides conclusions.

1. Literature review

The digital economy and its influence on national competitiveness on the examples of Switzerland, Russia, and Azerbaijan is analyzed by Miethlich and others [Miethlich et al., 2020]. Research revealed that Switzerland and Russia are good in IT education services, but Russia has a weakness in protecting intellectual property rights. A cluster analysis was implemented based on data of exporting telecommunications and computer and information (TCI) services. This analysis reveals that Russia is placed in the third cluster, which means that does not aim to export TCI. The development of the digital economy in Russia and 28 European Union countries using I-DESI database was compared by Banhidi, Dobos and Nemeslaki [Bánhidi, Dobos, Nemeslaki, 2019]. According to the results Russia is strong in the dimension of Human Capital.

Ermolaev, Trubetskaya and others [Ermolaev et al., 2018] assessed the development of Russia's digital economy based on international indexes. The results show the necessity to develop infrastructure to increase the percentage of Internet users, create digital infrastructure, and provide accessibility to ICT technologies. Revinova and Lazanyuk [Revinova, Lazanyuk, 2018] researched the level of digitalization in Russia's regions. The difference in the development of digital technologies varies highly in regions; among the Federal Districts the leading position belongs to Northwestern and among subjects to Moscow and St. Petersburg. The reason of this lagging between regions is lack of digital infrastructure and funding.

Stavytskyy, Kharlamova and Stoica [Stavytskyy, Kharlamova, Stoica, 2019] analyzed three hypotheses and revealed that a high level of consumption and low unemployment could be factors increasing the score of the DESI index. For countries not present in DESI, such as Ukraine, a number of steps are needed to increase the value of the DESI score.

Moroz [Moroz, 2017] presented research using two indices, NRI (Networked Readiness Index) and DESI (Digital Economy and Society Index), to compare the digital development of Poland in comparison with other EU countries. The results of both indexes are similar, although an analysis of NRI presents Poland as relatively good, while an analysis of the DESI index demonstrates a lag in development of Poland's digital economy in comparison to those of EU countries.

Risks of digitalization of Russian industry, in terms of lack of technologies, equipment, and software products, were presented by Korovin [Korovin, 2018]. Labor productivity is also crucial in achieving leading positions in Russian industrial digitalization. Statistical research demonstrates a positive trend in increasing the number of university graduates in such specialties as automation, IT, and communications, but the demand for these specialists in Russian industry remains low.

Baskakova and Soboleva [Baskakova, Soboleva, 2019] analyzed functional illiteracy in Russia based on access to the Internet and the level of computer literacy. Research shows that older adults, those with less educated, people with a low income, and the rural population are in a risk group of functional illiteracy. The regional factor for Russia also contributes to the unequal development of its digital economy.

Mironova, Bogdanova and Kolesnikov [Mironova, Bogdanova, Kolesnikov, 2019] argue that digital education and digital literacy in Russia as the main factors shaping the development of the economy and society. The difference between generations for development should be considered in the transition to digitization.

Kuvayeva [Kuvayeva, 2019] assesses Russia's readiness for digital integration. The author mentions the lack of a unified statistical measure for assessing digital readiness of all countries, including Russia. Analyzed dimensions, such as investments in technologies, high-tech industry development, and readiness to digital transitions, are low in comparison with developed countries.

In analyzing the relationship between the development of digital economy and industry, Vishnevsky's [Vishnevsky, 2019] statistical analysis showed that the digital economy is dependent on fixed assets investment. Investments in technologies increase the level of income and economic development. Finally, Akberdina [Akberdina, 2018] indicates that industrial digitalization is impossible without a developed industrial sector, and the level of digitalization and automation determines the degree of using high technologies in industry. Certain regions in Russia are less developed, due to historical factors, and the author proves that concentration of high technologies shapes differentiation in digital development.

2. Research questions and methodology of statistical analysis

A multivariate statistical analysis of I-DESI dimensions is presented using two analytic techniques. One group of methods explores the relationship between the five principal dimensions (as variables). Three analyses were performed for this. On the one hand, we investigate the correlation matrix of dimensions. We then use principal component analysis, which divides the variables into groups according to the strength of the linear relationship between them. The principal component analysis starts from the correlation matrix between the variables and returns its variance as much as possible. On the other hand, we map the causal relationships between the variables using partial correlation between the variables. In the latter case, we only explore the causal relationships, but no longer the direction of the causal relationship, because the method does not allow this.

The other method of analysis in the space of variables, i. e. dimensions, analyzes how objects, in our case countries, are arranged. There are two methods for this as well. Using cluster analysis, groups can be formed between countries using a defined distance measure. The other method is multidimensional scaling (MDS), which starts from the basic assumption that we have determined the distances between our objects (i. e. the 29 countries) and mapped them into a lower dimensional space in which these distances are kept where possible.

We test three hypotheses using the mathematical-statistical methods described above. Among our hypotheses, we divide two into two sub-hypotheses. The first hypothesis explores the probabilistic linear relationship between the digital dimensions. This hypothesis is presented using two sub-hypotheses (*H1a*. and *H1b.*).

H1a. There is a strong linear relationship between the digital dimensions, i. e., the dimensions contain redundant information.

Hypothesis *H1a*. also draws attention to the fact that it is likely that a reduction in digital dimension numbers better expresses the relationships between dimensions. Hypothesis *H1b.* draws attention to this question.

H1b. It is sufficient to express the information content of the five digital dimensions with two latent variables, i. e. components.

After exploring the linear relationships between the variables, we undertake to map the causal relationships between them. This causal relationship system is described in hypothesis H2.

H2. The digital dimensions can be divided into two causal groups using partial correlations.

Sub-hypotheses *H3a.* and *H3b.* explore the place of the Russian Federation in digital terms among the countries of the European Union *H3a.* and *H3b.* For this we use grouping procedures.

H3a. The Russian Federation bears a resemblance to the countries of Central and Eastern Europe and Southern Europe in terms of digital development.

The analysis also draws attention to the fact that Russia's digital development differs significantly from countries of the European Union, which calls attention to the fact that it itself forms a separate group in the examined group of countries, leading to sub-hypothesis *H3b*.

H3b. In terms of digital development, the Russian Federation is at the same distance from the developed and less developed countries of the Union.

Hypotheses *H1a.* with question RQ1, *H1b.* with RQ2, and *H2.* hypothesis are tested with question RQ3. Sub-hypotheses *H3a.* and *H3b.* are verified by analysis of questions RQ4 and RQ5. With the five analysis methods we want to use, we can formulate five research questions (RQ1–RQ5). These are the following.

RQ1. What are the linear relationships between the DESI dimensions? Do the digital dimensions measure different variables?

RQ2. How can the information content of I-DESI dimensions be reduced? What new latent variables (components) are created and what do they mean?

RQ3. What causal relationship can be revealed between the I-DESI dimensions? Which dimensions can be interpreted as causes and which as consequences?

RQ4. What clusters/groups can the 29 countries we are examining be divided into, and where is the Russian Federation among them?

RQ5. Based on the five-dimensional space of the I-DESI dimensions, how can countries be represented if, while maintaining distances, we map the countries into two dimensions, i. e. the plane, and where is Russia located?

Hypotheses *H1a.* is tested with question RQ1, *H1b.* with RQ2, and *H2.* with question RQ3. Sub-hypotheses *H3a.* and *H3b.* are verified by analysis of questions RQ4 and RQ5. The basic data used in SPSS 26 is listed in the Appendix.

2.1. RQ1: Exploring linear relationships with Correlational matrix

Table 1 shows moderate to strong correlations between the digital dimensions. Each of these correlation coefficients has a positive sign, suggesting that the movement of variables goes to one direction. We consider this to be a very important consequence to any further use of the DESI dimension, in which case the dilemma arises of how orthogonal, i. e. collinear, these dimensions are in terms of our understanding of the digital economy and society. In other words, how does each of these dimensions add value to our assessment of the status in the European Union and Russia regarding the digital transition? A high correlation could be alarming, pointing towards little additional value of some variables in the DESI model, and so among the dimensions there is multicollinearity.

Table 1. **Correlation matrix between DESI dimensions**

		Human Capital (*HC*)	**Use of Internet (*UI*)**	**Integration of Digital Technology (*IT*)**	**Digital Public Services (*PS*)**
Connectivity (*CN*)	Pearson Correlation	.492**	.773**	.699**	.454*
	Sig. (2-tailed)	0.007	0.000	0.000	0.013
Human Capital (*HC*)	Pearson Correlation		.753**	.665**	.501**
	Sig. (2-tailed)		0.000	0.000	0.006
Use of Internet (*UI*)	Pearson Correlation			.826**	.594**
	Sig. (2-tailed)			0.000	0.001
Integration of Digital Technology (*IT*)	Pearson Correlation				.681**
	Sig. (2-tailed)				0.000

Notes: ** — correlation is significant at the 0.01 level (2-tailed); * — correlation is significant at the 0.05 level (2-tailed).

Of the correlations, the highest correlation is between Use of the Internet and Integration of Digital Technology, 0.826. This means that there is probably some causal relationship between the two variables/dimensions. We calculated 0.773 and 0.753 correlation coefficients between Use of Internet, and Connectivity and Human Capital, respectively. These two high correlations may suggest that Use of the Internet may have some causal relationship to the two dimensions mentioned, namely Connectivity and Human Capital. However, this may also mean that the consequence of the latter two dimensions is the Use of Internet, which can be decided by examining the partial correlation. The Integration of Digital Technology dimension shows a strong medium correlation with the Connectivity, Human Capital, and Digital Public Services variables. The values of the correlations are 0.699, 0.665, and 0.681, respectively. At this time we do not go into the exploration of causal chains, which we do when determining the partial correlation. The other correlation coefficients show a weaker correlation, so we will not go into that.

In summary, we can conclude that correlation analysis suggests that the analysis revealed a strong linear relationship between the digital dimensions. This may also suggest that the number of variables can be significantly reduced by latent variables. To get around this, we will cover it in the next section.

2.2. RQ2: Search for latent variables with Principal component analysis

The studies that can be performed in principal component analysis start from the correlation matrix between variables, in our case between dimensions. In the method, we try to generate the correlation matrix using the eigenvalues and eigenvectors of the correlation matrix. To do this, eigenvalues are arranged in descending order of magnitude. The eigenvalues of the correlation matrix show the amount of information found in the correlation matrix, i. e. the proportion of variance. In an acceptable principal component analysis model, it is not necessary to return the whole of the variance with the eigenvectors; it is sufficient to produce two-thirds of it. The component matrix thus produced shows which variables show the greatest correlation with which components. It also shows what groups the digital dimensions can be grouped into, that is, what are the highly correlated variables. To make the principal components more orthogonal, we used the rotation method, more specifically, the Varimax rotation in this study. The goodness of such principal component models can be determined by the Kaiser — Mayer — Olkin measure.

Principal component analysis shows how the number of dimensions, which are strongly correlated, could be reduced. For this analysis, the principal components method with Varimax rotation and without rotation is used to determine linear relationships between the I-DESI dimensions.

The Kaiser — Mayer — Olkin measure of sampling adequacy shows that with the 0.812 value, the model is meritorious, which confirms it to be significant. Bartlett's test of sphericity is also significant, proving that the results of the model are acceptable. The results with components are presented in Table 2 (extraction method: Principal Component Analysis; rotation method: Varimax with Kaiser Normalization).

Table 2. **Rotated Component Matrix**

	Component	
	1	**2**
Connectivity	.908	.149
Use of Internet	.843	.444
Integration of Digital Technology	.703	.595
Human Capital	.625	.509
Digital Public Services	.233	.931

The communalities values of the analysis are between 0.8 and 0.9, which means that estimated components preserved a large proportion of the variance in the original dataset. In this analysis two components are considered to be sufficient and they have returned 83.459 % of variance. In the principal components with rotation, the value of first component is 49.476 %, and the second component gives 33.983 % of variance back. The first four weights of the first component are above 0.625, which means that each variable is strongly correlated with this component, except Digital Public Services, which shows weak correlation with 0.233 value. The second component without rotation explained 11.463 % of variance and with rotation all variables are strongly correlated except one variable, Connectivity, with a value of 0.149. Method with Varimax rotation shows more specified results representing 83.459 % of variance on two components.

Before turning to the meaning of the two components, note that three dimensions, namely Human Capital, Use of Internet, and Integration of Digital Technology, correlate strongly with both components, i. e. they explain it to some extent. The first component, Connectivity, and the second component, Digital Public Services, show an extremely strong correlation, meaning that those two variables can only be assigned to one of these components. Based on these observations, the first component can be interpreted as preparedness for the digital economy, as it includes Connectivity, which can be interpreted as technical readiness, as well as Human Capital, which also presents the educational level. The second component can measure government digital preparedness along the Digital Public Services dimension. Because the Use of Internet and Integration of Digital Technology dimensions are strongly correlated with both components, they can be considered outcome variables rather than inputs, as evidenced by the partial correlation coefficients.

The vectors of the two components are also used for multidimensional scaling. These compressed variables are chosen to describe the location of the countries.

2.3. RQ3: Exploring causal relationships with Analysis of partial correlations

Partial correlations were calculated by filtering out the effects of the other three of the five dimensions of I-DESI. The partial correlation matrix is shown in Table 3. Each of the partial correlations marked in grey is significant at a level of at least 6 %. The other correlations, however, are not significant, so they can be considered zero.

After filtering out background effects, the Connectivity and Human Capital dimensions have the highest partial correlation with Use of Internet. These two correlations show values above 0.500. However, there is a medium partial correlation between Use of Internet and Integration of Digital Technology, and between Integration of Digital Technology and Digital Public Services. These logical observations can be summarized in Fig. 1. The figure confirms the result of a multicollinearity analysis. The Connectivity and Human Capital dimensions explain Use of Internet, while Integration of Digital Technology is explained by the Digital Public Services dimension. There is a medium partial relationship between Use of Internet and Integration of Digital Technology.

The result thus obtained can also be interpreted as meaning that two of the five digital dimensions, namely Use of Internet and Integration of Digital Technology, can be expressed by the other three, i. e. Connectivity, Human Capital, and Digital Public Services. The linear regression of the two dimensions (Use of Internet and Integration of Digital

Table 3. **Partial correlation matrix between the I-DESI dimensions**

		Human Capital (*HC*)	Use of Internet (*UI*)	Integration of Digital Technology (*IT*)	Digital Public Services (*PS*)
Connectivity (*CN*)	Pearson Correlation	−.237	.529**	.214	−.079
	Sig. (2-tailed)	0.244	0.005	0.293	0.702
Human Capital (*HC*)	Pearson Correlation		.521**	.128	.038
	Sig. (2-tailed)		0.006	0.535	0.854
Use of Internet (*UI*)	Pearson Correlation			.373	.074
	Sig. (2-tailed)			0.060	0.718
Integration of Digital Technology (*IT*)	Pearson Correlation				.419*
	Sig. (2-tailed)				0.033

Notes: ** — correlation is significant at the 0.01 level (2-tailed); * — correlation is significant at the 0.05 level (2-tailed).

Based on the grey partial correlations in Table 3, causal relationships between the variables can be mapped. As could already be seen from the components of principal component analysis, at least three "independent variables" can be identified. These dimensions are Connectivity, Human Capital, and Digital Public Services. The "dependent variables" are Use of Internet and Integration of Digital Technology.

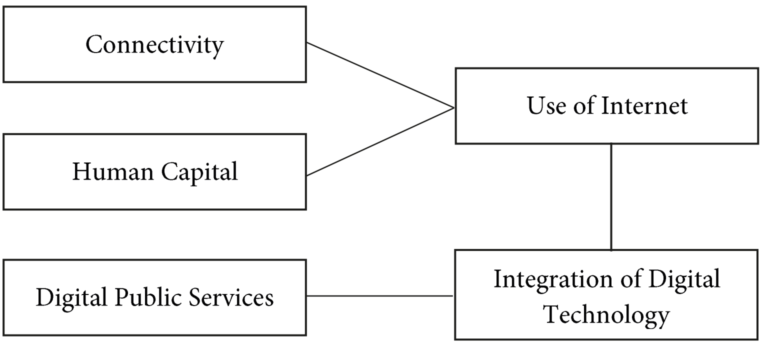

Fig. 1. Causal relationships between variables

Technology) showed a high *R*-square in both cases, with a value above 0.700. The details are omitted in this article and will be presented in a subsequent paper.

We should make one more comment. Bánhidi, Dobos and Nemeslaki [Bánhidi, Dobos, Nemeslaki, 2019] constructed a causal chain similar to that in Fig. 1 based on the DESI EU-28 dimensions. The figure now proposed corrects the erroneously constructed logical chain there. If temporal causality is taken into account [Pearl, 2009], technology, education, and the solution of public digital tasks must first be available in time for Use of Internet and Integration of Digital Technology to evolve. In summary, it is not dimension Digital Public Services that is at the end of the causal chain, but a consequence of residential and enterprise applications as the three premises.

2.4. RQ4: Grouping with Cluster analysis

Cluster analysis is a multivariate method which allows grouping objects, in our case 29 countries of the EU and the Russian Federation using the five I-DESI dimensions in a set of clusters. This method of analysis seems to be more objective, as countries could be divided to clusters based on similar points. The results of the cluster analysis are summarized in Fig. 2. The figure also shows how each group is formed.

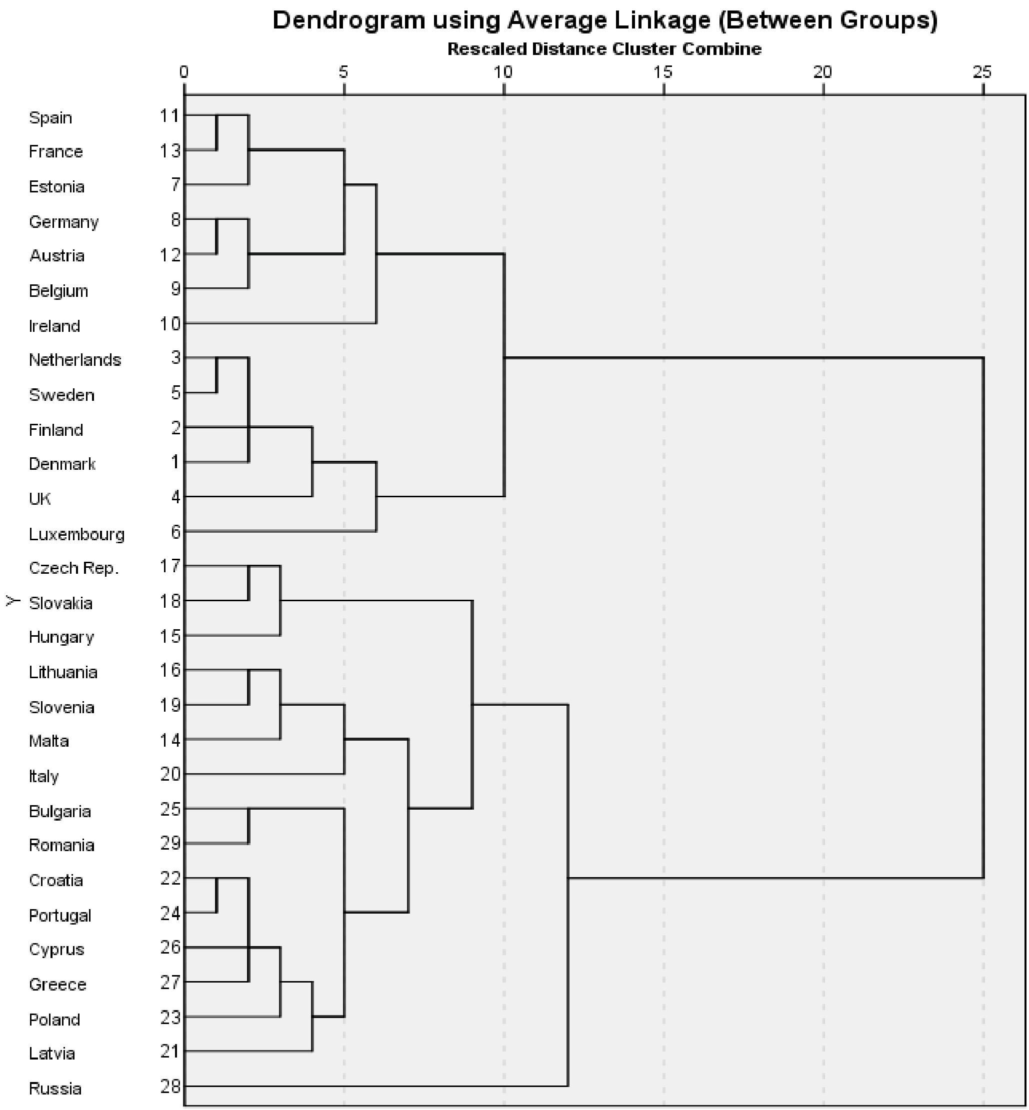

Fig. 2. Dendrogram of cluster analysis

In the first step, the countries were divided into two clusters. The first of these comprised top countries in the digital economy: Spain, France, Estonia, Germany, Austria, Belgium, Ireland, Netherlands, Sweden, Finland, Denmark, the UK, and Luxembourg.

In the second step, the groups were partitioned into three groups, separating Russia from the second cluster. Further, in the third step, a group consisting of the Netherlands, Sweden, Finland, Denmark, the UK, and Luxembourg was separated from the top-performing countries. In the fourth step, a Central-European cluster was created comprising the Czech Republic, Slovakia, and Hungary. The residual group would consist of Lithuania, Slovenia, Malta, Italy, Bulgaria, Romania, Croatia, Portugal, Cyprus, Greece, Poland and Latvia. We consider this an ideal end point to the clustering process, but the dendrogram also shows how these clusters could be subdivided into even smaller groups. The Russian Federation belongs to the group of moderately developed countries in terms of digital competitiveness, as shown above in Fig. 2. This is because when there are only two clusters, it is located in this group. At the same time, if we choose three clusters, it will be sharply detached from the moderately developed countries. This might be attributed to the fact that Russia occupies a prominent position in terms of Human Capital and Digital Public Services dimensions. Russia ranks 29th among 29 countries in Connectivity, 10th in Human Capital, 23rd in Use of Internet, 28th in Integration of Digital Technology and 19th in Digital Public Services.

As regards Connectivity, it should be noted that it is not only extremely challenging to cover the vast territory of the country, it is also financially unviable on a for-profit basis due to some sparsely populated areas (and could only be achieved with substantial state aid). Nevertheless, the country has set quite ambitious coverage targets to overcome its connectivity gap. In accordance with its modernization policy, the Ministry of Communications and Mass Media announced national broadband targets in 2012 to make 100 Mbps broadband available to 80 % of Russian residents by 2018, with the ultimate goal to provide conditions — physically and financially — for affordable broadband access for all residents of Russia. Since the 2018 I-DESI dataset comprises data from an earlier four-year period (2013–2016), it might not capture the progress made in this area in the subsequent years. At the same time, the corporate application of digital technology leaves much to be desired.

2.5. RQ5: Mapping with Multidimensional Scaling (MDS)

The application of multidimensional scaling can be said to be successful. Because we took two selected components of the principal component analysis, our map perfectly reproduced the location of each country in plane. Fig. 3 illustrates each country on this plane. The results are very similar to the dendrogram of the cluster analysis. We do not repeat the results of the cluster analysis; we only explain the location of Russia on the resulting map.

The Russian Federation is shown on the bottom right of the map. It is relatively far from moderately developed countries, but also from countries that can be considered digitally developed. What is interesting is that the closest country to Russia based on these is Italy. This draws attention to the fact that Russia's digital development is on a different path than that of the countries of the European Union.

Another interesting phenomenon is the two groups of countries that fall close to the Russian Federation outside Italy:

— Spain, France and Estonia, and
— Poland, Slovenia, Croatia, Lithuania and Greece.

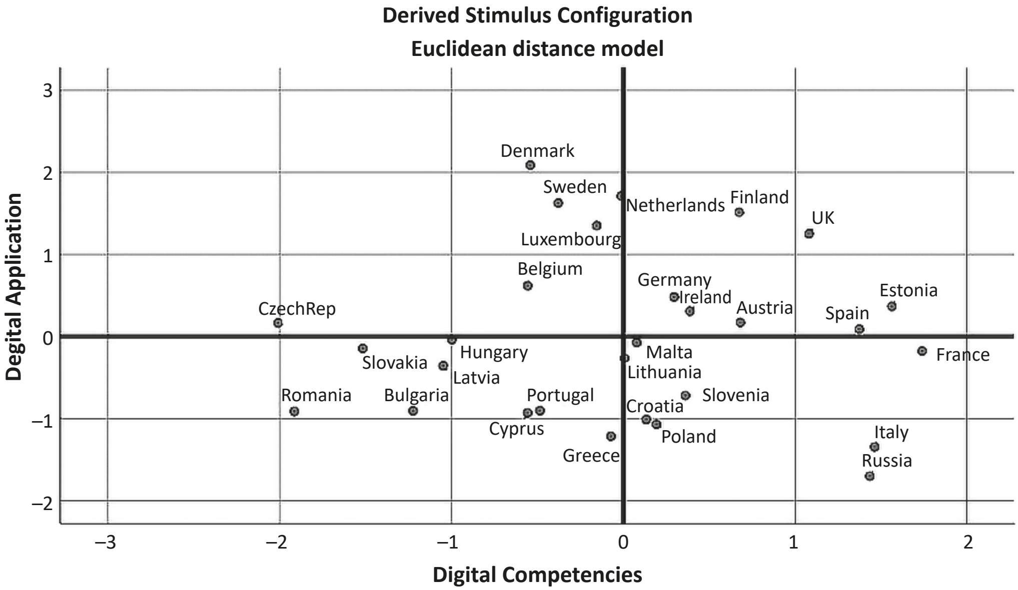

Fig. 3. Location of countries on the plane of principal components

The two groups of countries are roughly equidistant from the Russian Federation.

For a very simple interpretation of Fig. 3, we conclude that the DESI measurement distributes the Russian Federation and EU 28 countries into four quadrants based on their level of competency and application of digital technologies. The first quadrant in the upper right corner contains countries where both components are higher than average, as they perform well in both competencies (connectivity, use) and application (integration and public services). Moving clockwise, the lower right corner collects countries where applications lag the opportunities of competencies — the value of this latter component is negative. Countries which are lower than average in both components can be found in the lower left corner, which roughly corresponds to the residual cluster of the hierarchical cluster analysis showing the refined situation, as in Romania and Bulgaria; although lagging in terms of the competency component, they perform better in terms of the application dimension. Finally, the last quadrant is the upper left segment with countries that exploit applications of technology to a certain extent, beyond the level of competency or their actual readiness.

Conclusions

Our mathematical-statistical analyzes led to the conclusion that we should accept each of our hypotheses, *H1a., H2b., H2., H3a.* and *H3b.* This means that the five dimensions of I-DESI are redundant, i. e., two of the dimensions (Use of Internet and Integration of Digital Technology) do not carry significantly more information, and these can be captured through the other three dimensions. The Russian Federation can be classified into a separate group among the 29 countries studied at a substantially equal digital distance from the more developed and less developed countries of the European Union.

In this paper, we sought answers to two major questions. Our first question was how the information stored in the five dimensions of I-DESI can be reduced by the number of variables, and what causal relationship can be revealed between the dimensions. These two problems were answered by research questions RQ1, RQ2, and RQ3. The first research question (RQ1) draws attention to the fact that in the case of I-DESI as well, the correlation between the five digital dimensions is high, i.e. the variables are highly correlated. This proves that the European Commission's thesis about the relationship between the five principal dimensions of (EU-28) DESI, that these "are not isolated areas that contribute separately to digital development but in fact interconnected," is also true for the dimensions of the extended I-DESI dataset. The Commission's related interpretation is that "developments in the digital economy cannot be achieved through isolated improvements in particular areas but through concerted improvement in all areas"[2]. Although we can accept this as a sensible policy recommendation, we also feel that the high degree of collinearity might be inconsistent with the (statistical part of the) Commission's requirement that the "index should not contain indicators that are redundant, either *statistically* or in terms of interpretation".

The answer to RQ2 can be explained by a relatively high proportion of variance with two latent factors. The two components were also named. One measures technical digital readiness (competence), while the other component displays government digital readiness (application). In mapping causal relationships, we concluded that the Connectivity, Human Capital, and Digital Public Services dimensions explain the Use of Internet and Integration of Digital Technology dimensions as reasons. This also means that advancing the three dimensions can promote growth of the other two. As regards the relationship between the dimensions, the Commission argues that the "use of Internet (by citizens) and digital public services are enabled by the infrastructure" (Connectivity and Human Capital), "and their contribution is strengthened by the quality of such infrastructure"[3]. While this assertion has been confirmed by partial correlations for the Use of Internet dimension of I-DESI, the statistical evidence seems far weaker in the case of Digital Public Services, which had a much closer relationship with the Integration of Digital Technology than either of these two dimensions.

The answers to research questions RQ3 and RQ4 are extremely consistent. The digital development of the Russian Federation can be considered good along the dimensions of Human Capital and Digital Public Services; therefore, it can be classified as moderately highly developed compared to other EU countries. However, its location on the plane of principle components differs from the group of moderately developed EU countries, reflecting its unique characteristics (strengths and weaknesses). The Russian Federation's score in the Connectivity dimension is fairly low. In our opinion, this might be explained by the country's unique geographic and demographic characteristics (its vast territory and relatively low population density). On the other hand, the Russian Ministry of Communications and Mass Media has set quite ambitious national broadband coverage targets to overcome its connectivity gap, which are not yet reflected in our I-DESI database. However, it is questionable what development would be expected along the Integration of Digital Technology dimension.

[2] Digital Economy and Society Index (DESI) 2020 — Methodological Note. (2020) European Commission. URL: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=67082 (accessed: 10.11.2020).

[3] Ibid.

Further research would be needed to map the digital development of the Russian Federation to other groups of countries in the world, not just the countries of the European Union. This could also answer the question of how Russia stands in international competition, e.g. compared to the United States or China. To do this, all forty-five countries of I-DESI would have to be analyzed together. However, this study also showed very similar results in how these digital dimensions of the reduced EU-28 DESI and I-DESI relate to each other. We now believe that the two indicator systems are relatively homogeneous, so they can be considered statistically identical. However, examining the ranking between countries would require further analysis.

References

Akberdina V.V. (2018) Digitalization of industrial markets: Regional characteristics. *Upravlenets (The Manager)*, no. 6, pp. 78–87.

Bánhidi Z., Dobos I., Nemeslaki A. (2019) Comparative Analysis of the Development of the Digital Economy in Russia and EU Measured with DEA and Using Dimensions of DESI. *St Petersburg University Journal of Economic Studies*, no. 4, pp. 588–604.

Bánhidi Z., Dobos I., Nemeslaki A. (2020) What the overall Digital Economy and Society Index reveals: A statistical analysis of the DESI EU28 dimensions. *Regional Statistics*, no. 2, pp. 42–62.

Baskakova M., Soboleva I. (2019) New Dimensions of Functional Illiteracy in the Digital Economy. *Voprosy obrazovaniya (Educational Studies Moscow)*, no. 1, pp. 244–263.

Ermolaev K.N., Trubetskaya O.V., Shnyakin K.V., Pavlova J.A. (2019) Platforms and tendencies for the development of the digital economy in Russia. *SHS Web of Conferences*, vol. 62. https://doi.org/10.1051/shsconf/20196201001

Korovin G.B. (2018) Problems of Industrial Digitalisation in Russia. *Journal of New Economy. Ural State University of Economics*, no. 3, pp. 100–110.

Kuvayeva Yu. V. (2019) Digital economy: Concepts and Russia's readiness to transition. *Journal of New Economy. Ural State University of Economics*, no. 1, pp. 25–40.

Miethlich B., Belotserkovich D., Abasova S., Zatsarinnaya E., Veselitsky O. (2020) Digital economy and its influence on competitiveness of countries and regions. *Revista ESPACIOS*, no. 12, pp. 20–31.

Mironova O.A., Bogdanova R.M., Kolesnikov Yu. A. (2019) Aspects of the generational theory application in terms of digital education development in Russia. *Media Education (Mediaobrazovanie)*, no. 1, pp. 93–104.

Moroz M. (2017) The Level of Development of the Digital Economy in Poland and Selected European Countries: A Comparative Analysis. *Foundations of Management*, no. 1, pp. 175–90.

Pearl J. (2009) *Causality*. Cambridge University Press, Cambridge. 484 p.

Revinova S., Lazanyuk I. (2018) Prerequisites for Development of the Digital Economy in Russia: Differentiation of Regions. *International Journal of Business and Management Studies*, no. 2, pp. 607–617.

Stavytskyy A., Kharlamova A., Stoica E.A. (2019) The analysis of the digital economy and society index in the EU. *Baltic Journal of European Studies*, no. 3, pp. 245–261.

Vishnevsky V.P. (2019) The Digital Economy in the Context of the Fourth Industrial Revolution: Opportunities and Limitations. *St Petersburg University Journal of Economic Studies*, no. 4, pp. 606–627. (In Russian)

Received: 09.11.2020
Accepted: 30.03.2021

Authors' information:

Madina Tokmergenova — MA in Economics; tokmergenova.madina@gtk.bme.hu
Zoltán Bánhidi — MA in Economics; banhidi.zoltan@gtk.bme.hu
Imre Dobos — Dr. Sci. in Economics, Professor; dobos.imre@gtk.bme.hu

Анализ измерений I-DESI развития цифровой экономики в Российской Федерации и ЕС-28 с использованием многомерной статистики

М. Токмергенова, З. Банхиди, И. Добош

Будапештский университет технических и экономических наук,
Венгрия, 1117, Будапешт, бул. Венгерских Ученых, 2

Для цитирования: Tokmergenova M., Bánhidi Z., Dobos I. (2021) Analysis of I-DESI dimensions of the digital economy development of the Russian Federation and EU-28 using multivariate statistics. *Вестник Санкт-Петербургского университета. Экономика.* Т. 37. Вып. 2. С. 189–204. https://doi.org/10.21638/spbu05.2021.201

Авторы статьи продолжают предыдущую работу. В предыдущей работе, используя пять основных параметров Международного индекса цифровой экономики и социального развития (I-DESI) 28 стран ЕС и Российской Федерации, мы рассмотрели, как развитие России соотносится с развитием других стран ЕС. Целью данной статьи является не ранжирование, а определение взаимосвязи между каждым параметром и группами, на которые можно разделить 29 исследованных стран с помощью инструментов многомерного статистического анализа, а также анализ группы, к которой принадлежит Российская Федерация. Мы используем анализ главных компонентов (Principal Component Analysis, PCA) для сопоставления наших данных с пространством более низкого измерения (выявляя два скрытых измерения) и анализируем причинно-следственные связи между основными измерениями с использованием частных коэффициентов корреляции, делая вывод, что два из пяти основных измерений можно объяснить тремя независимыми измерениями. После этого мы используем кластерный анализ, чтобы сгруппировать наши объекты (т. е. 29 стран) в кластеры, и многомерное масштабирование (multidimensional scaling, MDS), чтобы визуализировать расположение этих групп и стран на плоскости двух компонентов (PCA), акцентируя внимание на Российской Федерации. По нашим результатам Российская Федерация может быть отнесена к умеренно развитой стране по показателю I-DESI, но ее положение в плоскости основных компонентов отличается от группы умеренно развитых стран ЕС, образующих отдельную «группу», во многом благодаря уникальным особенностям цифрового развития страны.

Ключевые слова: измерение цифровой трансформации, индекс DESI, анализ главных компонент, частичный корреляционный анализ, кластерный анализ, многомерное масштабирование.

Контактная информация:

Токмергенова Мадина — магистр экономики; tokmergenova.madina@gtk.bme.hu
Банхиди Золтан — магистр экономики; banhidi.zoltan@gtk.bme.hu
Добош Имре — д-р экон. наук, проф.; dobos.imre@gtk.bme.hu

Статья поступила в редакцию: 09.11.2020
Статья рекомендована в печать: 30.03.2021

Appendix

The basic data of digital dimensions

Country	Connectivity	Human Capital	Use of Internet	Integration of Digital Technology	Digital Public Services
Austria	0.63	0.59	0.60	0.59	0.72
Belgium	0.68	0.60	0.62	0.61	0.61
Bulgaria	0.61	0.47	0.42	0.36	0.45
Croatia	0.54	0.45	0.49	0.46	0.56
Cyprus	0.54	0.45	0.54	0.39	0.49
Czechia	0.67	0.58	0.58	0.39	0.43
Denmark	0.77	0.80	0.79	0.71	0.71
Estonia	0.62	0.66	0.70	0.53	0.85
Finland	0.72	0.73	0.78	0.67	0.83
France	0.59	0.62	0.59	0.53	0.82
Germany	0.64	0,62	0.66	0.59	0.69
Greece	0.50	0.48	0.46	0.45	0.48
Hungary	0.60	0.62	0.55	0.51	0.46
Ireland	0.63	0.77	0.56	0.51	0.66
Italy	0.51	0.50	0.42	0.47	0.68
Latvia	0.65	0.47	0.58	0.32	0.56
Lithuania	0.61	0.53	0.58	0.46	0.63
Luxembourg	0.65	0.67	0.79	0.77	0.64
Malta	0.64	0.48	0.57	0.57	0.66
Netherlands	0.75	0.69	0.76	0.75	0.76
Poland	0.53	0.53	0.51	0.33	0.57
Portugal	0.60	0.43	0.47	0.39	0.55
Romania	0.61	0.43	0.48	0.27	0.39
Russia	0.39	0.64	0.49	0.30	0.57
Slovakia	0.57	0.65	0.59	0.40	0.38
Slovenia	0.60	0.44	0.53	0.43	0.67
Spain	0.64	0.62	0.58	0.55	0.82
Sweden	0.75	0.69	0.78	0.65	0.73
United Kingdom	0.74	0.65	0.72	0.68	0.90

Source: International Digital Economy and Society Index 2018. URL: https://ec.europa.eu/digital-single-market/en/news/international-digital-economy-and-society-index-2018 (accessed: 05.06.2019).