Wiadomości Statystyczne. The Polish Statistician, 2021, vol. 66, 8, 24–45 DOI: 10.5604/01.3001.0015.2305
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 22.02.2021, poprawiony/revised: 6.07.2021, zaakceptowany/accepted: 28.07.2021

Principal components of innovation performance in European Union countries

Agnieszka Kleszcz[a]

Abstract. Innovation is one of the main determinants of economic development. Innovative activity is very complex, thus difficult to measure. The complexity of the phenomenon poses a great challenge for researchers to understand its determinants. The article focuses on the problem of innovation-related geographical disparities among European Union countries. Moreover, it analyses the principal components of innovation determined on the basis of the European Innovation Scoreboard (EIS) dimensions. The aim of the paper is to identify the principal components of the innovation index which differentiate countries by analysing the structure of the correlation between its components. All calculations were based on indicators included in the EIS 2020 Database, containing data from the years 2012–2019. A comparative analysis of the studied countries' innovation performance was carried out, based on the principal component analysis (PCA) method, with the purpose of finding the uncorrelated principal components of innovation which differentiate the studied countries.

The results were achieved by reducing a 10-dimensional data set to a 2-dimensional one, for a simpler interpretation. The first principal component (PC1) consisted of the human resources, attractive research systems, and finance and support dimensions (understood as academia and finance). The second principal component (PC2), involving the employment impacts and linkages dimensions, was interpreted as business-related. PC1 and PC2 jointly explained 68% of the observed variance, and similar results were obtained for the 27 detailed indicators outlined in the EIS. We can therefore assume that we have an accurate representation of the information contained in the EIS data, which allows for an alternative assessment and ranking of innovation performance. The proposed simplified index, described in a 2-dimensional space, based on PC1 and PC2, makes it possible to group countries in a new way, according to their level of innovation, which offers a wide range of application, e.g. PC1 captures geographic disparities in innovation corresponding to the division between the old and new EU member states.

Keywords: innovation, European Innovation Scoreboard, EIS, principal component analysis, PCA

JEL: O30, C10, O52

Główne składowe innowacyjności w krajach Unii Europejskiej

Streszczenie. Innowacyjność należy do głównych wyznaczników rozwoju gospodarczego. Działalność innowacyjna jest bardzo złożona, a przez to trudna do zmierzenia. Dużym wyzwaniem dla badaczy jest także poznanie uwarunkowań tego zjawiska. W artykule skupiono się na problemie zróżnicowania terytorialnego innowacyjności wśród krajów Unii Europejskiej, a także

[a] Uniwersytet Jana Kochanowskiego w Kielcach, Wydział Nauk Ścisłych i Przyrodniczych, Instytut Geografii i Nauk o Środowisku, Polska / Jan Kochanowski University of Kielce, Faculty of Natural Sciences, Institute of Geography and Environmental Sciences, Poland. ORCID: https://orcid.org/0000-0002-0450-5247.
E-mail: agakleszcz87@gmail.com.

na analizie głównych składowych innowacyjności wyznaczonych przy wykorzystaniu wskaźników uwzględnianych w Europejskim Rankingu Innowacyjności (European Innovation Scoreboard – EIS). Celem badania omawianego w artykule jest identyfikacja głównych składowych innowacyjności różnicujących kraje na podstawie analizy struktury korelacji. Obliczenia oparto na wskaźnikach zawartych w bazie EIS 2020, obejmujących 2012–2019. Przeprowadzono analizę porównawczą krajów pod kątem wydajności innowacyjnej przy użyciu metody analizy głównych składowych (PCA), aby znaleźć nieskorelowane główne składowe innowacji różnicujące kraje.

Wyniki uzyskano dzięki zredukowaniu 10-wymiarowego zestawu danych do zestawu 2-wymiarowego, łatwiejszego do interpretacji. Pozwoliło to wyróżnić pierwszą główną składową (PC1) zawierającą zasoby ludzkie, atrakcyjne systemy badawcze, finanse i wsparcie rozumiane jako środowisko akademickie i finanse. Druga główna składowa (PC2), obejmująca wpływ na zatrudnienie i sieć powiązań, jest interpretowana jako związana z biznesem. Składowe PC1 i PC2 wyjaśniły łącznie 68% wariancji; podobne wyniki uzyskano dla zestawu 27 szczegółowych wskaźników uwzględnianych w EIS. Można zatem uznać, że daje to dokładną reprezentację danych EIS, która zapewnia alternatywną ocenę i ranking wyników w zakresie innowacji. Zaproponowany uproszczony indeks innowacyjności, opisany w przestrzeni dwuwymiarowej, opierający się na PC1 i PC2, umożliwia nowy sposób grupowania krajów i może mieć szerokie zastosowanie, np. PC1 przedstawia geograficzne zróżnicowanie innowacji odpowiadające podziałowi na kraje członkowskie starej i nowej Unii.

Słowa kluczowe: innowacje, Europejski Ranking Innowacyjności, EIS, analiza głównych składowych, PCA

1. Introduction

In the literature, innovation is presented as the most important factor in achieving economic and employment growth and one of the main determinants of economic development in modern societies (Lee & Lee, 2020; Szopik-Depczyńska et al., 2018). According to the modern economic theory, innovation plays a significant role in the field of economic geography and regional science in terms of interpreting the organisation of economic activity in space and the growth and stagnation of regions over time (e.g. Acs, 2002; Makkonen & van der Have, 2013). It is worth noting here that the terms economic growth and economic development cannot be used interchangeably. Economic development describes all the changes in humanity's economic, social, and natural environments, making it a wider concept than economic growth, which refers solely to increases in material output *per capita* (Van den Berg, 2016).

One of the most important initiatives defined by the EU in its Europe 2020 Strategy was to create an innovation-friendly environment supporting the generation, emergence and diffusion of innovations (Zabala-Iturriagagoitia et al., 2020). Supporting innovation is also one of the 17 Sustainable Development Goals of the 2030 Agenda for Sustainable Development (Goal 9. Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation).

Within the modern globalised economy, innovation is considered the key driver of a country's productivity growth, competitiveness, and economic development

(Kontolaimou et al., 2016). According to the Organisation for Economic Co-operation and Development (2005, p. 46), 'An innovation is the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organisational method in business practices, workplace organisation or external relations'. Generally, innovation is understood as the invention of new products, the development of processes or services, and is perceived from the perspective of commercial activity (Dziallas & Blind, 2019).

It should also be mentioned that extensive research is conducted on the role of smart specialisation and on smart specialisation policies supporting the sustainable development policy in the face of new global challenges. The European Commission (EC) emphasises that the identification of smart specialisations will be crucial for achieving the smart growth priority outlined in the Europe 2020 Strategy, i.e. the development of a knowledge-based economy including innovations. Malik et al. (2020) identify the innovative industries being the potential drivers of regional development. Moreover, taking into account the fact that small and medium enterprises form the largest group of enterprises in the EU, there is a need to provide instruments such as processes which would allow the realisation of a sustainable development model. Economic growth based on the sustainable development concept aims to gradually bridge the gap between emerging and developed economies (Malik & Jasińska-Biliczak, 2018).

The complicated character of innovation has led to the development of a variety of measurement approaches and methodologies presented in the theory of innovation and applied in practice. However, no single universal method has yet been agreed upon, although numerous approaches are used (Kowalski, 2020). No consensus as to which indicator should be used to measure innovation has been reached either, although extensive discussions have been held in this respect. In consequence, several innovation indicators are used in the literature to measure and evaluate countries' innovative performance at a regional and national level. Makkonen and van der Have (2013) used the regional Finnish Innovations database – SFINNO. Another work presents a more in-depth innovation research on Swedish regions based on 44 collected variables (Holgersson & Kekezi, 2018).

An example of a national index is the composite Global Innovation Index (GII), published by Cornell University, INSEAD, and the World Intellectual Property Organization in partnership with other organisations and institutions measuring economies' innovation performance. Each year the GII presents a thematic component which tracks global innovation. It encompasses 81 different indicators for 143 countries, using different sub-indexes to aggregate them (Dutta et al., 2018).

Another popular indicator for measuring innovation at the country level is the annually published European Innovation Scoreboard (EIS) report, within which the Summary Innovation Index (SII) is developed (EC, 2019b).

The significance of innovation is also evidenced in the publication trends of research papers from the fields of management and business, economics, and geography. In the literature, innovation indexes are presented and analysed in various forms and may, e.g. focus on changes over time (e.g. Kowalski, 2020) or include research on the causal relationship between innovation, financial development, and economic growth (e.g. Mtar & Belazreg, 2020). Stojanovska and Madzova (2018) presented the differences in the efficiency of innovation performance between two groups of countries: EU candidate countries, including Macedonia, Serbia, and Turkey and the average level of innovation performance observed among the EU-28 countries in the years 2010–2017 using 12 EIS indicators. The issue of the multidimensionality of innovation has also been analysed by e.g. Holgersson and Kekezi (2018), who pointed out that the economic space of innovation generated by a large number of unrelated and independent, yet relevant economic factors seems somewhat unreasonable. Rather, one may expect the existence of only a few factors determining most of the unique variation in variables of innovation. Edquist et al. (2018) asserted that the SII does not constitute a meaningful measure of innovation performance and that this indicator is not useful from the point of view of innovation-related policy design. Bielińska-Dusza and Hamerska (2021) analysed the determinants affecting the level of the SII through stepwise regression, proving that reducing the number of indicators used within the EIS ranking procedure from 27 to 22 is in fact possible.

Researchers are still facing the issue of how to quantify and measure innovation due to its complex nature. Despite the existence of numerous studies addressing this matter, further research into the phenomenon and its complex character is crucial. The aim of this paper is to identify the principal components of the innovation index which differentiate the countries by analysing the structure of the correlation between its components. The research hypothesis asserts that some of the information resulting from the EIS indicators and dimensions is redundant due to the correlations between the variables; on the other hand, the current aggregation of the dimensions into one general index (the SII) fails to include some important features differentiating countries. Components of innovation, however, can be aggregated without losing much information about these features.

2. Methodology

The presented research examines the EU-28 countries' 2019 innovation scoreboard obtained from the EIS database. Ten dimensions of the EIS were taken into account, which aggregate the performance of a wide range of different indicators to measure the countries' innovation level. A common method for assessing a country's innovation level is to use aggregated indexes consisting of multiple indicators, e.g.

the SII, using the arithmetic mean as the aggregating function, in which all the indicators are given the same weight. This paper proposes the application of principal component analysis (PCA) to calculate a more meaningful representation of the innovation performance data, by reducing the 10-dimensional dataset into a low-dimensional one which differentiates countries to the greatest possible extent.

Thanks to this simpler representation, the EIS principal components may be assigned interpretations. Finally, the paper presents the assessments of the new EU countries based on the uncorrelated principal components data. The obtained results clearly demonstrate the existence of geographical disparities in innovation and provide more precise information on what determines the tendency to innovate.

All calculations and visualisations were performed with the use of the R software. The PCA calculations involved the application of the *prcomp* function.

2.1. European Innovation Scoreboard

The EIS is frequently used in the literature for ranking and assessing the level of innovation in European countries (e.g. Kowalski, 2020; Onea, 2020; Stojanovska & Madzova, 2018). The annual EIS provides a comparative assessment of the research and innovation performance of the EU member states and selected non-EU countries.

The overall performance of each country's innovation system has been summarised in a composite index – the SII. It helps countries to determine which areas require intensified efforts for their innovation performance to boost.

The EIS measurement framework distinguishes between four main types of indicators:

- framework conditions – captures the main drivers of innovation performance outside the enterprise;
- investments – captures investments made in both the public and business sectors;
- innovation activities – captures different aspects of innovation in the business sector;
- impacts – captures the effects of firms' innovation activities.

In 2019, those four main types of indicators consisted of 10 innovation dimensions, capturing a total of 27 different indicators (Table 1). All the indicators are stimulants of innovation, and so expressed as a percentage or a dimensionless quantity; however, normalised indicators are commonly used by researchers. The indicators are grouped into 10 dimensions of similar aspects and their values are obtained by averaging appropriate indicators (EC, 2019a, 2020b, 2020c). Methodology for calculating composite scores is available in EC (2020c, p. 13). The methodology for calculating the SII consists of eight steps. Step 7 (calculating composite innovation indexes) states that for each year a composite SII is calculated as the unweighted average of the rescaled scores for all indicators where all the indicators receive the same weight (1/27 if data are available for all the 27 indicators) (Bielińska-Dusza & Hamerska, 2021, p. 8; EC, 2020c).

Table 1. European Innovation Scoreboard

Types of indicators	Innovation dimensions	Indicators
Framework conditions	Human resources	1.1.1. New doctorate graduates[a] 1.1.2. Population aged 25–34 with tertiary education[b] 1.1.3. Lifelong learning[b]
	Attractive research systems	1.2.1. International scientific co-publications[a] 1.2.2. Top 10% most cited publications[b] 1.2.3. Foreign doctorate students[b]
	Innovation-friendly environment	1.3.1. Broadband penetration[a] 1.3.2. Opportunity-driven entrepreneurship[a]
Investments	Finance and support	2.1.1. Research and development (R&D) expenditure in the public sector[b] 2.1.2. Venture capital investments[b]
	Firm investments	2.2.1. R&D expenditure in the business sector[b] 2.2.2. Non-R&D innovation expenditure[b] 2.2.3. Enterprises providing training to develop or upgrade the information and communications technology (ICT) skills of their personnel[a]
Innovation activities	Innovators	3.1.1. Small-and medium-sized enterprises (SMEs) with product or process innovations[b] 3.1.2. SMEs with marketing or organisational innovations[b] 3.1.3. SMEs innovating in-house[b]
	Linkages	3.2.1. Innovative SMEs collaborating with others[b] 3.2.2. Public-private co-publications[a] 3.2.3. Private co-funding of public R&D expenditure[b]
	Intellectual assets	3.3.1. Patent Cooperation Treaty (PCT) patent applications[a] 3.3.2. Trademark applications[a] 3.3.3. Design applications[a]
Impacts	Employment impacts	4.1.1. Employment in knowledge-intensive activities[b] 4.1.2. Employment in fast-growing firms' innovative sectors[b]
	Sales impacts	4.2.1. Medium & high-tech product exports[b] 4.2.2. Knowledge-intensive services exports[b] 4.2.3. Sales of new-to-market and new-to-firm product innovations[b]

a, b Expressed as: a – a dimensionless quantity, b – a percentage.

Note. Single underscore – the first principal component (PC1) and the newPC1, double underscore – the second principal component (PC2) and the newPC2. Details are explained in section 3. Statistical analysis of the principal components of innovation – research results.

Source: author's work based on EC (2019b, 2020c).

Despite its annual publication, there are some limitations of the EIS index highlighted in its methodology report (EC, 2019a, p. 20, 2020c, p. 22). The authors of the report stress that the results presented in different EIS reports are not fit for comparison for the following reasons: (i) in relation to several indicators, data have been revised in external sources from which the data have been extracted, (ii) the time period covered in various reports is different, e.g. the oldest data used in the EIS 2019 no longer used in the EIS 2020, (iii) data transformations have been applied to a slightly different set of indicators.

EIS uses the following classification structure when arranging countries into particular performance groups (see Map A):

- Innovation Leaders are all countries whose relative performance in 2018 exceeded 120% of the EU average reported the same year (Sweden, Finland, Denmark, the Netherlands);
- Strong Innovators are countries whose relative performance in 2018 reached between 90% and 120% of the EU average reported the same year (Luxembourg, Belgium, the United Kingdom, Germany, Austria, Ireland, France, Estonia);
- Moderate Innovators encompasses countries whose relative performance in 2018 totalled between 50% and 90% of the EU average reported the same year (Portugal, the Czech Republic, Slovenia, Cyprus, Malta, Italy, Spain, Greece, Lithuania, Slovakia, Hungary, Latvia, Poland, Croatia);
- Modest Innovators are those countries whose relative performance in 2018 was below 50% of the EU average reported the same year (Bulgaria, Romania) (EC, 2019a).

2.2. Dataset

The research examines the 2019 innovation scoreboard relating to the EU member states. This research is based on normalised values of indicators which are therefore dimensionless numbers from the $[0, 1]$ interval, obtained from the EIS 2020 Database (composite indicators and normalised scores for 27 indicators in 2019 were extracted from sheets 1.1.1 to 4.2.3) (EC, 2020a).

The dataset is comprised of data relating to all 28 EU member states in 2019. These are: Austria, Belgium, Bulgaria, Croatia, Cyprus the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the United Kingdom (the last report is based on 2019 data, so the United Kingdom is also included).

2.3. Principal component analysis

Often a small number of objects are described by a large number of features. When the features describe just a few aspects of an object, the information is repeated among the features which, in result, become correlated. Dimensionality reduction involves searching for a new representation of data preserving as much information as possible (and in this case optimally minimising the number of features). The new representation has therefore a lower number of dimensions. Many techniques have been developed for this purpose, e.g. factor analysis, autoencoders (Bishop, 2006), t-Distributed Stochastic Neighborhood Embedding (t-SNE) (van der Maaten & Hinton, 2008), but PCA remains one of the oldest and most popular methods (Jolliffe & Cadima, 2016). Among its numerous applications, the most prominent one is face recognition (Turk & Pentland, 1991). In economics, PCA has been successfully used in the construction of an index system (Davidescu et al., 2015; Zihao et al., 2020).

Mathematically, PCA is a linear projection of original variables onto lower dimensions called principal components (PCs). The first PC is chosen to minimise the total distance between the data and their projection onto the PC. By minimising this distance, we also maximise the variance of the projected points.

For a p-dimensional feature vector $\boldsymbol{x} = \left(x_1, x_2, \dots, x_p\right)$, the first PC is obtained as a linear combination which maximises the variance of

$$z_1 = \phi_{11}x_1 + \phi_{21}x_2 + \cdots + \phi_{p1}x_p \tag{1}$$

and is constrained so that $\sum_i \phi_{i1}^2 = 1$. Vector $\boldsymbol{\phi}_1 = (\phi_{11}, \phi_{21}, \dots, \phi_{p1})$ represents the direction of the maximum variability in the data. The subsequent PCs are selected similarly, with an additional requirement assuming that they remain uncorrelated with all the previous PCs. For example, the second PC is the one maximising the variance of

$$z_2 = \phi_{12}x_1 + \phi_{22}x_2 + \cdots + \phi_{p2}x_p, \tag{2}$$

constrained so that $\sum_i \phi_{i2}^2 = 1$ and z_2 is uncorrelated with z_1 ($\boldsymbol{\phi}_1$ is orthogonal to $\boldsymbol{\phi}_2$). The direction along $\boldsymbol{\phi}_1$ has the second-highest variability in the dataset. A useful interpretation of PCA is that ρ^2 of the regression is the percent variance (of all the data) explained by the PCs.

Numerically, PCA is computed as an eigenvector decomposition of the covariance or correlation matrix ($\mathbf{R}$) for n observations: $\mathbf{X} = [\boldsymbol{x}_1^T, \dots, \mathbf{x}_n^T]^{\mathrm{T}}$ ($\boldsymbol{\phi}_k$ is the k-th eigenvector of the appropriate matrix). A more numerically stable implementation is based on the singular value decomposition (SVD) of column-centered data matrix $\mathbf{X}^* = [X_{i,j} - \bar{X}_j]$. In geometric interpretation, $\{\boldsymbol{\phi}\}$ is the new set of basis vectors in

$\mathbb{R}^p$ and PCA can be interpreted as a rotation of the coordinate system so that the basis gets the directions of the maximal variability (Gareth et al., 2013; Lever et al., 2017).

In standard PCA terminology, the elements of eigenvectors ϕ_k are commonly referred to as PC loadings, whereas the elements of linear combinations $\mathbf{X}^*\boldsymbol{\phi}_k$ are called PC scores, because they are the values that each individual element would score on a given PC. Both the scores and loading are commonly visualised. In particular, a joint representation of scores and loading projected onto the first two PCs is known as a biplot (Gray, 2017; Murphy, 2021).

PCA can be applied to any data; however, it gives meaningful results only if the variables are dependent. For independent variables, all directions are equivalent and principal components are determined by the noise in data, which can be tested by Bartlett's test of sphericity. The null hypothesis of the test assumes that the variables are orthogonal, i.e. not correlated. The alternative hypothesis is that the variables are not orthogonal, i.e. they are correlated to the extent to which the correlation matrix diverges significantly from the identity matrix as measured by the test statistics:

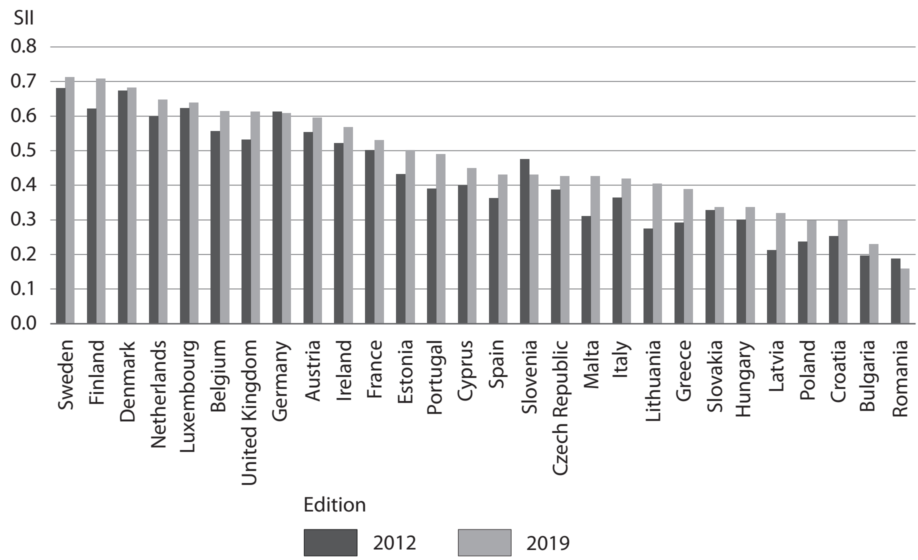
$$\chi^2 = -\left(n - 1 - \frac{2p+5}{6} \cdot \log(|\mathbf{R}|)\right). \tag{3}$$

The PCA is able to perform a compression of the available information only if the null hypothesis is rejected. Bartlett's χ^2 is asymptotically χ^2-distributed with $df =$ $= p(p-1) / 2$ under the null hypothesis (Snedecor & Cochran, 1989).

3. Statistical analysis of the principal components of innovation – research results

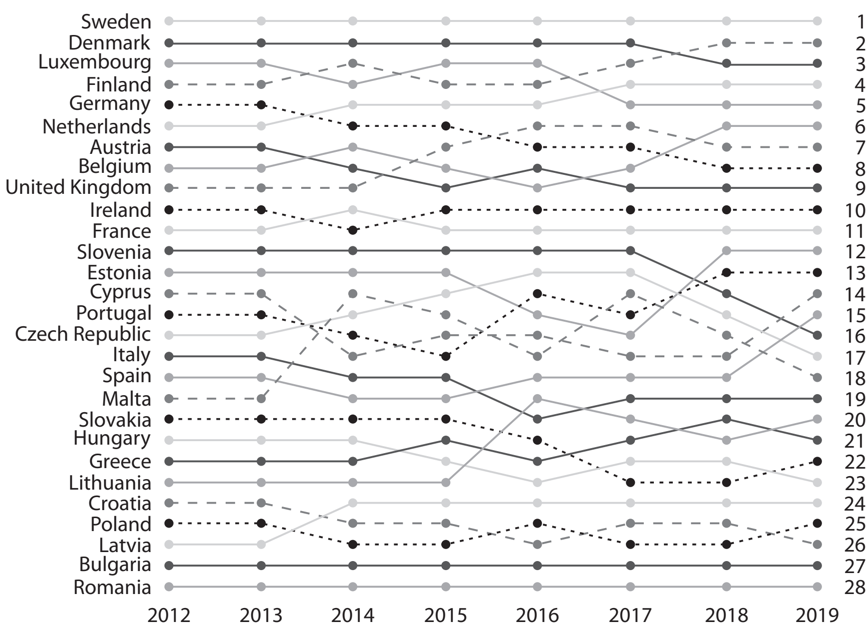
Before presenting the PCA results for the 10 dimensions of the EIS innovation index, let us begin with a brief visualisation of the dataset. Figure 1 shows an EU-28 countries ranking in 2012 and 2019. The countries are ranked in a descending order based on the 2019 SII. The results of the analysis of the 2019 edition of the SII indicate that some countries are highly developed, while others lag considerably. The countries with the highest value of the indicator in 2019 were: Sweden, Finland, Denmark, and the Netherlands, while Poland, Croatia, Bulgaria, and Romania had the lowest value. Innovation performance increased to the largest extent in Lithuania (+0.129), Malta (+0.115), Latvia (+0.107), Portugal (+0.100), Greece (+0.096), and decreased most considerably in Slovenia (–0.046), Romania (–0.027), and Germany (–0.002).

Figure 1. Performance of EU member states' innovation system according to SII



Source: author's work based on EC (2020a).

Figure 2. SII ranking within EU-28

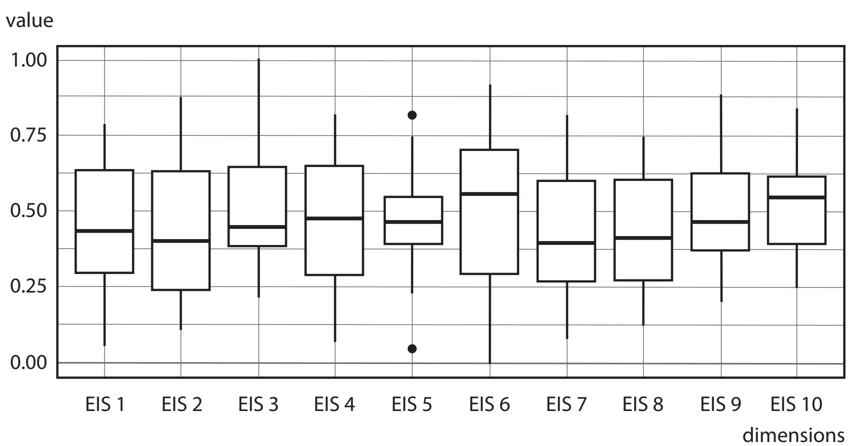


Note. The scale of the SII ranking is from 1 to 28, where 1 is the best score and 28 is the weakest score.
Source: author's work based on EC (2020a).

Since the SII cannot be compared between years, the time analysis is limited to the ranking (Figure 2). No significant changes are observed in the years 2012–2019 (the maximum movement is three positions), however, some countries improved their position: Finland (2), the Netherlands (2), Belgium (2), the United Kingdom (2), Estonia (1), Portugal (2), Spain (3), Malta (1), Lithuania (3), Greece (1), and Latvia (2).

Innovation dimensions are reported as normalised 10-dimensional vectors. The basic statistical visualisation depicted in Figure 3 (based on a box-plot) shows that the dimensions are of a similar scale. Regarding the company investment (EIS 5) dimension, an upper outlier – Germany, and lower outlier – Romania, are observed. The majority of dimensions point to disparities between countries in the form of noticeable 'whiskers' (especially in relation to innovation-friendly environment and innovators).

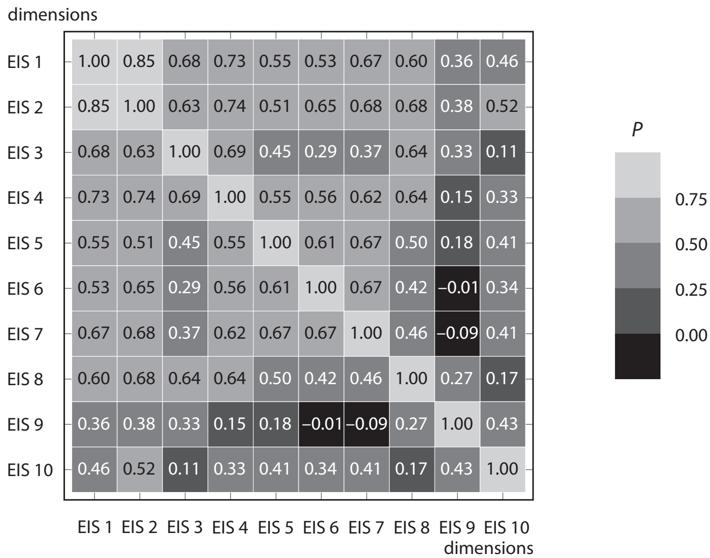
Figure 3. Innovation dimensions concerning EU-28 countries in 2019



Note. EIS 1 – human resources, EIS 2 – attractive research systems, EIS 3 – innovation-friendly environment, EIS 4 – finance and support, EIS 5 – firm investments, EIS 6 – innovators, EIS 7 – linkages, EIS 8 – intellectual assets, EIS 9 – employment impacts, EIS 10 – sales impacts.
Source: author's work based on EC (2020a).

Figure 4 presents a correlation matrix, showing correlation coefficients between variables. We can observe that some dimension variables are strongly correlated with each other (e.g. EIS 1 – Human resources with EIS 2 – Attractive research systems). However, the correlation of the dimensions is mostly moderate and weak. A weak negative correlation appeared only in a few cases. The off-diagonal entries in the correlation matrix form groups of similar values, which may indicate the presence of a common factor underlying innovation dimensions, which motivated the author to use PCA.

Figure 4. Correlations between innovation dimensions

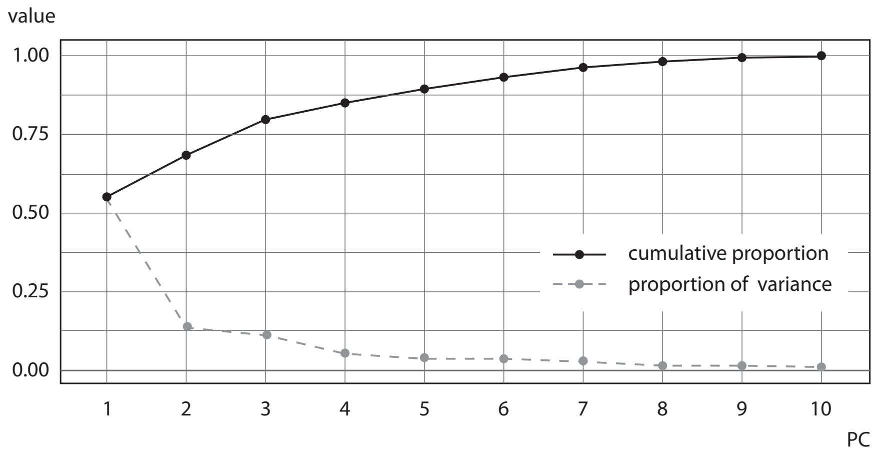


Note. As in Figure 3.
Source: author's work based on EC (2020a).

Figure 5 presents the PCA of 10 dimensions of the EIS components. Reducing the dimensionality facilitates the identification of the patterns among the EIS components. In this particular case, we performed the correlation matrix-based PCA (scale=TRUE in R) on normalised EIS data (each variable has a mean of zero and a unit standard deviation). As expected, the principal components are a good representation of the data, with only four of them explaining over 85% of the variance (see Figure 5 presenting the scree plot).

The application of PCA is also justified by Bartlett's test of sphericity. The test rejects the null hypothesis assuming that the EIS components are orthogonal (not correlated) at the 0.05 (p-value $= 8.592 \times 10^{-18}$) significance level. An uncorrelated representation in principal components can be obtained by a linear transformation with a rotation matrix (loadings matrix) shown in Table 2 (p. 37). The table entries are the loadings of all of the variables for each of the studied principal components. The values of those loadings (in bold) are considered the most important for our interpretation of the first two principal components (top three loadings according to the absolute value).

Figure 5. Fraction of the explained variance as a function of the number of principal components



Source: author's work based on EC (2020a).

Despite most of the variance being explained by the first four PCs, they are difficult to visualise, so we focused on the first two of them (PC1 and PC2). They are visualised in Figure 6 as a biplot, and jointly explain over 68% of the variance (PC1 – 55% and PC2 – 13.5%). The countries represented by PCA scores are scattered across the PC1–PC2 plane. The visualisation also informs about the value of the EIS dimensions, as well as of the fact of belonging to the old EU (countries which joined the EU before 2004), or to the new EU (countries which joined the EU after 2004), to facilitate the interpretation.

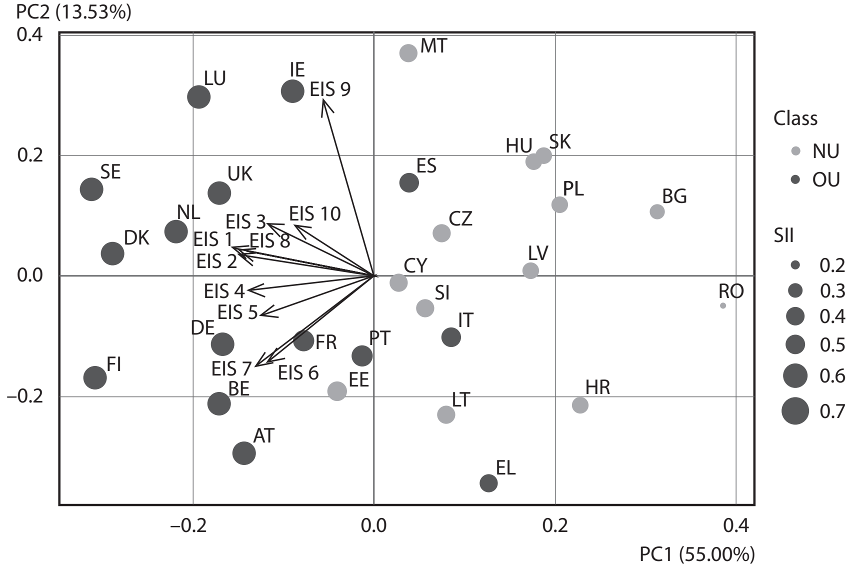
Most of the loadings, along with PC1, are influenced to the greatest extent by the following dimensions: attractive research systems, human resources, and finance and support. Attractive research systems is made of three indicators and measures the international competitiveness of the science base, by focusing on international scientific co-publications, most cited publications, and the number of foreign doctorate students. The human resources dimension consists of three indicators and measures the availability of a high-skilled and educated workforce. Human resources captures new doctorate graduates, the population aged 25–34 with completed tertiary education, and population aged 25–64 involved in education and training. Finance and support is comprised of two indicators, and measures the availability of finance for innovation projects by venture capital expenditure, and the support of governments for research and innovation activities by R&D expenditure in universities and government research organisations. Upon inspection of those components, we conclude that they are mostly related to academia, finance and high-skilled workforce. In conclusion, the interpretation of PC1 provides a measure of the condition of those areas.

Table 2. Rotation matrix: the matrix of variable loadings (columns are eigenvectors of the correlation matrix)

Innovation dimensions	Symbol	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Human resources	EIS 1	**–0.380**	0.105	–0.024	0.288	–0.171	0.167	–0.422	0.146	–0.685	–0.188
Attractive research systems	EIS 2	**–0.391**	0.092	0.028	0.333	0.266	0.012	–0.144	–0.119	0.156	0.773
Innovation-friendly environment	EIS 3	–0.307	0.213	–0.484	–0.031	–0.342	0.285	0.102	–0.588	0.216	–0.153
Finance and support	EIS 4	**–0.362**	–0.067	–0.201	0.215	–0.184	0.005	0.652	0.562	0.053	0.000
Firm investment	EIS 5	–0.317	–0.175	0.156	–0.780	–0.295	0.041	–0.003	0.053	–0.180	0.336
Innovators	EIS 6	–0.305	**–0.374**	0.186	–0.083	0.607	0.465	0.188	–0.163	–0.042	–0.278
Linkages	EIS 7	–0.333	**–0.394**	0.154	0.101	–0.228	–0.151	–0.455	0.157	0.568	–0.265
Intellectual assets	EIS 8	–0.318	0.091	–0.371	–0.220	0.414	–0.695	–0.032	–0.055	–0.102	–0.190
Employment impacts	EIS 9	–0.143	**0.741**	0.178	–0.246	0.192	0.239	–0.132	0.340	0.291	–0.162
Sales impacts	EIS 10	–0.226	0.217	0.687	0.160	–0.184	–0.330	0.323	–0.362	–0.069	–0.147

Source: author's work based on EC (2020a).

Figure 6. Projection of countries and dimensions onto the plane spanned by the first two principal components



Note.
PC1: academia, finance and high-skilled workforce, PC2: business.
AT – Austria, BE – Belgium, BG – Bulgaria, CY – Cyprus, CZ – the Czech Republic, DE – Germany, DK – Denmark, EE – Estonia, EL – Greece, ES – Spain, FI – Finland, FR – France, HR – Croatia, HU – Hungary, IE – Ireland, IT – Italy, LT – Lithuania, LU – Luxembourg, LV – Latvia, MT – Malta, NL – the Netherlands, PL – Poland, PT – Portugal, RO – Romania, SE – Sweden, SI – Slovenia, SK – Slovakia, UK – the United Kingdom.
OU – old EU, NU – new EU.
The negative PC1 represents the countries' high EIS value (above the EU-28 average). This is just an artifact of the numerical implementation of PCA and can be inverted by multiplying the basis vector by – 1.
Source: author's work based on EC (2020a).

PC2 is mostly influenced by employment impacts with an additional contribution from linkages and innovators. Employment impacts consists of two indicators measuring the employment in knowledge-intensive activities and in fast-growing firms in innovative sectors. The innovators dimension encompasses three indicators measuring the share of firms which have introduced innovations into the market or within their organisations, covering both product and process innovators, marketing and organisational innovators, and SMEs that innovate in-house. Linkages is comprised of three indicators measuring innovation capabilities by looking at collaboration efforts between innovating firms, research collaboration between the private and public sector, and the extent to which the private sector finances public R&D activities. Therefore, the interpretation of PC2 provides a measure of the condition of business-related areas.

Our claims are supported by an additional PCA of 27 indicators included in the EIS. The new principal components (referred to as newPCs to avoid ambiguity) mostly belong to the same categories as the PC1 and PC2 analysed above. Scientific publications from among the top 10% most cited publications and international scientific co-publications (corresponding to the attractive research systems being one of the most influential components of PC1), and additionally PCT patent applications and public-private co-publications constitute the highest contribution to the newPC.

The highest contribution to the newPC2 includes private co-funding of public R&D expenditure and innovative SMEs collaborating with others (which corresponds to the linkages system being one of the most influential components of PC2). In general, newPC1 and newPC2 explain over 56% of the variance, while the first ten newPCs explain over 90%. Thus, we can conclude that the remaining 17 observations (explaining less than 10%) can be considered redundant. In cases requiring a multivariate description of innovation, the use of 10 newPCs instead of 10 EIS dimensions is recommended, as they contain the same information, yet they are uncorrelated, thus easier to interpret.

The further analysis of the biplot in Figure 6 shows that a few patterns emerge from this representation. Thirteen countries have PC1 below zero (representing the average for EU-28), in this particular orientation of the PC1 axis, and these are mostly old EU member states. The lowest values are noted in Sweden, Denmark, and Finland. The other low values are observed in the Netherlands, Luxembourg, the United Kingdom, Germany, Belgium, Austria, Ireland, France and Portugal. Only one country from the new EU appeared in those groups – Estonia. Three countries from the old EU, namely Spain, Italy, and Greece have values above the mean for EU-28. This group encompasses almost all new EU countries (except Estonia). The highest values were noted for Bulgaria and Romania.

Since the countries with the highest innovation index are located in the upper left quarter, we can notice that a negative direction in PC1 and a positive one in PC2 is the direction of increasing innovation (PCA finds only directions and not orientations). Based on this observation, we can categorise countries depending on which quarter they are located in. Each quarter covers a different group of countries; as a result, we arrange quarters in a descending order 3–0, where 3 is the strongest and 0 is the weakest group (see Map B). The upper left-hand quarter (PC1 below 0 and PC2 above 0) on the biplot is the highest and the bottom right-hand quarter (PC1 above 0 and PC2 below 0) is the lowest. The bottom left-hand quarter (PC1 below 0 and PC2 below 0) ranks 2, while the upper right-hand quarter (PC1 above 0 and PC2 above 0) ranks 1, and both fall in the middle of the ranking. Since PC1 explains most of the variance, it is assigned a higher weight and 0 is the natural discrimination level. The class assignment is depicted in Map B, which presents a new picture of the studied countries' innovation assessment. The new ranking groups the countries as follows:

- 3: Innovation Leaders – Denmark, Luxembourg, the Netherlands, Ireland, the United Kingdom, and Sweden perform significantly above the EU average in PC1 and PC2;

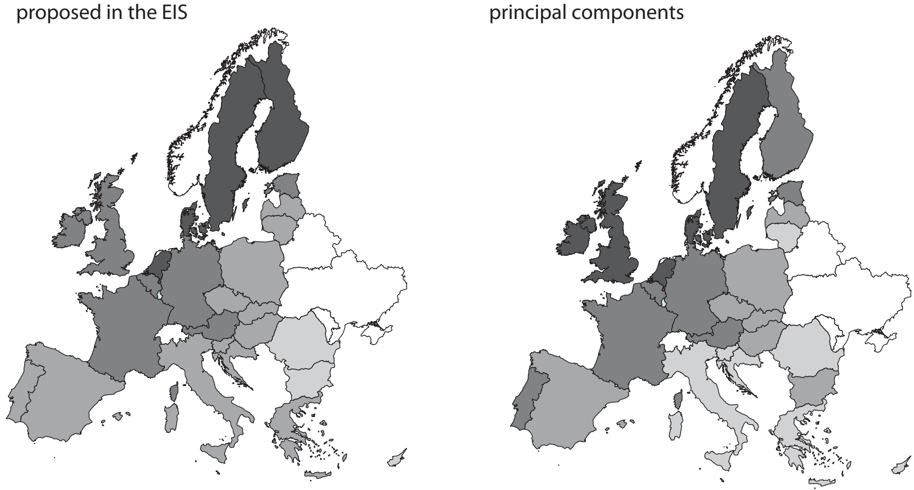
- 2: Strong Innovators – the innovation performance of Finland, Austria, Belgium, Estonia, France, Germany, and Portugal is above the EU average for PC1 and below for PC2;
- 1: Moderate Innovators – the innovation performance of the Czech Republic, Hungary, Latvia, Malta, Poland, Slovakia, Spain, and Bulgaria is below the EU average for PC1 and above average for PC2;
- 0: Modest Innovators – the innovation performance of Lithuania, Slovenia, Italy, Cyprus, Croatia, Greece, and Romania is below the EU average both for PC1 and PC2.

Map B presents the distribution of innovation among the studied countries based on PC1 and PC2. The results from the new rankings based on PC1 and PC2 are similar to the mainstream classification proposed in EIS reports Map A. Note that the proposed EIS' methodology (based on SII) has a slightly different classification structure (e.g. Innovation Leaders are all the countries whose relative performance exceeds the EU average by over 20%).

Map. Comparison maps of innovation performance of EU-28 countries based on the EIS classification and PC1 and PC2

A. Countries classification based on SII proposed in the EIS

B. Countries classification based on the PC1 and PC2 principal components



Innovation Leaders
Strong Innovators
Moderate Innovators
Modest Innovators

3: the upper left-hand quarter on Figure 6 (PC1 below 0 and PC2 above 0)
2: the bottom left-hand quarter on Figure 6 (PC1 below 0 and PC2 below 0)
1: the upper right-hand quarter on Figure 6 (PC1 above 0 and PC2 above 0)
0: the bottom right-hand quarter on Figure 6 (PC1 above 0 and PC2 below 0)

Source: author's work based on EC (2019a, 2020a).

A few countries were assigned a different class assignment, but the difference is at most one step. From among those countries Portugal, the United Kingdom, Ireland, and Bulgaria achieved better results, while Finland, Italy, Greece, Lithuania, Slovenia, and Croatia are lower compared to the EIS ranking based on SII. The advantage of the proposed ranking's mechanism is its simpler interpretability compared to the EIS ranking. The biplot quarter together with the interpretation of the PCs provide not only a country's ranking but also information about its strongest and weakest aspects. For example, Malta is in the Moderate Innovators group 1, but its business-related performance is the best and the employment impacts dimension is the second highest among the EU-28. However, the country has some slight deficiencies in PC1, and an improvement in the academia and finance aspects would enable it to be classified in the Innovation Leaders group.

4. Conclusions

Many researchers of innovation agree that innovative activity is very complex. The complexity of this phenomenon poses a great challenge for researchers to understand its determinants.

The EIS provides a yearly comparative assessment of research and innovation performance in the EU countries. The EIS, based on the scores for 27 separate indicators, largely corresponds to the key fields of an innovation system which consists of an appropriate level of public and private investment in education, research and skills development, efficient innovation partnerships among companies and with academia, and an innovation-friendly business environment. However, assigning the same weight to all the indictors is impossible in EIS, in contrast to the SII (using the arithmetic mean as the aggregating function).

In this paper, we indicate the principal components influencing the innovation performance by using the PCA method. The PCA uses the same dimensions as SII but with different weights, and allows the aggregation of dimensions with a lower number of representative variables which collectively explain most of the variability in the EIS data.

Two principal components of innovations, explaining over 68% of the variance, have been identified. PC1 tends to be mostly impacted by the following dimensions: research system, human resources, finance and support, while PC2 is influenced by the employment impacts to the largest extent. The proposed interpretation for the first two PCs as academia and finance and business can be useful for a model builder, since the variables are independent and allow for an easier interpretation of linear models when used as predictors. Furthermore, we proposed a country innovation performance ranking based on PCA scores. In a majority of cases our results corresponded with the classification into four performance groups based on the SII presented in the EIS report. However, we obtained similar results based on only two principal components, which proves their significance and shows that they may be applied instead of the average 27 indicators. The PCA methodology can be extended to cover over 90% of the variance by using 10 principal components of the EIS indicators. Those components can be considered as an alternative to the 10 EIS dimensions.

The presented results also emphasise the geographic disparities of innovation between countries. In particular the old and new EU countries are on the opposite sides of the first principal component, which shows that countries from the new EU lag behind the old EU member states. The reasons behind this innovation gap can be observed more precisely based on PC1. In order to achieve a high level of innovation performance, countries need an innovation system based on a research system, human resources, as well as finance and support.

This study demonstrates that the PCA method is a useful tool applied to compare innovation performance across EU countries. Considering the fact that governments play an important role in enhancing innovation capacities of an economy, the research results may enable policy-makers to assess and identify the member states' priority areas requiring the improvement of innovation performance, which concerns the new EU states to the largest extent. Moreover, the presented results may be useful for scientists and institutions who develop innovation indicators.

One limitation of the presented innovation ranking for EU countries is receiving different results when applying different innovation indexes, e.g. SII vs. GII. However, the methodology can also be used to combine information from multiple indicators. The second limitation is that the presented countries' ranking is based on variables explaining 68% of the general variance. The remaining 32% are excluded from the analysis. These limitations can, nevertheless, be overcome by using more principal components and applying the clustering method.

Exogenous factors such as economic crises, governmental changes, and political uncertainties may of course affect innovation, as is likely to do the current global crisis caused by the COVID-19 pandemic. On the other hand, the pandemic has also highlighted the importance of innovation, especially within the digital infrastructure and academia – industry cooperation involving, e.g., a multidisciplinary research on the COVID-19 vaccine. The complexity of innovation means the study on this phenomenon requires continuous development and additional research. The presented approach can be a useful tool for assessing countries' innovation performance and can be considered an alternative to unweighted averaging.

References

Acs, Z. J. (2002). *Innovation and the Growth of Cities*. Cheltenham, Northampton: Edward Elgar Publishing. http://www.stellenboschheritage.co.za/wp-content/uploads/Innovation-and-the-Growth-of-Cities-.pdf.

Bielińska-Dusza, E., & Hamerska, M. (2021). Methodology for Calculating the European Innovation Scoreboard – Proposition for Modification. *Sustainability*, *13*(4), 1–21. https://doi.org/10.3390/su13042199.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer Science+Business Media. http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf.

Davidescu, A. A., Paul, A. M. V., Gogonea, R. M., & Zaharia, M. (2015). Evaluating Romanian Eco-Innovation Performances in European Context. *Sustainability*, *7*(9), 12723–12757. https://doi.org/10.3390/su70912723.

Dutta, S., Lanvin, B., & Wunsch-Vincent, S. (Eds.). (2018). *Global Innovation Index 2018: Energizing the World with Innovation* (11th edition). Ithaca, Fontainebleau, Geneva: Cornell University, INSEAD, World Intellectual Property Organisation. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2018.pdf.

Dziallas, M., & Blind, K. (2019). Innovation indicators throughout the innovation process: An extensive literature analysis. *Technovation*, *80–81*, 3–29. https://doi.org/10.1016/j.technovation.2018.05.005.

Edquist, C., Zabala-Iturriagagoitia, J. M., Barbero, J., & Zofío, J. L. (2018). On the meaning of innovation performance: Is the synthetic indicator of the Innovation Union Scoreboard flawed. *Research Evaluation*, *27*(3), 196–211. https://doi.org/10.1093/reseval/rvy011.

European Commission. (2019a). *European Innovation Scoreboard 2019*. Luxembourg: European Union. https://doi.org/10.2873/877069.

European Commission. (2019b). *European Innovation Scoreboard 2019: Methodology Report*. Maastrich: Maastricht Economic and Social Research Institute on Innovation and Technology.

European Commission. (2020a). *European innovation scoreboard 2020 – Database* [Data set]. https://ec.europa.eu/docsroom/documents/41864.

European Commission. (2020b). *European innovation scoreboard 2020 – Executive summary*. https://ec.europa.eu/docsroom/documents/41903.

European Commission. (2020c). *European Innovation Scoreboard 2020: Methodology Report*. Maastricht: Maastricht Economic and Social Research Institute on Innovation and Technology.

Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer Science+Business Media. http://doi.org/10.1007/978-1-4614-7138-7.

Gray, V. (Ed.). (2017). *Principal Component Analysis: Methods, Applications and Technology*. New York: Nova Science Publishers.

Holgersson, T., & Kekezi, O. (2018). Towards a multivariate innovation index. *Economics of Innovation and New Technology*, *27*(3), 254–272. https://doi.org/10.1080/10438599.2017.1331788.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 1–16. http://doi.org/10.1098/rsta.2015.0202.

Kontolaimou, A., Giotopoulos, I., & Tsakanikas, A. (2016). A typology of European countries based on innovation efficiency and technology gaps: The role of early-stage entrepreneurship. *Economic Modelling*, *52, Part B*, 477–484. https://doi.org/10.1016/j.econmod.2015.09.028.

Kowalski, A. M. (2020). Dynamics and Factors of Innovation Gap Between the European Union and China. *Journal of Knowledge Economy*. https://doi.org/10.1007/s13132-020-00699-1.

Lee, K., & Lee, J. (2020). National innovation systems, economic complexity, and economic growth: country panel analysis using the US patent data. *Journal of Evolutionary Economics*, *30*(4), 897–928. https://doi.org/10.1007/s00191-019-00612-3.

Lever, J., Krzywinski., M., & Altman, N. (2017). Principal component analysis. *Nature Methods*, *14*(7), 641–642. https://doi.org/10.1038/nmeth.4346.

van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(86), 2579–2605. https://www.jmlr.org/papers/volume9/vandermaaten08a /vandermaaten08a.pdf.

Makkonen, T., & van der Have, R. P. (2013). Benchmarking Regional Innovative Performance: Composite Measures and Direct Innovation Counts. *Scientometrics*, *94*(1), 247–262. http://doi.org /10.1007/s11192-012-0753-2.

Malik, K., & Jasińska-Biliczak, A. (2018). Innovations and Other Processes as Identifiers of Contemporary Trends in the Sustainable Development of SMEs: The Case of Emerging Regional Economies. *Sustainability*, *10*(5), 1–17. https://doi.org/10.3390/su10051361.

Malik, K., Mach, Ł., Szewczuk-Stępień, M., & Bebenek, P. (2020). Specialization Integrated Strategy of Innovations: Effective Model for Emerging Regional Economy Development? *European Research Studies Journal*, *23*(2), 78–97. https://doi.org/10.35808/ersj/1581.

Mtar, K., & Belazreg, W. (2020). Causal Nexus Between Innovation, Financial Development, and Economic Growth: the Case of OECD Countries. *Journal of the Knowledge Economy*, *12*(1), 310–341. http://doi.org/10.1007/s13132-020-00628-2.

Murphy, K. P. (2021). *Probabilistic Machine Learning: An introduction*. Cambridge: MIT Press.

Onea, I. A. (2020). Innovation Indicators and the Innovation Process – Evidence from the European Innovation Scoreboard. *Management* & *Marketing. Challenges for the Knowledge Society*, *15*(4), 605–620. http://doi.org/10.2478/mmcks-2020-0035.

Organisation for Economic Co-operation and Development. (2005). *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data* (3rd edition). Paris. https://ec.europa.eu/eurostat /cache/metadata/Annexes/inn_cis8_esms_an3.pdf.

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods* (8th edition). Ames: Iowa State University Press.

Stojanovska, S., & Madzova, V. (2018). Differences in efficiency of innovation performance among EU candidate countries and EU 28 countries. *ILIRIA International Review*, *8*(1), 117–144. https://doi.org/10.21113/iir.v8i1.388.

Szopik-Depczyńska, K., Kędzierska-Szczepaniak, A., Szczepaniak, K., Cheba, K., Gajda, W., & Ioppolo, G. (2018). Innovation in sustainable development: an investigation of the EU context using 2030 agenda indicators. *Land Use Policy*, *79*, 251–262. https://doi.org/10.1016/j.landusepol. 2018.08.004.

Turk, M. A., & Pentland, A. P. (1991). Face recognition using eigenfaces. *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 586–591). https://doi.org/10.1109/CVPR.1991.139758.

Van den Berg, H. (2016). *Economic growth and development* (3rd edition). Singapore: World Scientific Publishing Company.

Zabala-Iturriagagoitia, J. M., Aparicio, J., Ortiz, L., Carayannis, E. G., & Grigoroudis, E. (2020). The productivity of national innovation systems in Europe: Catching up or falling behind?. *Technovation*, *102*. https://doi.org/10.1016/j.technovation.2020.102215.

Zihao, G., Qibin, J., Sheng, C., & Ben, N. (2020). A New Innovation Index System Based on Principal Component Analysis: Evidence from Shenzhen. *ICEME 2020: 2020 The 11th International Conference on E-business, Management and Economics*. New York: Association for Computing Machinery. https://doi.org/10.1145/3414752.3414777.