Categorisation of regions in the European Union based on smart and inclusive growth indicators for the Europe 2020 strategy

Tímea Győri
Hungarian University of Agricultural and Life Sciences,
Hungary
E-mail: gyoritimi05@gmail.com

Interpreting regional differences is crucial in promoting regional and cohesion policies. The different levels of development of the regions, their various historical and geographical factors, as well as diverse political attitudes, institutional conditions, and their interactions fundamentally determine the economic convergence processes in the regions. Recent research has emphasised a pluralistic approach in which composite measures of progress and well-being are prioritised.

The primary goal of the research was to create a more complex region categorisation model than the traditional gross domestic product (GDP-) based categorisation. This study analyses the changes and correlations of the Europe 2020 Strategy's smart and inclusive growth indicators at the regional level between 2009 and 2019. To create the region categorisation model, the indicators of the strategy were concentrated and then expanded with other background variables into composite indicators, thus creating the main components: 'Relative deprivation' and 'Innovation environment'. The regions were categorised along the dimensions of the new latent variables to form the outstanding, catching-up, and lagging groups, which can be generally characterised by development, catching-up, and lagging, respectively. The outstanding group typically includes the capital of the member states and the regions of their agglomeration. The catching-up and lagging groups in chiefly consist of member states that joined after 2004. The lagging group are mostly external border regions and can be considered underdeveloped along both dimensions; in addition to their below-average innovation environment, their labour market position also lags behind. The model developed in this research provides a more nuanced picture of regional development differences than the traditional GDP-based categorisation.

Keywords:
Europe 2020 Strategy,
cohesion policy,
region categorisation,
principal component analysis,
cluster analysis,
spatial autocorrelation

Introduction

Several studies have examined the relationship between economic growth and regional inequality. According to Henderson et al. (2018), in addition to geographical characteristics, territorial inequalities in the distribution of economic activities are determined by variables related to the level of development. The geographic features include the location of the region (central, peripheral) and its larger centre. The capital may appear as a positive factor as these cities tend to concentrate on a number of economic and social activities and act as the administrative and economic centre of the country (Butkus et al. 2018).

In contrast to the positive impact of the capital, being situated at a border can be a barrier. Border regions are typically characterised by a geographical peripheral location (Egri–Kőszegi 2018, 2020) far from the country's centre of power and administration and close to regions belonging to other countries. Consequently, they may suffer from the border effect, i.e. limited interregional economic interactions due to differences in cultures, regulations, and business norms and higher transaction costs between countries (Capello et al. 2018).

Empirical research has yielded conflicting results on the effects of inequality on economic growth, depending on the data and methodology used. The negative impact of inequality on growth is mainly supported by cross-country studies (Alesina–Rodrik 1994, Persson–Tabellini 1994) while several other studies have shown a positive relationship (Li–Zou 1998, Forbes 2000).

In the European Union (EU), cohesion policy is the main instrument for reducing regional disparities, strengthening social and territorial cohesion, and promoting harmonious territorial development (EC 2007). Cohesion policy is based on the principle that everyone benefits from reducing the income gap between the richer and poorer regions and improving the well-being of their inhabitants (Nyikos 2017, Zdanowska et al. 2020). Assessing the impact of cohesion policy in general is not straightforward and involves a number of significant challenges as it is a rather diverse policy that seeks to address different economic and social objectives (Fratesi 2016). Several studies emphasise the importance of context and 'conditioning factors' (quality of government, absorption capacity of regions, territorial capital, etc.) in determining the impact of cohesion policy (Kearns–Reid-Henry 2009, Perger 2010, Dall'Erba–Fang 2017, Percoco 2017, Giordano 2017).

Enlargement in the 2000s (the accession of ten countries in 2004 and Bulgaria and Romania in 2007) necessitated a strengthening of the harmonisation process. Cohesion policy for 2007–2013 was reorganised according to three objectives: convergence, regional competitiveness and employment, and European territorial cooperation (Fratesi–Wishlade 2017).

As a result of the economic downturn in 2008, which began with the collapse of the US real estate market, economic growth in the EU lagged behind that of the rest of the world. As a result of the crisis, which also spread to the labour market, economic growth and employment growth trends in the EU have been interrupted, and inequalities between regions have worsened (EC 2010). GDP and employment growth in the 20–64 age group fluctuated significantly between 2000 and 2020 in the EU. The crisis hit member states to varying degrees. The Baltic countries, particularly Spain and Ireland, have been hit hardest by declining labour demand due to the economic downturn. The highest unemployment rate was recorded in Lithuania in 2009 (17.3%), followed by Spain (17.2%), Latvia (13.7%), Estonia (13.3%), and Ireland (12.0%). In Spain, which has been significantly impacted by the crisis, unemployment rose sharply, and by 2013 the unemployment rate had reached 27.3%.

Reform efforts introduced in the EU have contributed to job-creating economic growth. Unemployment rates for the 20–64 age group returned to pre-crisis levels in 2018 but remained high in many member states (EC 2018b). In the year before the coronavirus epidemic (2019), Greece (17.3%), Spain (13.8%) and Italy (9.9%) were still experiencing significant unemployment.

The economic crisis has highlighted the interdependence of our national economies (reforms implemented in one country also affect the performance of others) and our reflection on the crisis that we are much more effective in enacting change when we work together (EC 2013).

The next section explains the purpose of the research, followed by a description of the methodology of the study and the range of data used. Finally, the results of the study and the conclusions that can be drawn from them are presented.

Research objective

This study deals with the eligibility of the EU's Nomenclature of Territorial Units for Statistics (Nomenclature des Unités Territoriales Statistiques – NUTS) 2 level regions for support. Specifically, it aims to explore whether a more complex regional categorisation model based on the 'Europe 2020' strategy indicators for 2019 would result in a significant shift from GDP-based categories. According to the hypothesis related to the research question, a composite model can be created based on the indicators of the Europe 2020 Strategy, which expresses the development differences between regions in a more complex way and allows a more nuanced delimitation. It is posited that the constructed composite model will identify a larger number of regions as catching-up and lagging areas than the traditional entitlement categorisation.

For a sustainable future, the European Commission has set out its Europe 2020 growth strategy for the period 2010–2020. The Europe 2020 Strategy is the long-term orientation of EU countries toward smart, sustainable, and inclusive growth (Müller-Frączek 2019, Walesiak et al. 2021). The objectives of the strategy were formulated by considering the factors that influence economic growth. Numerous empirical studies have demonstrated that the human capital variable can explain a significant proportion of the variance in per capita GDP between countries (Mankiw et al. 1992). According to experts, the most important element of the long-term solution to economic problems is the drastic increase in the education and knowledge of the population (Jankó 2010, Alpek–Tésits 2019, Hajdú 2020). The growth of higher education has a positive impact on regional macroeconomic growth (Koltai–Filó 2021) and positively influences the development of regional economic and social cohesion (Canal Domínguez 2021). A higher level of education can provide better employment opportunities (Hajdú–Koncz 2021, 2022). Moreover, by increasing employment rates, poverty can be alleviated and general 'quality of life' increases (Egri et al. 2009, Egri 2017).

The EU cohesion policy and budget for the period 2014–2020 were set in line with the objectives of the Europe 2020 Strategy. More than a third of the European budget has been invested in key areas (EC 2015). Financial instruments associated with the objectives are the European Regional Development Fund (ERDF), the *European Social Fund (ESF)*, and the Cohesion Fund (CF).

The smart and inclusive growth targets set out in the strategy are as follows:

- spend 3% of EU GDP on research and development (R&D),
- reduce early school drop out to below 10% across the EU,
- raise the proportion of people with a tertiary degree to 40% in the 30–34 age group,
- raise the employment rate of the 20–64 age group to 75%, and
- reduce the number of people living in poverty and social exclusion by 20 million.

Given that inequalities between regions have been further exacerbated by the crisis, most of the Structural and Investment Funds (ERDF, ESF, CF, *European Agricultural Fund for Rural Development*, and *European Maritime and Fisheries Fund*) and programs have specifically targeted less-developed regions (Malloy 2010, Jones et al. 2020), promoting the convergence of these regions toward more developed regions (Panzera–Postiglione 2021). Several studies have shown that, although the impact is not uniform, European cohesion policy has positively contributed to economic growth in lagging areas (Gagliardi–Percoco 2016, Fratesi–Wishlade 2017). In particular, the performance of rural areas close to metropolitan agglomerations has improved. Favourable geography and the progressive suburbanisation of rural landscapes have created new opportunities for rural areas close to cities.

From a policy point of view, the categorisation of regions by level of development is certainly relevant. For the 2007–2013 budget periods, the eligibility of regions at the NUTS2 level was determined based on a threshold of 75% of the EU average GDP per capita (EC 2006). The allocation of resources for the 2014–2020 budget period continued to be based on the premise that support was provided to the most deprived regions (Nyikos 2017). However, in the name of a political compromise, the category of transition regions was introduced. The introduction of the transitional category sought to maintain an interest in cohesion policy in member states with essentially developed regions (and net contributors). Therefore, the regions in the 2014–2020 budget period were ranked (EC 2020a) and divided into three groups (Figure 1):

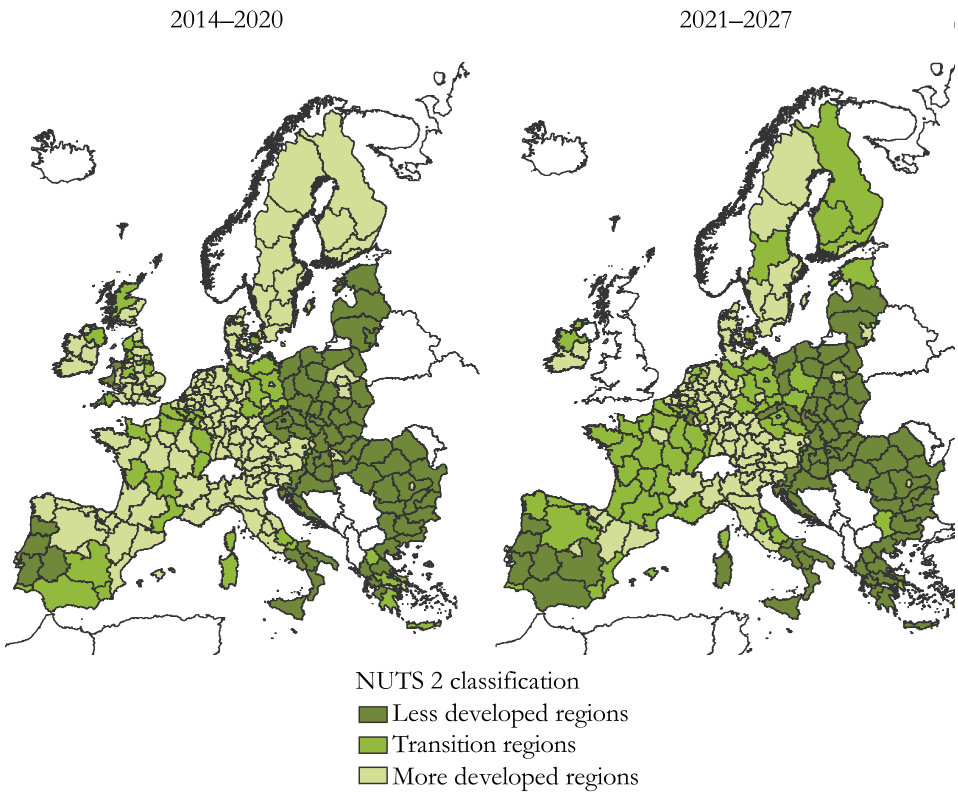
- less-developed regions (GDP per capita is less than 75% of the EU average),
- transition regions (GDP per capita 75–90% of the EU average), and
- more developed regions (GDP per capita exceeded 90% of the EU average).

When classifying regions according to their level of development, we must not forget that the EU's regional division has been revised and amended several times over the years. With the 2013 amendment, the number of regions was expanded to 276. At that time, the most significant change was the accession of Croatia to the NUTS classification. A total of 281 statistical regions were created in the EU with the 2016 revision of NUTS (EC 2018a, 2020b). These changes affected six member states: Ireland, France, Lithuania, Poland, the United Kingdom, and Hungary, where Central Hungary (HU10) split into Budapest (HU11) and Pest (HU12). The latest amendments cover the budget period 2021–2027. The new nomenclature covers 27 member states of the EU and the United Kingdom. The NUTS2021 classification includes 283 regions at the NUTS2 level, and a new amendment concerning regions in Croatia (Eurostat 2021). In addition, a number of amendments have been made which have changed not only the number of regions but also their boundaries and, simultaneously, their population.

For the budget period 2021–2027, the European Commission has proposed changing the previous threshold for transition and more developed regions. According to the proposal, regions with a GDP per capita between 75% and 100% of the EU average should be considered transition regions. This change will significantly increase the population in the transitional category from approximately 15% to more than 25% of the EU's total population (EC 2020a). As a result of these new regulations, a significant number of previously identified regions have been transferred to the transitional category (e.g. France, Germany, Spain, and Finland). The rearrangements between the categories were determined not only by changes in the regulations but also by changes in the economic situation of the regions. The economic performance of several regions in the Czech Republic and Poland has improved whereas the situation in some regions in Spain and Greece has deteriorated. In Hungary, after the division of the Central Hungary region, which previously belonged to the more developed category, only Budapest was considered more developed while Pest County was ranked among the less-developed regions.

Figure 1

ERDF and ESF eligibility



González et al. (2014) stated that NUTS2 level GDP maps are a good illustration of regional disparities, but territorial cohesion policy needs to go beyond GDP. Recent research emphasises the plural approach (Jones et al. 2020), with composite measures of progress and well-being gaining prominence (Walesiak et al. 2021). The structure of complex indicators involves several subjective decisions. Pasimeni (2012) proposed the development of a synthetic, composite 'Europe 2020 index' to assess the effects of the Europe 2020 Strategy and to determine its level of development, which allows comparisons between member states. Rappai (2016) developed a complex indicator that also considers correlations, thus measuring progress in a complex manner. Fura et al. (2017) developed synthetic indicators such as a positioning measure based on the median of the indicators. The composite indicator developed by Becker et al. (2020) aggregates the differences between the achieved values of each member state or region and the target values set in the strategy.

As part of this research, I examine the change in the smart and inclusive indicators of the Europe 2020 Strategy (intramural R&D expenditure by percentage of GDP, employment rate, early leavers from education and training, people at risk of poverty or social exclusion, and tertiary educational attainment) between 2009 and 2019. The year 2009 reflects a state of the crisis in the labour market and the return to pre-crisis levels in 2019.

As a first step in the categorisation of regions based on 2019 data, a correlation study will explore the strength and direction of the correlations between the indicators and other selected background variables. Subsequently, I compress the information content of the variables involved in the research into composite indicators using principal component analysis (PCA). I examine the degree of clustering determined by the new latent variables and the spatial arrangement of the possible clusters using a spatial autocorrelation procedure. Finally, I arrange the regions into homogeneous groups using cluster analysis along the dimensions of the new latent variables. Ultimately, the results obtained are compared with traditional GDP-based categorisation.

Data and method

Methodology

During the preparation of this study, the main focus was on the analysis of statistical data collected from secondary sources. The data used were obtained from the 2009 and 2019 NUTS2 data files of the Eurostat database. An exception was the intramural R&D expenditure (percentage of GDP) database, for which only 2018 data were available. The territorial basis for the analysis was provided by the regions of 28 EU member states. This study was based on the NUTS2 (regional) territorial levels defined according to the NUTS 2016 statistical regionalisation system.

To perform regional typing based on the 2019 data, I examined the possibilities of condensing the European 2020 indicators and other background variables into a composite indicator. To group the variables of the imputed basic database and compress the content of the variables with minimal information loss, I performed a factor analysis after Z standardising the data. PCA was used as the extraction method. In the analysis, the results of the correlation study were the starting point because PCA is based on the Pearson correlation matrix. The essence of the method is the existence of linear correlations between the individual variables, based on which new indicators are formed by linear regression. From the set of linearly correlated variants in pairs, uncorrelated principal components are generated by orthogonal transformation (Zou et al. 2006, Kovács 2014, Liu et al. 2019) such that the first few components describe a fairly large fraction of the total variance squared of the variables (Jolliffe 2005).

A variable should be omitted from the analysis if the proportion it explains is too low. If the communality (multiple coefficients of determination) is less than 0.25, the variable does not correlate strongly with any main component (Kovács 2014).

PCA was performed using several parameters and components. To ensure the validity of the factorisation, the Kaiser–Meyer–Olkin (KMO) measurement and Bartlett's spherical test were also performed. The KMO criterion was used to assess the suitability of the variables for factor analysis. Kaiser (1981) originally recommended that the baseline criterion for factorability should be 0.50. Based on Kaiser's recommendation, the KMO index was interpreted as follows: KMO≥0.9 excellent, KMO≥0.8 very good, KMO≥0.7 adequate, KMO≥0.6 moderate, KMO≥0.5 poor, and KMO<0.5 unacceptable. The basic hypothesis of Bartlett's chi-square test is that the original variables are independent (Arsham–Lovric 2011); our variables are suitable for factor analysis if the homogeneity test hypothesis can be rejected.

The number of significant principal components was operationalized directly on the one hand, and by providing own values (Eigenvalues), based on the Kaiser criterion, on the other. For a factor to have positive Kuder–Richardson reliability (Cronbach's alpha), it is necessary and sufficient for the associated eigenvalue to be greater than 1 (Kaiser 1960). Finally, variables with a Cronbach's alpha greater than 0.70 were included in the PCA.

The clustering of latent variables generated by PCA was also examined using spatial autocorrelation and hierarchical cluster analysis. In the research framework, I also compiled global and local Moran I statistics from spatial autoregressive modelling procedures. The basic assumption of spatial autocorrelation statistics is that the spatial distribution of the data follows a random pattern (Tóth 2014). As a result of the test, an index and a Z value were obtained. A statistically significant positive Z value indicates a spatial grouping of data whereas a negative Z value indicates data scatter. Values close to zero indicate a random spatial distribution of the data (Dusek 2004).

Local Moran I is a spatial autocorrelation study that informs about the spatial distribution of inequalities together with GIS visualisation (Tóth 2003), ignoring the extent of the differences. Local Moran I assigns a numerical value to each region, the expected value of which is 0, so if we get a value significantly different from 0, it indicates a regularity that can be discovered in the spatial arrangement. Two types of regular arrangements have been distinguished (Anselin 1995). In the case of a positive autocorrelation, the data of the adjacent area units are similar; in the case of a negative autocorrelation, the adjacent areas differ (Nemes Nagy 2005). In the case of a non-autocorrelation, the individual values are distributed randomly and the spatial differences do not draw a regular spatial shape. Neighbourhood relations are based on queen contiguity, with 999 permutations. The Moran scatter plots collectively represent the standardised values of the variables and their associated local Moran I values. The four plane quarters of the scatterplot represent the high–high, low–low, high–low, and low–high groups (Tóth 2003).

In the empirical research, the regions were grouped along the dimensions of the established principal components using a hierarchical cluster analysis procedure. Cluster analysis is suitable for arranging (clustering) data arrays into homogeneous groups; therefore, it essentially functions as a dimension-reducing method (Rao 1971, Tésits et al. 2021, Kökény–Kiss 2021). The essence of clustering is that the data within each cluster are similar in some dimensions; in this respect, they also differ from the elements of other clusters (Bardhoshi et al. 2021). The hierarchical cluster analysis procedure was performed using the Ward method, which is less sensitive to outliers, the essence of which is to reduce the internal heterogeneity of the clusters to be created (Székelyi–Barna 2008). The square Euclidean distance was used to measure the distance between clusters. The new variable obtained by clustering (CLU) was nominal, which allows the use of the nonlinear association coefficient (ETA). ETA^2 is the quotient of the sum of squares between the groups and the total deviation. The distinctive power of the variables was also assessed using an analysis of variance (ANOVA) standard deviation resolution table. As a control, the rank-based Kruskal–Wallis test was performed as an effective alternative to the one-way ANOVA (Breslow 1970, Vargha–Delaney 1998).

Database

The initial basic database of the research included five main indicators of smart and inclusive growth of the Europe 2020 strategy and 17 other indicators (background variables). The background variables were selected in relation to the topics of the five main indicators, considering the correlations between them.

For four of the selected indicators of the Europe 2020 Strategy (intramural R&D expenditure by percentage of GDP, tertiary educational attainment, early leavers from education and training, and people at risk of poverty or social exclusion), data were missing for 2019, which is 8.732% of the total database. In the case of the examined variables, a general lack of data patterns can be detected, i.e. the pattern does not have any specialty (Oravecz 2008).

When compiling the initial basic database, there was a significant lack of data on several background variables. Data gaps typically stemmed from the voluntary nature of data collection, which affected a significant proportion of the entire database; therefore, the final database operated with only a very narrow dataset. The potential background variables for the analysis were filtered based on the missing amount of data; the lack of data classified as significant (> 20%) was an exclusionary reason. The lack of data was significant for the following indicators: regional GDP (purchasing power standard [PPS] per inhabitant), job vacancy, impact of social transfers, motorways (km/1000 km^2), less than primary or primary education rate (25–64 years), people living in households with very low work intensity, severely materially deprived people, at risk of poverty, and individuals regularly using the Internet.

Based on the above criteria, 8.099% of the values were missing from the basic database containing 14 variables. The method of dealing with data gaps was multiple imputation (MI), which was first proposed by Rubin (1987). Multiple imputation is one of the most widely used missing data management techniques (Chung–Cai 2018) and can be applied to virtually any data structure or model type (Allison 2003). Ginkel et al. (2014) demonstrated that MIs can be safely applied in the context of PCA. Multiple imputation has already been shown to be effective with a small number of imputations, depending on the percentage of missing data (Allison 1999). In general, we can use imputation for variables where a maximum of 30%–40% of the data per variable is missing, but the lack of data in the entire database does not exceed 10%–15%. During imputation, a linear regression model was created to address missing data, with other complete variables as predictors.

In the research framework, correlations between the indicators of the basic database and their strengths were explored by means of correlation analysis. Pearson's correlation analysis was performed to examine the correlation between the baseline and imputed database variables. The value of the coefficient varies between +1 and -1. The coefficient is strong in the range of 0.7 to 1 in absolute values, medium at intervals of 0.3 to 0.7, and weak at 0–0.3 intervals (Nemes Nagy 2005). Depending on the results obtained, the selected significance levels were 1% and 5% ($p= 0.01$ and $p= 0.05$), respectively. The strength of the relationships showed a minimal improvement of a few hundred percent due to imputation of some indicators. In the next stage of the research, the basic data for the analysis were provided by the imputed database.

Descriptive statistics

The basic descriptive statistics (minimum, maximum, mean, and standard deviation) of the variables from the basic imputed database are listed in Table 1.

The extent of the differences between the minimum and maximum values of the variables suggests significant differences in development between regions. Based on the extent of deviation from the expected value, the values in some regions differ significantly from the average.

Table 1

Descriptive statistics of examined indexes, 2019

Examined indicators	Minimum	Maximum	Mean	Std. dev.
Intramural R&D expenditure (% of GDP)	0.09	8.52	1.639	1.205
GDP at current market prices (EUR/capita)	5,400	102,200	30,961	14,758
Income quintile share ratio, S80/S20 (%)	2.80	9.10	4.686	0.958
Life expectancy at birth (years)	73.70	85.80	81.116	2.440
Early leavers from education and training (% of population aged 18–24)	1.70	27.20	10.405	4.728
Tertiary educational attainment (% of population aged 30–34)	16.30	78.60	40.206	11.221
Lifelong learning – Adult participation age group 25–64 (% of population)	0.60	35.80	11.408	7.139
Employment rate age group 20–64 (% of population)	43.30	85.10	74.074	7.957
Activity rate age group 20–64 (% of population)	57.70	91.0	80.457	5.202
Employed in the high-technology sector (% of total employed)	0.80	11.90	3.832	2.045
People at risk of poverty or social exclusion (% of population)	7.90	49.70	20.279	7.362
Unemployment rate age group 20–64 (% of population)	0.84	29.80	6.270	5.147
Long-term unemployment, 12 months or more (% of unemployed)	11.30	84.40	36.422	14.361
Less than primary, primary and lower secondary education, levels 0–2 (% of unemployed)	1.96	45.50	11.411	7.179

Results

Change and correlation of Europe 2020 indicators

The achievements of the targets set in the Europe 2020 Strategy are diverse. Key indicators improved significantly between 2009 and 2019. Among the indicators relevant to the research (Table 2), the educational target values (the proportion of early school leavers and those with tertiary education) were met in 2019 at the EU level. Expenditures on R&D as a share of GDP rose to 2.14% while the employment rate stood at 73.9%.

At the regional level, a positive change can be observed in the field of employment: the employment rate for the 20–64 age group shifted predominantly to 70%–85% by 2019. The proportion of people with tertiary education also rose and was concentrated at approximately 30%–60%. Overall, the rate of early school leavers also improved compared to that in 2009, at around 10% in most regions.

Table 2

Headline indicators of Europe 2020 Strategy at national level

Headline indicators	EU28 (2009)	EU targets	EU28 (2019)
Intramural R&D expenditure (% of GDP)	1.93	3	2.14
Early leavers from education and training (% of population aged 18–24)	14.2	<10	10.3
Tertiary education attainment (% of population aged 30–34)	32.3	≥40	41.6
Employment rate age group (% of population aged 20–64)	68.9	75	73.9
Poverty and social exclusion (% of population)	23.8	–	21.4

The correlation calculation between the values of the Europe 2020 indicators examined at the regional level in 2009 and 2019 naturally showed a strong positive correlation, with all correlation coefficients above 0.7.

The coefficients of the indicators of R&D, employment, skills, and impoverishment defined in the strategy and the selected background variables for 2019 are presented in Table 3. In the correlation matrix, all coefficients were significant at the $p= 0.01$ level. No explicit correlations were found between these indicators. A moderately strong positive correlation ($r= 0.373$) was found between intramural R&D expenditures (% of GDP) and employment. It thus follows that the coefficient of determination is also very low, $R^2= 13.91\%$; that is, R&D expenditures explain only 13.91% of the variance in employment. Among the Europe 2020 indicators, the people at risk of poverty or the social exclusion indicator have the most correlations, and the employment rate shows the strongest correlations. There is a negative, moderately strong correlation between poverty and R&D expenditures ($r= –0.306$) and employment ($r= –0.646$) and the proportion of those with tertiary education and early school leavers ($r= –0.453$).

The coefficients calculated for the background variables were significant at the 95% and 99% confidence intervals. For most variables, medium correlations were observed. The employment rate clearly shows a strong correlation with the activity rate, unemployment rate, and the indicator of less than primary, primary, and lower secondary education (levels 0–2) percentage of the unemployed. There was a moderately strong correlation between the proportion of people with tertiary education and the proportion of people employed in the high-tech sector. There was also a moderately strong correlation between people at risk of poverty or social exclusion and the unemployment rate, as well as the income quintile share ratio (S80/S20). Life expectancy (years) showed the lowest number of and weakest correlations.

Table 3

Correlation of Europe 2020 indicators, 2019

Indicators	Intramural R&D expen-ditures	Early leavers from education and training	Tertiary educa-tional attainment	Employ-ment rate	People at risk of poverty or social exclusion
Intramural R&D expenditures	1	–0.275**	0.291**	0.373**	–0.306**
Early leavers from education	– 0.275**	1	–0.453**	–0.417**	0.493**
Tertiary educational attainment	0.291**	–0.453**	1	0.300**	–0.353**
Employment rate	0.373**	–0.417**	0.300**	1	–0.646**
People at risk of poverty	–0.306**	0.493**	–0.353**	–0.646**	1
GDP at current market prices	0.562**	–0.231**	0.457**	0.475**	–0.409**
Income quintile share ratio (S80/S20)	–0.175**	0.487**	–0.221**	–0.282**	0.607**
Life expectancy (years)	0.321**		0.259**		–0.196**
Activity rate (20–64 age group)	0.393**	–0.330**	0.366**	0.896**	–0.528**
Employed in high-tech sector	0.468**	–0.364**	0.601**	0.396**	–0.399**
Unemployment rate (20–64 age group)	–0.255**	0.452**	–0.150 *	–0.861**	0.611**
Long-term unemployment	–0.275**	0.309**	–0.416**	–0.709**	0.540**
Levels 0–2 education % of unemployed	–0.159**	0.380**	–0.147*	–0.675**	0.474**
Lifelong learning (25–64 age group)	0.362**	–0.190**	0.419**	0.345**	–0.334**

* Correlation is significant at the 0.05 level.
** Correlation is significant at the 0.01 level.

Principal component analysis

Possible data gaps limiting the factor-level analysis of Europe 2020 indicators and potential background variables were filled in by pre-filtering the variables and imputing gaps not exceeding 20%. The imputation had a minimal effect on the relationship between the original variables. The composite models in Table 4 were constructed using a narrow range of background variables that were not or were only slightly affected by the lack of data.

For the first model, which contains only the Europe 2020 indicators, the KMO value is 0.751, indicating that the variables are suitable for factor analysis. Medium-strength correlations existed between the variables. In the analysis, one main component was formed, the explanatory power of which was 51.3%, which was below the limit.

The variables of the second model were supplemented with indicators employed in the high-technology sector, unemployment rate, and less than primary or primary education levels (0–2) of unemployment. The KMO index was greater than 0.7; thus, the variables were proven suitable for factor analysis. This was also confirmed by Bartlett's chi-square test with a 99% confidence interval rejecting that the original variables were independent. In terms of communality, the two extremes were unemployment rate and R&D spending. The unemployment rate retains 91% of the original information and 43% of R&D expenditures. Two main components were generated in the composite model. The combined variance quotient of the main components was 68.1%. The first component included variables related to unemployment, employment, and the risk of poverty, and the second included indicators related to skills, R&D expenditures, and employment in the high-technology sector. Based on the information they compress, I refer to the main components as 'Relative deprivation' and 'Innovation environment'.

Table 4

Regional PCA models, 2019

Parameters	Model_1	Model_2	
KMO	0.751	0.766	
Sums of Squared Loadings	Component 1	Component 2/1	Component 2/2
Total	2.587	3.106	2.339
% of variance	51.741	38.821	29.231
Cumulative %	51.741	68.053	
Components	Principal component	Relative deprivation	Innovation environment
People at risk of poverty	–0.808	0.666	–0.415
Employment rate	0.785	–0.861	0.293
Early leavers from education	–0.744	0.443	–0.532
Tertiary educational attainment	0.648	0.049	0.836
Intramural R&D expenditures	0.587	–0.169	0.632
Employed in high-tech sector		–0.182	0.825
Levels 0–2 education % of unemployed		0.876	–0.035
Unemployment rate (20–64 age group)		0.945	–0.134

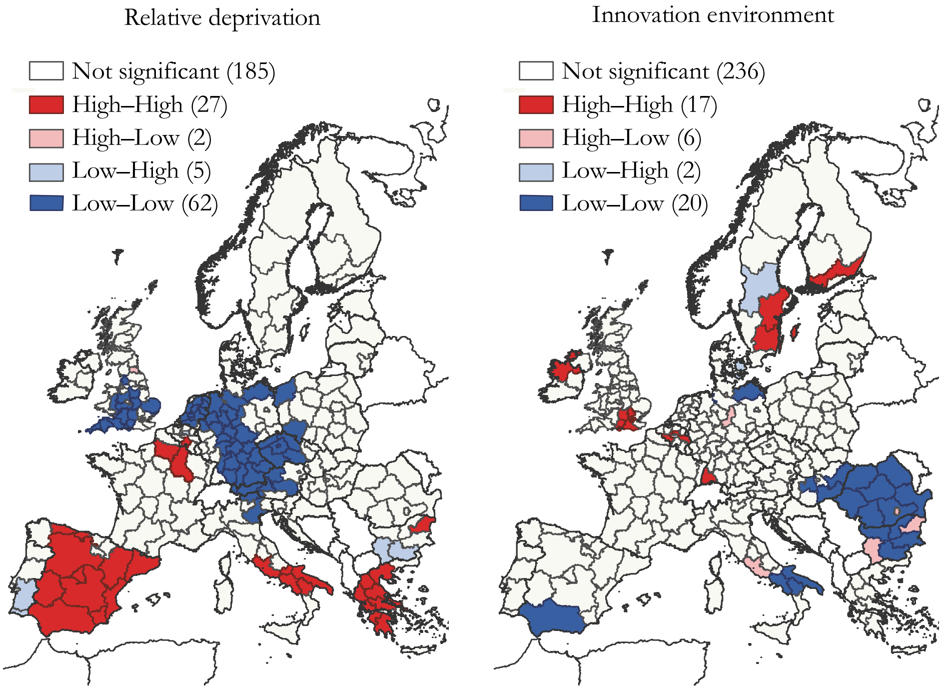
Spatial autocorrelation of principal component

The regional spatial inequality determined by the innovation environment and relative deprivation variables was revealed by global and local Moran's I analysis (Figure 2). The value of global Moran I for the principal component of relative deprivation was I=0.609, indicating strong clustering. In the case of the innovation environment, I=0.308, which confirms the existence of a smaller degree of clustering.

The spatial pattern of the main components of the relative deprivation and innovation environment was mapped using the local test function of spatial autocorrelation (local Moran I statistics). Only regions with a local Moran I value considered significant at the 99% confidence interval are shown on local indicator of spatial autocorrelation (LISA) maps. The maps can be used to identify the regions that contributed the most to the high Moran's index, which indicated high spatial autocorrelation. These are hot spot (high–high) regions that, with their high value, have similar above-average neighbours. Regions with lower-than-average values and neighbouring regions appear as cold spots (low–low).

Figure 2

LISA maps – 'Relative deprivation' and 'Innovation environment', 2019



For the relative deprivation principal component, 27 and 62 regions can be considered hot and cold spots, respectively. The hot spots were located in Spain, southern Italy, Greece, and France, indicating above-average unemployment in these regions. The regions identified as cold spots were concentrated in Central Europe and the United Kingdom. The analysis identified a minimal number of regions that are in stark contrast to neighbouring regions in terms of their labour market situation, of which five regions are low–high outliers (Portugal: Algarve [PT15], Alentejo [PT18], Belgium: Prov. Vlaams-Brabant [BE24], Bulgaria: Yugozapaden [BG41], Yuzhen tsentralen [BG42]), namely low-value regions for which neighbouring regions have a high value. These include the two southernmost regions of Portugal, the capital and neighbouring regions of Bulgaria, and the region surrounding the Belgian capital. In these regions, there is clearly a prevailing positive effect of capital. A further two regions appear as high–low outliers (Germany: Bremen [DE50], United Kingdom: Tees Valley and Durham [UKC1]); these are high-value regions for which the neighbouring regions have a low value, namely Tees Valley and Durham in the United

Kingdom and Bremen in Germany. While it is true that both regional centres were in their heyday during the industrialisation period (Durham – coal mining, Bremen – a free Hanseatic city), they have since been mainly stagnant.

The lower global Moran I value of the innovation environment principal component has a lower number of hot and cold spots. Only 17 were classified as high to high, ten of which were in the London and southern regions of the United Kingdom. The other high–high regions are isolated from each other, typically adjacent to the capital regions of Finland, Sweden, Ireland, and Belgium. This study identified a significant number of cold spots in Bulgaria, Romania, and Hungary. In addition, some regions in southern Spain and southern Italy had below-average innovation environments. Of the outlier regions identified, two belonged to the low–high (Sweden: Norra Mellansverige [SE31], Denmark: Sjælland [DK02]) cluster and six to the high–low cluster (Germany: Hamburg [DE60], Braunschweig [DE91], Italy: Lazio [ITI4], Bulgaria: Yugozapaden [BG41], Romania: Bucureşti–Ilfov [RO32], Croatia: Kontinentalna Hrvatska [HR04]). There were significant regional differences between the capital regions of Denmark and Sweden, and those adjacent to these regions. High–low outliers are typically regions that include capital, so their innovation environment is more developed than that of neighbouring regions.

Cluster analysis of principal components

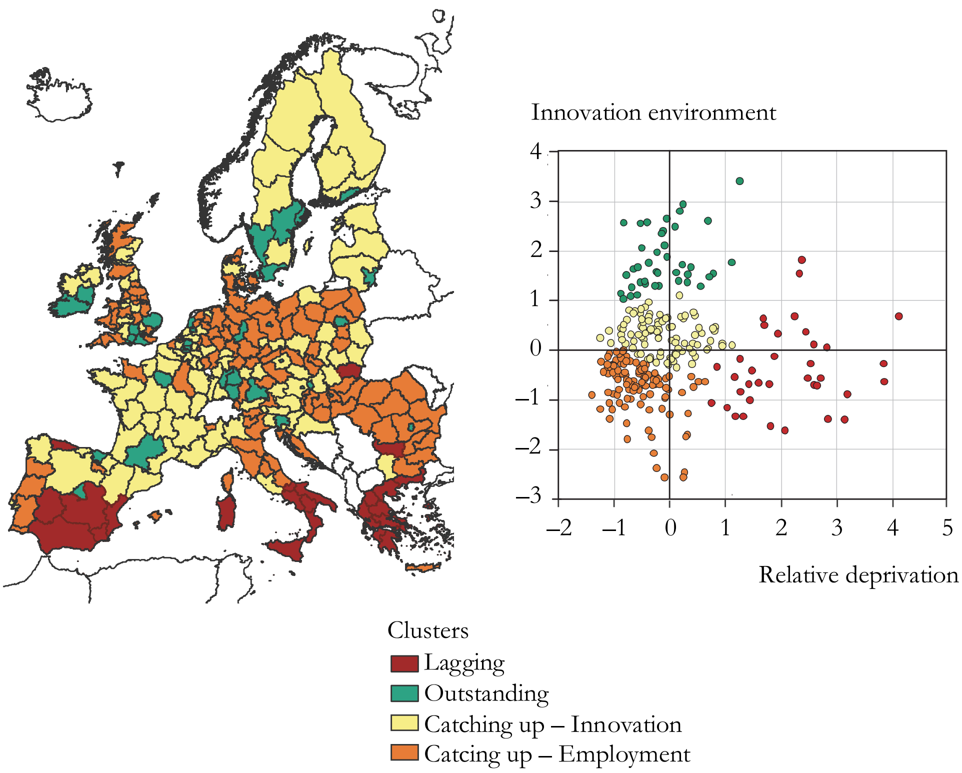
To facilitate the interpretation of the obtained results, hierarchical cluster analysis was performed to arrange and typify the regions into homogeneous groups using latent variables created by PCA. After structural analysis of the variables, four clusters were identifiable. Based on the F-test and the empirical significance level in the ANOVA table, the distinctive power of the principal components was found to be significant at $p < 0.01$ level. The Kruskal–Wallis test, performed as an alternative to the test, yielded the same result. The ETA coefficient calculated for the nominal index of the cluster was 0.841 and 0.852 for the first and second principal components, respectively. Based on the within-group variance, the formed clusters can be considered homogeneous. The variance in the latent variables within each group did not exceed the total standard deviation in either case.

The four clusters created were named outstanding, catching-up–innovation, catching-up–employment, and lagging, based on their characteristics according to the relative deprivation and innovation environment (Figure 3). When interpreting the data of the two latent variables, it should be noted that while the above-average value of the variable 'Innovation environment' is a measure of development, the above-average value of the variable 'Relative deprivation' is an indicator of lag.

The outstanding group includes regions that are considered to have developed in both dimensions. The regions belonging to the lagging group are characterised by a below-average innovation environment and a worse than average labour market situation. According to the relative deprivation dimension, catching-up–employment regions are not considered to be lagging behind; namely, their labour market situation is favourable while their innovation environment is below average. The regions in the catching-up–innovation group perform slightly above average in the innovation environment dimension, but significantly lag behind the outstanding regions. It is also important to point out that these regions, unlike the catching-up–employment group, show average performance in the other dimension.

Figure 3

Clusters of 'Relative deprivation' and 'Innovation environment' in EU, 2019



The spatial arrangement of the created regional categories was already indicated by spatial autocorrelation with latent variables. The outstanding group overlapped with hot spots with above-average innovation environments while cold spots indicated regions in the catching-up–employment and lag clusters. The area of the cold spot identified by spatial autocorrelation analysis of the relative deprivation factor can also generally be fitted to regions of catching-up groups while hot spots show a high degree of identity with the lagging cluster.

The catching-up and lagging groups typically consist of member states that joined after 2004. Of the clusters created, the number of catching-up groups was the highest, at 205 regions, 97 of which had an above-average innovation environment and 108 of which had a higher level of employment. The catching-up–innovation category typically includes Belgium, Luxembourg, Germany, Denmark, the Netherlands, Ireland, the United Kingdom, Spain, France, and Poland. Most of the regions in the catching-up–employment category are in Germany, Poland, Hungary, Bulgaria, Romania, northern Italy, and Portugal. The 32 regions of the Visegrad Group (V4) were included in the catch-up group. Hungary's regions, with the exception of Budapest, are part of the catching-up–employment cluster.

In terms of the dimensions of the main components, 40 regions were considered outstanding. These regions are typically the capitals of member states and their agglomerations (London, Brussels, Madrid, Berlin, Vienna, Prague, Warsaw, Budapest, Bucharest, Helsinki, Stockholm, etc.). The V4 has five outstanding regions: Budapest, Praha, Warszawski stoleczny, Makroregion Województwo Mazowieckie, and Bratislavský kraj.

The lagging group is considered to be underdeveloped along both dimensions; in addition to a below-average innovation environment, the labour market position also lags behind. These are typically regions in Greece, southern Italy, and southern Spain. Only one of the 38 regions belonged to the V4.

The regional categories developed in the research framework were compared with the GDP-based categorisation for the periods 2014–2020 and 2021–2027. While the regional categories overlap at several points, differences can be detected. The categorisation of regions based on the 2019 data created in the research framework identified a narrow range of regions as outstanding and lagging regions; therefore, these categories can be interpreted as extreme values. In both periods, almost all regions classified in the outstanding category were more developed while regions in the lagging category were less developed. A significant portion of the regions classified in the catching-up–innovation category was more developed in the period 2014–2020. The period 2021–2027 is considered a transitional region, so the amendment of the regulation mainly affected this category. Regions in the catching-up–employment category were typically underdeveloped in both funding periods.

Conclusions

Strengthening territorial cohesion is an important element of the Europe 2020 Strategy that seeks to extend the benefits of economic growth to peripheral areas. Exploring regional differences based on strategy indicators is of key importance for resource allocation decisions in future budget periods.

In the examined period, a slight spatial equalisation of strategy employment and qualification indicators can be observed; the scattering of the regions will occur at much smaller intervals in 2019, which confirms the decrease in regional differences. Thus, the reduction in inequalities between regions is expected to strengthen cohesion.

Unlike traditional GDP-based assessments, composite indicators can provide a more complex assessment of differences in regional development. PCA using the indicators of the Europe 2020 Strategy and involving additional background variables produced two composite indicators. Based on their information content, new latent variables called relative deprivation, and the innovation environment interact were created. However, synergy between dimensions does not mean that progress in one dimension can only occur at the expense of the other.

The existence of regional spatial inequalities, determined by the two latent variables, was confirmed by a spatial autocorrelation study. The analysis identified significant high–high and low–low clusters for both variables.

In the clusters of the final model created by cluster analysis of latent variables, the regions are well separated based on their differences in development (outstanding, catching-up, and lagging). Overall, the regions of the outstanding, catching-up, and lagging groups are characterised by development, catching-up, and lagging, respectively. The compiled map is an eclectic example of a polycentric pattern in which capital regions typically show strong performance. Regional differences in development level, an array of historical and geographical factors, as well as the different political attitudes, institutional conditions, and their interaction, basically govern the economic convergence processes in the region.

The outstanding group typically includes the capital of the member states and the regions of their agglomeration. These results are in line with the findings of previous studies that confirm the presence of the capital in the region has a positive impact on its economic growth.

The lagging group is considered to be underdeveloped along both dimensions; in addition to their below-average innovation environment, their labour market position is lags behind. Typical examples include Greece, southern Italy, and southern Spain, a significant portion of which are external border regions. The results also show a strong correlation with the findings in the literature on peripheral border regions; namely, a border location can negatively impact economic growth.

The more advanced and transition regions identified by the European Commission based on GDP per capita for the period 2014–2020 overlap significantly with the outstanding and catching-up–innovation groups established in the research. The less-developed regions that benefit most from EU structural and cohesion funds tend to fit into the lagging and catching-up–employment groups established by the research. The difference is striking in the case of southern Spain and Portugal, where areas identified as more developed and transitional regions according to the traditional classification have been identified as catching-up and lagging areas by research based on composite indicators.

The European Commission has proposed changes to previous thresholds for transition and more developed regions for the 2021–2027 budget period. According to this proposal, the threshold for GDP per capita in the transitional category will increase by 10%. The amendment of thresholds significantly rearranged the previous, more advanced and transitional region categories. Downgrading mainly affects regions in Finland, Germany, France, and Spain, which have been classified as catching-up–innovation. The regions in the catching-up–employment and lag categories remained underdeveloped in the period 2021–2027, according to the GDP-based categorisation.

In addition to the changes in the regulations, the rearrangements between the categories were determined on the one hand by changes in the NUTS system and on the other by changes in the regions' economic situation. Most of the changes in the NUTS system aimed at separating the more developed areas from their wider agglomeration (e.g. the division of the Central Hungarian region). The effects of the 2008 economic crisis varied to some extent in member states. In 2019, the majority of regions in the member states severely affected by the crisis (Greece, Spain, and Italy) were still experiencing significant unemployment. The deterioration in economic performance between 2009 and 2019 affected the classification of some regions in Spain and Greece, which subsequently affected the classification of several regions in the Czech Republic and Poland.

Research has shown that along the main dimensions of the Europe 2020 Strategy, a fault line can be identified in Europe, which is linked to the different effects of the 2008 economic crisis on individual regions.

Ultimately, the regional categorisation developed in the framework of the present study also confirmed the need for reclassification. With this change, the determination of eligibility for resources will be based on a categorisation that is more in line with the current developmental disparities. Awareness of regional disparities is of paramount importance in promoting regional and cohesion policies. The key to success lies in a complex approach that focuses on the specificities and needs of the areas. Exploring regional disparities based on European 2020 indicators will contribute to the development of sounder policies to strengthen resilience and will be key to resource allocation decisions in forthcoming budget periods, which will promote sustainable catching-up.

Overall, the EU has the capacity to act in times of crisis and adapt its economies and societies to change. The experience gained during crisis management can be used to address recessions in the near future. Europeans today must once again prepare for transformation to cope with the effects of crises, overcome the EU's structural weaknesses, and address growing global challenges. It is indisputable that coronavirus outbreaks in 2020 will affect the EU economy. As a result of the crisis, regional performance fragmentation may intensify. Due to the economic recession caused by the epidemic, meeting more of the Europe 2020 targets will be more challenging for member states than ever before.

REFERENCES

ALESINA, A.–RODRIK, D. (1994): Distributive politics and economic growth *The Quarterly Journal of Economics* 109 (2): 465–490. https://doi.org/10.2307/2118470

ALLISON, P. (1999): Multiple imputation for missing data: A cautionary tale *Sociological Methods & Research* 28 (3): 301–309. https://doi.org/10.1177/0049124100028003003

ALLISON, P. (2003): Missing data techniques for structural equation models *Journal of Abnormal Psychology* 112 (4): 545–557. https://doi.org/10.1037/0021-843X.112.4.545

ALPEK, B. L.–TÉSITS, R. (2019): A foglalkoztathatóság térszerkezeti és települési dimenziói Magyarországon *Területi Statisztika* 59 (4): 381–399. https://doi.org/10.15196/TS590402

ANSELIN, L. (1995): Local indicators of spatial association—LISA *Geographical Analysis* 27 (2): 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

ARSHAM, H.–LOVRIC, M. (2011): Bartlett's test. In: LOVRIC, M. (ed.): *International encyclopedia of statistics science* pp. 87–88., Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_132

BARDHOSHI, G.–UM, B.–ERFORD, B. T. (2021): Conducting a cluster analysis in counseling research: Four easy steps *Counseling Outcome Research and Evaluation* 12 (1): 54–62. https://doi.org/10.1080/21501378.2020.1768522

BECKER, W.–NORLEN, H.–DIJKSTRA, L.–ATHANASOGLOU, S. (2020): Wrapping up the Europe 2020 Strategy: A multidimensional indicator analysis *Environmental and Sustainability Indicators* 8: 100075. https://doi.org/10.1016/j.indic.2020.100075

BRESLOW, N. (1970): A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship *Biometrika* 57 (3): 579–594. https://doi.org/10.1093/biomet/57.3.579

BUTKUS, M.–CIBULSKIENE, D.–MACIULYTE-SNIUKIENE, A.–MATUZEVICIUTE, K. (2018): What is the evolution of convergence in the EU? Decomposing EU disparities up to the NUTS 3 level *Sustainability* 10 (5): 1–37. https://doi.org/10.3390/su10051552

CANAL DOMÍNGUEZ, J. F. (2021): Higher education, regional growth and cohesion: Insights from the Spanish case *Regional Studies* 55 (8): 1403–1416. https://doi.org/10.1080/00343404.2021.1901870

CAPELLO, R.–CARAGLIU, A.–FRATESI, U. (2018): Measuring border effects in European cross-border regions *Regional Studies* 52 (7): 986–996. https://doi.org/10.1080/00343404.2017.1364843

CHUNG, S.–CAI, L. (2018): Alternative multiple imputation inference for categorical structural equation modeling *Multivariate Behavioral Research* 53 (1): 323–337. https://doi.org/10.1080/00273171.2017.1406324

DALL'ERBA, S.–FANG, F. (2017): Meta-analysis of the impact of European Union structural funds on regional growth *Regional Studies* 51 (6): 822–832. https://doi.org/10.1080/00343404.2015.1100285

DUSEK, T. (2004): *A területi elemzések alapjai* Regionális Tudományi Tanulmányok 10., ELTE TTK Regionális Földrajzi Tanszék, Budapest.
EGRI, Z. (2017): Regional health inequalities in the European macroregion from the East Central European perspective *Regional Statistics* 7 (1): 197–224. https://doi.org/10.15196/RS0711
EGRI, Z.–KŐSZEGI, I. R. (2018): A gazdasági-társadalmi (komplex) térszerkezet kelet-közép-európai képe *Területi Statisztika* 58 (1): 27–56. https://doi.org/10.15196/TS580102
EGRI, Z.–KŐSZEGI, I. R. (2020): A közúti elérhetőség szerepe a kelet-magyarországi gazdasági teljesítményben és gazdaságfejlesztésben *Területi Statisztika* 60 (6): 653–687. https://doi.org/10.15196/TS600603
EGRI, Z.–TÖRŐCSIK, V.–TÁNCZOS, T. (2009): Regional HDI as a territorial and social differentiation index in Central Europe. In: KÁPOSZTA, J. (ed.): *New elements and research in spatial economy* pp. 167–180., Research Institute of J. Selye University, Komarno.
EUROPEAN COMMISSION (EC) (2010): Labour market and wage developments in 2009. In: *European Economy* Brussels. https://doi.org/10.2765/36582
EUROPEAN COMMISSION (EC) (2013): *Sustainable development in the European Union – 2013 monitoring report of the EU sustainable development strategy. Eurostat Statistical Books.* Publications Office of the European Union, Luxembourg. https://doi.org/10.2785/11549
EUROPEAN COMMISSION (EC) (2018a): *Regions in the European Union nomenclature of territorial units for statistics – NUTS 2016/EU-28* Publications Office of the European Union, Luxembourg. https://doi.org/10.2785/475524
EUROPEAN COMMISSION (EC) (2020a): *Eurostat regional yearbook – 2020 edition* Publications Office of the European Union, Luxembourg. https://doi.org/10.2785/98733
FORBES, K. J. (2000): A reassessment of the relationship between inequality and growth *American Economic Review* 90 (4): 869–887. https://doi.org/10.1257/aer.90.4.869
FRATESI, U. (2016): Impact assessment of EU cohesion policy: Theoretical and empirical issues. In: PIATTONI, S.–POLVERARI, L. (eds.): *Handbook on Cohesion Policy in the EU.* pp. 443–460., Cheltenham: Edward Elgar. https://doi.org/10.4337/9781784715670
FRATESI, U.–WISHLADE, F. G. (2017): The impact of European cohesion policy in different context *Regional Studies* 51 (6): 817–821. https://doi.org/10.1080/00343404.2017.1326673
FURA, B.–WOJNAR, J.–KASPRZYK, B. (2017): Ranking and classification of the EU countries regarding their levels of implementation of the Europe 2020 Strategy *Journal of Cleaner Production* 165: 968–979. https://doi.org/10.1016/j.jclepro.2017.07.088
GAGLIARDI, L.–PERCOCO, M. (2016): The impact of European cohesion policy in urban and rural regions *Regional Studies* 51 (6): 857–868. https://doi.org/10.1080/00343404.2016.1179384
GINKEL, J. R. V.–KROONENBERG, P. M.–KIERS, H. A. L. (2014): Missing data in principal component analysis of questionnaire data: A comparison of methods *Journal of Statistical Computation and Simulation* 84 (11): 2298–2315. https://doi.org/10.1080/00949655.2013.788654

GIORDANO, B. (2017): Exploring the role of ERDF in regions with specific geographical features: Islands, mountainous and sparsely populated areas *Regional Studies* 51 (6): 869–879. https://doi.org/10.1080/00343404.2016.1197387

GONZÁLEZ, R. C. L.–PAÜL, V.–TRILLO-SANTAMARÍA, J. M. (2014): The territorial dimension of "Europe 2020": Mapping European regions under the strategy to emerge from the crisis *Social Sciences* 3 (4-1): 13–27. https://doi.org/10.11648/j.ss.s.2014030401.13

HAJDÚ, D. (2020): A munkanélküliség és felnőttképzés területi eloszlása az Észak-magyarországi régióban *Észak-magyarországi Stratégiai Füzetek* 17 (2): 62–69. https://doi.org/10.32976/stratfuz.2020.14

HAJDÚ, D–KONCZ, G. (2021): Megélhetési tanulók a Borsod-Abaúj-Zemplén megyei felnőttképzésben *Területi Statisztika* 61 (2): 229–254. https://doi.org/10.15196/TS610206

HAJDÚ, D.–KONCZ, G. (2022): Employment data of participants in supported adult training for jobseekers and their territorial pattern in Hungary, 2010–2020 *Regional Statistics* 12 (2): 117–148. https://doi.org/10.15196/RS120205

HENDERSON, J. V.–SQUIRES, T. L.–STOREYGARD, A.–WEIL, D. N. (2018): The global spatial distribution of economic activity: Nature, history and the role of trade *The Quarterly Journal of Economics* 133 (1): 357–406. https://doi.org/10.1093/qje/qjx030

JANKÓ, B. (2010): A tudás szerepe a gazdasági növekedésben – az alapmodellek bemutatása *Statisztikai Szemle* 88 (1): 47–72.

JOLLIFFE, I. T. (2005): *Principal component analysis* 2nd ed. John Wiley & Sons. Wiley Online Library, New York. https://doi.org/10.1002/0470013192.bsa501

JONES, R.–GOODWIN-HAWKINS, B.–WOODS, M. (2020): From territorial cohesion to regional spatial justice: The well-being of future generations act in Wales *International Journal of Urban and Regional Research* 44 (5): 894–912. https://doi.org/10.1111/1468-2427.12909

KAISER, H. F. (1960): The application of electronic computers to factor analysis *Educational and Psychological Measurement* 20: 141–151. http://dx.doi.org/10.1177/001316446002000116

KAISER, H. F. (1981): A revised measure of sampling adequacy for factor-analytic data matrices *Educational and Psychological Measurement* 41 (2): 379–381. https://doi.org/10.1177/001316448104100216

KEARNS, G–REID-HENRY, S. (2009): Vital geographies: Life, luck and the human condition *Annals of the Association of American Geographers* 99 (3): 554–574. https://doi.org/10.1080/00045600902967177

KOLTAI, Z.–FILÓ, CS. (2021): A magyarországi városok telephelyi tényezőinek vállalati megítélése *Területi Statisztika* 61 (1): 79–104. https://doi.org/10.15196/TS610104

KÖKÉNY, L.–KISS, K. (2021): There is a time and a place for everything (and for everyone): Examining main socio-demographic and territorial differences in use of leisure time *Regional Statistics* 11 (2): 136–164. https://doi.org/10.15196/RS110206

LI, H.–ZOU, H. (1998): Income inequality is not harmful for growth: Theory and evidence *Review of Development Economics* 2 (3): 318–334. https://doi.org/10.1111/1467-9361.00045

LIU, Y.–SINGLETON, A.–ARRIBAS-BEL, D. (2019): A Principal Component Analysis (PCA)-based framework for automated variable selection in geodemographic classification *Geo-spatial Information Science* 22 (4): 251–264. https://doi.org/10.1080/10095020.2019.1621549

MALLOY, T. H. (2010): Creating new spaces for politics? The role of national minorities in building capacity of cross-border regions *Regional & Federal Studies* 20 (3): 335–351. https://doi.org/10.1080/13597566.2010.484572

MANKIW, N. G.–ROMER, D.–WEIL, D. N. (1992): A contribution to the empirics of economy growth *The Quarterly Journal of Economics* 107 (2): 407–437. https://doi.org/10.2307/2118477

MÜLLER-FRĄCZEK, I. (2019): Dynamic measurement of complex phenomena in assessing the Europe 2020 strategy effects *Regional Statistics* 9 (1): 32–53. https://doi.org/10.15196/RS090107

NEMES NAGY, J. (ed.) (2005): *Regionális elemzési módszerek* Regionális Tudományi Tanulmányok, 11., ELTE Regionális Földrajzi Tanszék, Budapest.

NYIKOS, GY. (2017): *Kohéziós politika 2014–2020, Az EU belső fejlesztéspolitikája a jelen programozási időszakban* Dialóg Campus Kiadó, Budapest

ORAVECZ, B. (2008): Hiányzó adatok és kezelésük a statisztikai elemzésekben *Statisztikai Szemle* 86 (4): 365–384.

PANZERA, D.–POSTIGLIONE, P. (2021): The impact of regional inequality on economic growth: A spatial econometric approach *Regional Studies* 56 (5): 687–702. https://doi.org/10.1080/00343404.2021.1910228

PASIMENI, P. (2012): Measuring Europe 2020: A new tool to assess the strategy *International Journal of Innovation and Regional Development* 4 (5):365–385. https://doi.org/10.1504/IJIRD.2012.048986

PERCOCO, M. (2017): Impact of European cohesion policy on regional growth: Does local economic structure matter? *Regional Studies* 51 (6): 833–843. https://doi.org/10.1080/00343404.2016.1213382

PERGER, É. (2010): Az EU kohéziós politika kormányzati irányításának magyar sajátosságai *Tér és Társadalom* 24 (1): 119–136. https://doi.org/10.17649/TET.24.1.1299

PERSSON, T.–TABELLINI, G. (1994): Is inequality harmful for growth? *American Economic Review* 84 (3): 600–621. https://doi.org/10.3386/w3599

RAO, M. R. (1971): Cluster analysis and mathematical programming *Journal of the American Statistics Association* 66 (335): 622–626. https://doi.org/10.1080/01621459.1971.10482319

RAPPAI, G. (2016): Europe en route to 2020: A new way of evaluating the overall fulfillment of the Europe 2020 strategic goals *Social Indicators Research* 129: 77–93. https://doi.org/10.1007/s11205-015-1092-5

RUBIN, D. B. (1987): *Multiple imputation for nonresponse in surveys* John Wiley & Sons. Wiley Online Library, New York.

SZÉKELYI, M.–BARNA, I. (2008): *Túlélőkészlet az SPSS-hez* (Negyedik kiadás), TYPOTEX, Budapest.

TÉSITS, R.–ZSIGMOND, T.–ALPEK, L.–HOVÁNYI, G. (2021): The role of endogenous capital factors in the territorial development of the Sellye District in Hungary *Regional Statistics* 11 (1): 58–77. https://doi.org/10.15196/RS110103

TÓTH, G (2003): Területi autokorrelációs vizsgálat a Local Moran I módszerével *Tér és Társadalom* 17 (4): 39–49. https://doi.org/10.17649/TET.17.4.914

TÓTH, G. (2014): *Térinformatika a gyakorlatban közgazdászoknak* (Egyetemi jegyzet), Miskolci Egyetemi Kiadó, Miskolc.

VARGHA, A.–DELANEY, H. D. (1998): The Kruskal-Wallis test and stochastic homogeneity *Journal of Educational and Behavioral Statistics* 23 (2): 170–192. https://doi.org/10.3102/10769986023002170

WALESIAK, M.–DEHNEL, G.–OBREBALSKI, M. (2021): Assessment of the Europe 2020 strategy: A multidimensional indicator analysis via dynamic relative taxonomy. *Energies* 14 (16): 4990. https://doi.org/10.3390/en14164990

ZDANOWSKA, N.–ROZENBLAT, C.–PUMAIN, D. (2020): Evolution of urban hierarchies under globalisation in Western and Eastern Europe *Regional Statistics* 10 (2): 3–26. https://doi.org/10.15196/RS100202

ZOU, H.–HASTIE, T.–TIBSHIRANI, R. (2006): Sparse principal component analysis *Journal of Computational and Graphical Statistics* 15 (2): 265–286. https://doi.org/10.1198/106186006X113430

INTERNET SOURCES

EUROPEAN COMMISSION (EC) (2006): *Guidelines on national regional aid for 2007–2013* Official Journal of the European Communities https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=LEGISSUM:g24242 (downloaded: February 2022)

EUROPEAN COMMISSION (EC) (2007): *Cohesion policy 2007–13, Commentaries and official texts* Office for Official Publications of the European Communities, Luxemburg. https://ec.europa.eu/regional_policy/sources/docoffic/official/regulation/pdf/2007/publications/guide2007_en.pdf (downloaded: February 2022)

EUROPEAN COMMISSION (EC) (2015): The Europe 2020 index: The progress of EU countries, regions and cities to the 2020 targets *Regional Focus* 2015(1) https://ec.europa.eu/regional_policy/sources/docgener/focus/2015_01_europe2020_index.pdf (downloaded: December 2021)

EUROPEAN COMMISSION (EC) (2018b): *Draft joint employment report from the Commission and the Council* Publications Office of the European Union, Luxembourg. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0761&from%20=EN (downloaded: December 2021)

EUROPEAN COMMISSION (EC) (2020b): *Reference and management of nomenclatures – NUTS (Nomenclature of territorial units for statistics), by regional level, version 2016* Publications Office of the European Union, Luxembourg. https://ec.europa.eu/eurostat/web/nuts/history (downloaded: February 2022)

EUROSTAT (2021): *Eurostat – Metadata, ESS reference metadata reporting standards.* https://ec.europa.eu/eurostat/data/metadata (downloaded: February 2022)

KOVÁCS, E. (2014): *Többváltozós adatelemzés* Budapesti Corvinus Egyetem, Operációkutatás és Aktuárius tudományok tanszék, Typotex, Budapest. http://etananyag.ttk.elte.hu/FiLeS/downloads/14_KOVACS_E_Tobbvalt_adatelemzes.pdf. (downloaded: December 2021)