Investigaciones Regionales – Journal of Regional Research
https://doi.org/10.38191/iirr-jorr.25.012

Accepted Article

A core-periphery comeback? Clustering European Union Social Progress Index (EU-SPI)

Albert Busca, Albert Carreras***

Received: 14 February 2025
Accepted: 29 October 2025

Abstract:

Regional cohesion in the European Union (EU) remains a major issue, shaping EU policies, EU budget allocation and regional welfare. We dig into the EU-Social Progress Index published in 2020 to explore its territorial/regional clustering patterns. Are there EU internal frontiers –clusters-? Do they coincide with the member States? Do they have other influences? What determines belonging to different clusters? Is there any resemblance between the GDP per capita and the SPI regional distribution? We divide the EU NUTS-2 map into clusters using hierarchical clustering. We look for the optimal number of clusters, and we compare the outcome with state borders, paying attention to discrepancies or to state combinations. Our main finding is that the optimal clustering is two, and that they are very robustly defined. Of course, this leads towards reassessing core-periphery approaches and the impact of the great recession and the sovereign debt crisis. Another major finding is that the EU-SPI clustering reveals major discrepancies with the per capita GDP clustering. Some NUTS-2 regions perform in SPI terms much better than expected by per capita GDP. On the contrary, some NUTS-2 regions perform in SPI much worse than expected by per capita GDP. The discrepancies suggest major public policy successes and failures.

Keywords: Clustering; social progress; European Union; regional studies; regional welfare.
JEL Classification: I31.

¿Regreso del centro-periferia? Clusterizando el Índice Europeo de Progreso Social (EU-SPI)

Resumen:

La cohesión regional en la Unión Europea (UE) sigue siendo visto como un asunto de la mayor importancia, moldeando políticas europeas, la asignación presupuestaria de la UE y el bienestar regional. Nos adentramos en el Índice de Progreso Social de la UE (EU-SPI) publicado en 2020 para explorar sus agrupaciones y patrones territoriales/regionales. ¿Hay fronteras internas -clústers- en la UE? ¿Coinciden con las de los Estados miembros? ¿Tienen otras influencias? ¿Qué determina pertenecer a un clúster diferente? ¿Hay alguna semejanza entre el PIB per cápita y la distribución regional del SPI? Dividimos el mapa de NUTS-2 en clústers usando clusterización jerárquica. Miramos el número óptimo de clústers y comparamos el resultado con las fronteras nacionales, prestando atención a las discrepancias o a las combinaciones de Estados. Nuestro resultado principal es que el número óptimo de clústers es dos, y que estos están definidos de manera robusta. Por supuesto, esto nos lleva a revalorizar las teorías centro-periferia y el impacto de la gran recesión y la crisis de la deuda soberana. Otro gran resultado es que los resultados de clusterizar el EU-SPI revelan discrepancias sustanciales con la clusterización basada en el PIB per cápita.

* Investigador en la Universidad Autónoma de Madrid, Facultad de Ciencias Económicas y Empresariales. Madrid, España. albertbuscalopez@gmail.com
** Departamento de Economía y Empresa, Universitat Pompeu Fabra. Barcelona, España. albert.carreras@upf.edu
Corresponding author: albertbuscalopez@gmail.com

Algunas regiones NUTS-2 tienen un mejor desempeño en SPI que el esperado por su PIB per cápita. Al contrario, otros NUTS-2 presentan valores del SPI mucho más bajos que los esperados por su PIB per cápita. Estas discrepancias sugieren importantes éxitos y fracasos de las políticas públicas.

Palabras clave: Clustering; progreso social; Unión Europea; estudios regionales; bienestar regional.
Clasificación JEL: I31.

1. Introduction

Assessing the welfare of European regions has been a goal pursued for a long time. While GDP per capita remains the most widely used indicator of economic performance, its limitations as a measure of well-being are well known. The data on disposable family income per capita at purchasing power parity is a much better approach to individual or family economic outcome, but it is hard to find such good quality data at regional level. Alternative ways to approach human well-being have been proposed, but they usually have some shortcomings that make difficult wide enough comparisons through time and space.

The Social Progress Index (SPI) is an initiative of Social Progress Imperative, under the leadership of Michael E. Porter that has jumped to a world view starting almost from scratch (Porter, Stern, & Artavia, 2013). Its first edition was published in 2013 for almost all the countries in the world. It manages to chart social progress without using economic data. It has been issued annually at the country level and globally. The European Commission caught the opportunity to promote the same index applied to EU NUTS-2 regions to assess social progress in Europe and to help policy decision-makers. The EU-SPI is thus created, in the context of the GDP and Beyond[1] initiative.

The EU-SPI is not the only one of its kind. In recent decades, various initiatives have appeared, trying to depart from hard economic measures such as the per capita GDP. Particularly, the Human Development Index (HDI) created by the United Nations Development Programme (UNDP) has made great progress in creating an alternative to GDP that embraces other noneconomic components and is widespread in its use. Others, such as the Better Life Index made by the OECD or the World Happiness Report, also follow this line of work. A summary and discussion of the available choices can be found in Greve, 2016.

Among all these, the EU-SPI, first published in 2016, has some particularities that make it stand out. Firstly, the EU-SPI makes a great effort to explicitly and totally depart from economic metrics, and particularly those measured in monetary values. Composed of 55 indicators in the 2020 (second) version, not one of these is economic in nature. It focuses solely on other aspects of societal progress, mostly welfare outcomes. Moreover, the EU-SPI is the only one to present its results at the subnational level, specifically at the second level of the Nomenclature of Territorial Units for Statistics (NUTS-2).

The EU-SPI's philosophy and expected applicability can be seen in the choice of its indicators. They are chosen to have a set of common characteristics. One of these is that they must "cover matters that can be addressed by policy intervention" (European Commission et al., 2020). This shows that one of the objectives of this index is to aid policymakers on the evaluation of public policies. Furthermore, a second characteristic to be shared by the indicators is that they must "measure outcomes, not inputs" (ibid), which in turn converts the EU-SPI into a tool that policymakers can use to measure the outcomes of the policies that are implemented in each region and especially its impact on social progress. To sum up, this proves that the EU-SPI is designed to be an important and innovative tool that can help any actor interested in better designed policies. At the same time, it promotes more efficiency at allocating public resources to maximize the EU-SPI or, in other words, the social progress of the region.

The EU-SPI is not only distinct in its nature and application, but it also has been shown to be robust and better suited than other alternatives. The latter argument has been proven in relation to the GDP per capita and the HDI, demonstrating that the EU-SPI performs better at predicting social issues and outcomes than the other indices (Siddique et al., 2017). More specifically against the GDP per capita, the

1 More information here: https://ec.europa.eu/environment/beyond_gdp/index_en.html

methodological paper presented by the European Commission delves into this relationship and states that, while some correlation can be found (0.62), the GDP per capita alone is unable to explain all variability in the EU-SPI (European Commission et al., 2020).

Furthermore, the internal consistency and robustness of the index have also been assessed. Recent literature has found that the EU-SPI is robust in its results across various methods of unbalanced penalization (Annoni and Scioni, 2022). Beltran-Esteve et al., 2023, also conclude that the EU-SPI is ~~also~~ robust against changes in the normalization or aggregation criterion. Worried about this issue we tested for robustness on opinion variables by extracting them and recomputing the EU-SPI but found no significant change. The original EU-SPI and the alternative version excluding opinion-based variables were highly correlated (0.95).

In conclusion, the EU-SPI does not only present methodological innovations but also constitutes a robust and consistent index that performs better to estimate social progress than traditional alternatives such as the GDP per capita or the HDI.

Nevertheless, it has attracted much less attention from scholars that could have been expected. Except for the University of Valencia–based team including Beltrán-Esteve et al (2023) and Peiró-Palomino et al (2023 & 2024), there has not been a rush to explore its richness and its potential to provide much more light to major debates in regional science such as those of Rodríguez-Pose (2018) on "The revenge of the places that don't matter" or others that share his co-authorship as Iammarino, Rodríguez-Pose & Storper (2019) on "Regional inequality in Europe: evidence, theory and policy implications" or Dijkstra, Poelman and Rodríguez-Pose (2020) on "The geography of EU discontent". Indeed, the absence of references to the EU-SPI in these major issues and papers reduces the measuring tools to purely economic factors. There seems to be no alternative but the political outcome: voting results. The EU-SPI allows for a complete switch in measurement strategies as its content is the welfare outcome of economic and political factors and decisions.

Taking all of this into account, the goal of this research is to study the regional similarities and disparities that might be present in social progress in Europe, excluding economic indicators. This will be achieved through the extensive analysis of the results of the EU-SPI second edition (2020). As such, the research question emerges as: <u>How do EU regions differ from each other on social progress?</u>

From this question, the initial hypothesis arises as:

<u>H1. There are geographical patterns that affect social progress, especially at the country level.</u> We expect geographical patterns to emerge in relation to the values of the EU-SPI, highlighting regional differences of multicausal nature.

Even more, we also state that:

<u>H2. Difference in social progress also reflect differences in the economic and political context of the regions.</u> Despite the EU-SPI not including any of those variables, we suspect that, inevitably, the differences between regions will also respond to those contexts to some degree.

In the following section, we start by focusing on the methodology that we will apply in order to obtain the most from the richness of SPI data without prejudging the outcomes. We decided to perform a cluster analysis, ideal to classify regional (NUTS-2) diversity and commonality. We discover that the optimal number of clusters is two. Starting from this point we extract as much information (data) as possible out of the two clusters to provide the best possible profile of each of them. After that, we explore alternative number of clusters, but each of them brings us back to the two major clusters, although suggesting potential venues to further explore the reasons for such simple clustering that brings us back to the core-periphery hypothesis. We finish by addressing the relation between the SPI two major clusters and GDP per capita, dividing the EU into above and below-average per capita GDP and linking it to cluster membership, suggesting new avenues for further research

2. Methodology

EU-SPI was first published in 2016. The second edition was presented in 2020 and the third edition -labelled as EU-SPI 2.0- is the last edition available at the time of this writing, published in 2024.

When choosing to analyse the second edition as opposed to either the first or the third, several factors were taken into account. Firstly, from the entirety of the indicators of the EU-SPI 2.0, three quarters use data from 2020 to 2022. The second edition mainly extracts its data from the 2016 to 2018 period while the first edition is mostly from the period of 2011 to 2013.

As such, much of the data that forms the third EU-SPI was collected during the COVID-19 pandemic in Europe and while the extraordinary lockdown and health protecting measures were still in action. The impact of both COVID mortality and lockdown is quite visible and distorts the "normal" results. The special impact of Covid mortality on European regions has been extensively studied, and expert research finds patterns that could alter substantially the SPI regional patterns (see, for instance, Burlina & Rodríguez-Pose, 2023). The first edition also represents abnormality, extracting its data from the years in the aftermath of the 2008 financial crisis and during the sovereign debt crisis, with its effects still impacting the EU-SPI results. The second edition is the only one that responds to a situation of normality, making the findings based on it expectedly more loyal to the true situation of social progress in Europe.

However, choosing the second edition also has its drawbacks. Compared to the first, the United Kingdom is not present in the database and consequently 40 NUTS-2 regions are lost. On the other hand, the EU-SPI 2.0 has obviously more recent data than the second one. Ultimately, a decision was made to use the second edition to eliminate the effect of both the pandemic and the financial and sovereign debt crisis at the expense of losing some observations (compared to first edition) and using slightly older data (compared to the third edition). For a first exploitation of its wealth of data, it seems to us the best start. Nevertheless, we acknowledge that a valuable first attempt to exploit the EU-SPI 2.0, with a focus on the East-West EU divide has already been published (Lacmanovic & Tijanic, 2025).

The impact of this decision is arguably limited. The EU-SPI 2.0 consists of 53 indicators. Of those, only 13 are new, while the other 40 (75%) are the same. The components and the dimensions remain equal. In turn, the second edition introduced 14 new indicators among its 55. Moreover, the methodology in the normalization and aggregation is mostly the same across the editions with minor changes. On top of this, the 2020 edition enjoys the tough methodological screening made by Beltrán-Esteve, Peiró-Palomino, Picazo-Tadeo and Ríos (2023) with very positive results. For all these reasons, the benefit of using the second edition exceeds the drawbacks discussed previously. See Annex 1 for details on the composition of the second edition of the EU-SPI by indicators, components, and dimensions.

To test for the existence of geographical patterns stemming from the EU-SPI, in line with our first hypothesis, a cluster analysis will be performed. The very nature of the data supports this choice. There are precedents of cluster analysis applied to similar sets of data and problems (Del Campo, Monteiro, Oliveira, 2008; Monfort, Cuestas & Ordóñez, 2013).

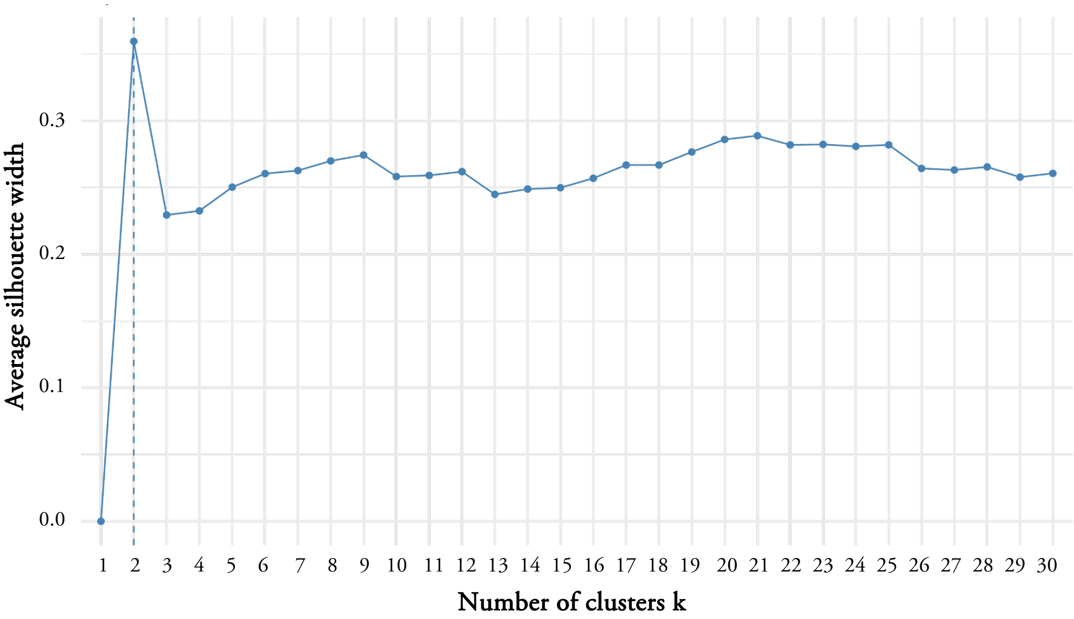
In our case, the twelve components of the EU-SPI will be used to compute the clusters. These indicators have been first normalised from 0 to 100 using a min-max transformation with indicator-specific boundaries (Annoni & Bolsi, 2020). They are then aggregated as the unweighted generalised power mean of order 0.5 by default to ensure that the index is partially non-compensatory (Annoni & Bolsi,2020).

More specifically, a hierarchical cluster method will be used. Hierarchical clustering has the benefit of not needing to specify the number of clusters when applied and allows to dynamically choose the granularity of the clusters by cutting the results at different levels. Among the methods available, the Ward method (Ward, 1963) was chosen.

For cluster validation, several steps were taken. First, the validity of the dendrogram was evaluated using the cophenetic correlation coefficient (Sokal and Rohlf, 1962) -or CPCC-, which has been proven to be a good tool to test for the global fit of the dendrogram on the data when validating clusters (Dubes and Jain, 1979). The dendrogram resulting from the Ward method not only had a satisfactory CPCC value of 0.64, but the value was higher than those of other methods tested. The Ward method was also chosen because, given the nature of this method, it maximizes internal variance. In other words, the observations that make up each cluster share similar characteristics across the twelve components of the EU-SPI, which aligns with our research objective.

The next step was to validate the best number of clusters to perform. Two indices were used: for its wide-spread use and good performance (Chouikhi et al., 2015; Arbelaitz et al., 2013), the Silhouette Index (Rousseeuw, 1987); and for being proven as one of the best performing indices (Chouikhi et al., 2015; Arbelaitz et al., 2013; Milligan and Cooper, 1985), the Calinski-Harabasz (CH) Index (Calinski and Harabasz, 1974). The mean Silhouette for every number of clusters ranging from 2 to 30 was computed. The resulting graph can be seen in Figure 1. The Calinski-Harabasz index was also computed for each number of clusters from 2 to 30.

FIGURE 1.
Mean Silhouette for each number of clusters



As can be seen in Figure 1, the best performing number of clusters is two, with an average Silhouette of (0.36). The CH index has similar results. The highest index results for two clusters (164.38). Following these results, the main analysis will be performed partitioning the dendrogram in two clusters. We must stress that this result -two clusters- is a major, although disturbing, contribution "per se".

After the first cluster analysis, we expect the existence of the geographical patterns to be either proven or disproven. After this, we aim to test the nature of these geographical patterns. For that, further analysis techniques will be used. First, a Linear Discriminant Analysis (Huberty, 1975) or LDA, secondly a Random Forest Analysis (Breiman, 2001; Genuer et al., 2010) or RFA, and lastly a Factor Analysis (Alhija, 2010) or FA.

The LDA finds a Linear Discriminant (LD1) as a combination of the twelve components so that it better separates the two clusters. With this, we can assess which components contribute the most to the separation between clusters.

While the LDA focuses on linear relationships, the RFA will be useful in understanding the impact on cluster formation when the relationships are non-linear.

Finally, the FA will result in the calculation of several factors, each of which will represent a latent variable in the dataset. The grouping of the components in these factors will aid in understanding if there are core aspects of the components that cannot be seen in the original variables.

Even though this main analysis is deemed to be the most relevant, further partitions are also believed to be important. Firstly, a partitioning at nine clusters will be performed. Both Silhouette index and CH index are relatively good compared to most alternatives and it allows seeing how those two initial clusters are further subdivided without atomizing the groups.

The final partition studied will be at 21 clusters. Despite the value being lower in the CH index, it has the second highest Silhouette index. It also represents the optimal subdivision close to the same number of countries included in the EU-SPI (27). This partition will, therefore, be of the utmost relevance to understand the impact of country boundaries to social progress.

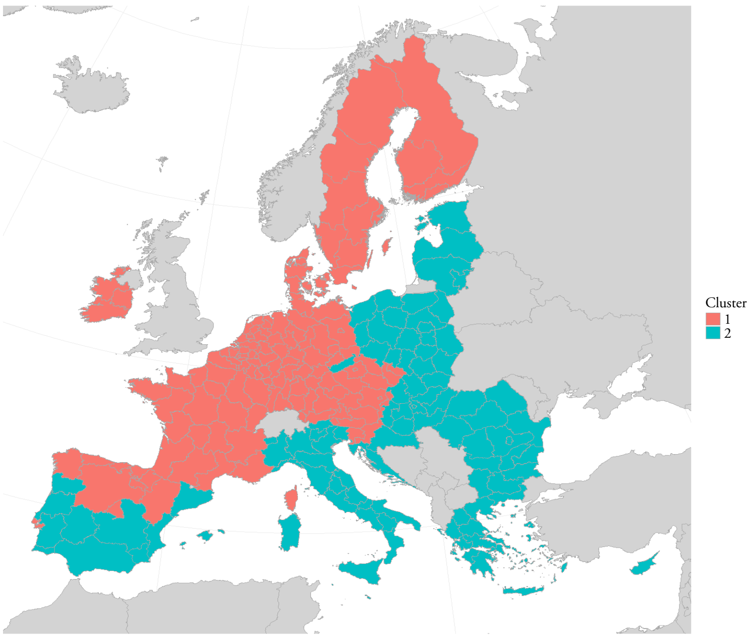
The cluster analysis, aligned with the two hypotheses, will be conducted using the components that make up the EU-SPI (H1), but also contrasted with other important contextual variables: the per capita GDP and the national borders (H2). By doing this, we will be able to test whether the clusters reflect economic, political or social characteristics.

3. Results

3.1. Two clusters

Firstly, a dendrogram was computed from the values of the 12 components across all NUTS-2 regions (Appendix 2). At this level, most national borders are preserved, with Spain as the main exception. Europe is divided into two regions, the first comprised by central and northwestern countries and the second comprised by eastern and southern countries. Notably, Italy is separated from the other EEC founding countries. This division calls for a reassessment of the "core-periphery" hypothesis, just as some authors (Magone, Laffan and Schweiger, eds., 2016) have done, but with some significant changes in cluster frontiers: most of Northern Spain (except Catalonia), Lisbon region, Slovenia and most of Czechia (but Severozápad) join the core, while Catalonia and all of Italy do not.

Figure 2.
EU Map at the NUTS-2 level of the SPI clustering results



The mean of each component by cluster can be seen in Table 1. Cluster 1 has higher average values in every component except for *Personal Security*, indicating overall better performance. The largest difference between clusters occurs in *Access to Advanced Education*.

Table 1.
Mean for each EU-SPI component by cluster

EU-SPI Components	Cluster 1	Cluster 2
Nutrition and Basic Medical Care	85.73	75.33
Water and Sanitation	93.83	83.80
Shelter	88.55	70.63
Personal Security	70.23	73.45
Access to Basic Knowledge	78.77	68.82
Access to ICT	82.64	65.60
Health and Wellness	71.19	58.70
Environmental Quality	52.84	37.57
Personal Rights	56.32	39.59
Personal Freedom and Choice	72.30	54.24
Tolerance and Inclusion	70.45	49.02
Access to Advanced Education	66.55	42.36

With all this, two distinct European areas are revealed (northwest and southeast) with differing social progress levels, in line with recent 'core-periphery' literature (Magone, Laffan and Schweiger, eds., 2016). Accordingly, Cluster 1 can be identified as the 'core' of the EU and Cluster 2 as its 'periphery'. The two clusters are well separated and there is homogeneity within the clusters. To further examine the nature of these patterns, we perform a Linear Discriminant Analysis (LDA), a Random Forest Analysis (RFA) and a Factor Analysis (FA).

The LDA aids in discovering the linear relationships between the components and cluster membership. The values shown in Table 2 are the scaling factor for each component, where larger absolute values indicate greater influence on the first linear discriminant (LD1) and, therefore, on cluster separation. Negative values correspond to Cluster 1 membership, while positive values correspond to Cluster 2.

Table 2.
First Linear Discriminant (LD1) loading for each EU-SPI component

EU-SPI Components	LD1
Nutrition and Basic Medical Care	0.055
Water and Sanitation	0.032
Shelter	-0.163
Personal Security	0.034
Access to Basic Knowledge	-0.031
Access to ICT	0.005
Health and Wellness	-0.062
Environmental Quality	0.005
Personal Rights	-0.038
Personal Freedom and Choice	-0.016
Tolerance and Inclusion	-0.011
Access to Advanced Education	-0.007

Shelter stands out as the most important component, with a scaling factor twice that of the second highest -*Health and Wellness*-. *Nutrition and Basic Medical Care* ranks third. These are the components that have a stronger linear relationship with cluster membership. Given their negative scaling factor, higher values of *Shelter* and *Health and Wellness* are associated with Cluster 1, whereas higher values of *Nutrition and Basic Medical Care* correspond to Cluster 2.

Next, we apply an RFA to account for non-linear relationship between the EU-SPI components and cluster membership, nuancing the findings in LDA. Table 3 presents the results. The first column presents global performance for each component, with *Shelter* again emerging as the most important variable, showing the largest Mean Decrease Accuracy, meaning the model relies on *Shelter* to predict cluster membership. *Access to ICT* ranks second, emerging as an influential variable for both clusters. The second and third column show the same metric but are disaggregated for each cluster. *Health and Wellness* and *Nutrition and Basic Medical Care* are particularly useful for classifying Cluster 1 regions, whereas *Personal Freedom and Choice* is more important for identifying Cluster 2.

TABLE 3.
Random Forest Analysis results

EU-SPI Components	Mean Decrease Accuracy	Cluster 1[a]	Cluster 2[a]
Nutrition and Basic Medical Care	12.408	11.646	6.007
Water and Sanitation	3.873	-1.038	4.220
Shelter	24.657	20.384	16.576
Personal Security	4.059	3.303	2.569
Access to Basic Knowledge	13.080	10.782	10.998
Access to ICT	18.763	17.734	10.549
Health and Wellness	11.910	11.232	6.806
Environmental Quality	6.514	5.359	3.855
Personal Rights	14.086	9.628	10.638
Personal Freedom and Choice	13.477	8.051	11.148
Tolerance and Inclusion	15.599	12.931	10.608
Access to Advanced Education	9.013	6.369	6.752
[a]Decrease in accuracy for the selected cluster			

In summary, *Shelter* is the most important component in both models. *Health and Wellness* and *Nutrition and Basic Medical Care* diminish their impact when non-linear relationships are considered in favour of *Tolerance and Inclusion* or *Personal Rights*. *Access to ICT* has a strong non-linear relationship with cluster membership. Overall, these variables seem to be key determinants of classification.

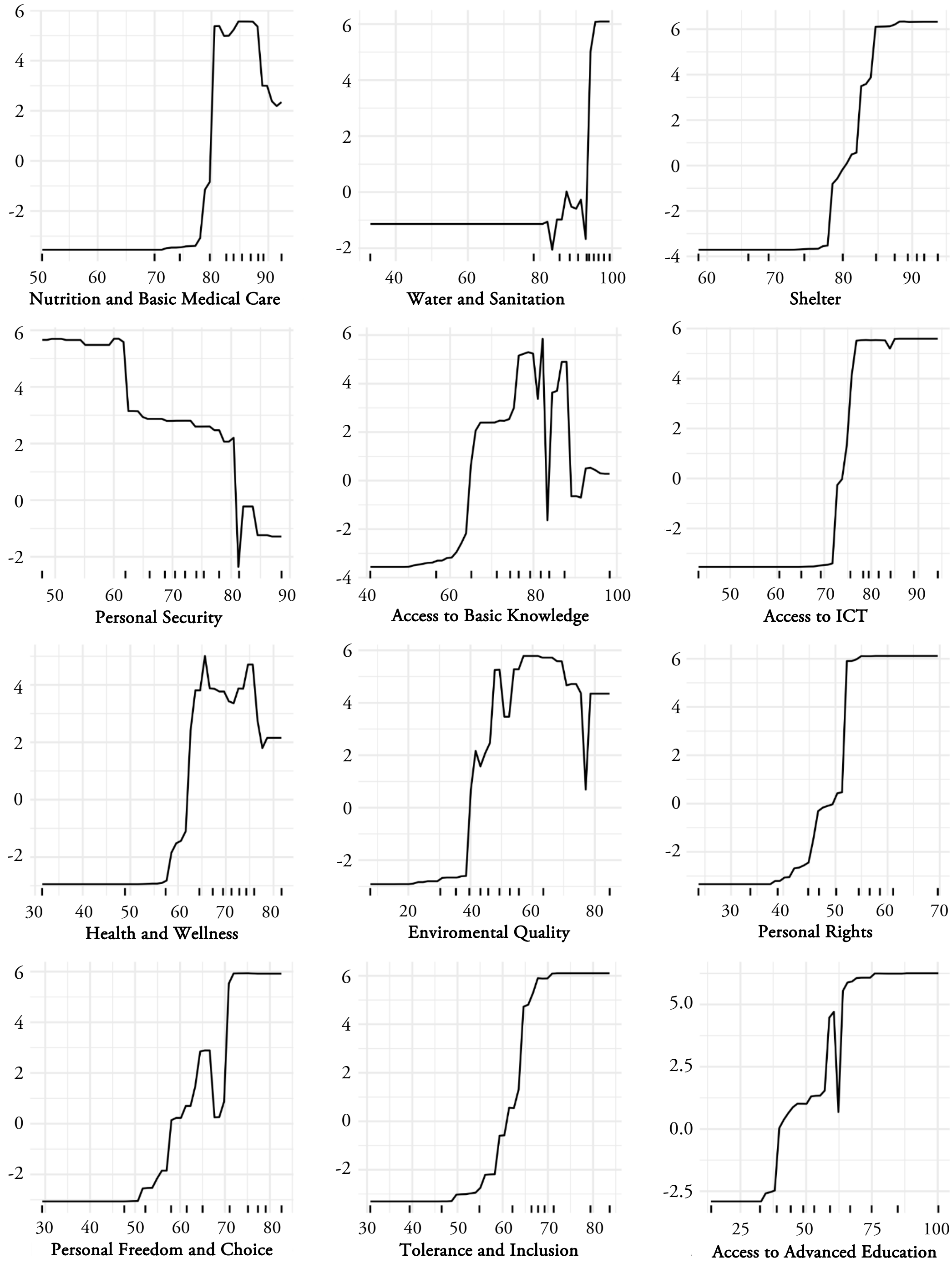
According to both LDA (Table 2) and RFA (Table 3), Cluster 1 is characterized by high levels of *Shelter*, the two components related to health or physical wellbeing, and *Access to ICT*. Conversely, Cluster 2 is defined by having low levels of *Shelter* -again as the most important variable- and *Personal Freedom and Choice*, which also includes employment related indicators.

With this in mind, it can be stated that the core of the EU (cluster 1) is defined by better housing and technological accessibility while the periphery (cluster 2) is characterised by poorer housing and a worse employment market.

To understand the linear and non-linear relationship between each component and cluster membership, partial dependence plots (PDPs) were computed. Based on the RFA model, these plots show the logs odds of belonging to Cluster 1 (the plots for Cluster 2 would be its mirror image) for any given component value. The results of the PDPs can be seen in Figure 3.

ISSN: 1695-7253 e-ISSN: 2340-2717

FIGURE 3.
Partial dependence plots for cluster 1 by components



As expected, the probability of being in Cluster 1 generally increases with higher component values, except for *Personal Security*. The short lines along the x-axis represent the data deciles.

Most graphs show very low probabilities of being in Cluster 1 until a certain threshold, where the probabilities rise sharply. This pattern, particularly evident in key components according to previous models such as *Shelter* or *Access to ICT*, implies the presence of critical values that distinctly separate the two clusters, reinforcing their differentiation.

It also highlights tipping points where policy interventions that enhance or reduce a component's value may critically influence cluster membership, while changes beyond certain ranges may have limited effects.

It also shows tipping points where the impact of public policies either improving or diminishing the component's value may have a very critical impact on cluster membership, while changes beyond certain ranges may have limited effects.

After assessing the importance of each component, a Factor Analysis (FA) was conducted to determine if there are any latent variables that affect the components and explain cluster membership. An oblique rotation was applied, allowing the factors to be correlated and consistent with the conceptual interrelation among social progress components. This approach enhances the model's robustness.

Prior to computing the model, the data's suitability for FA was assessed. The Kaiser-Mayer-Olkin (KMO) measure (Kaiser, 1970) reached a global value of 0.86, indicating very good sampling adequacy. All components except three exceeded the 0.8 threshold and only *Personal Security* fell below the minimum 0.5 with a value of 0.19. Bartlett's test (Bartlett, 1937) yielded a p-value close to 0, rejecting the null hypothesis of no correlation among variables and confirming their suitability for FA.

Based on these results, the model was calculated omitting *Personal Security*. To determine the number of factors, several models were estimated and their fit assessed using the Root Mean Square Error of Approximation or RMSEA (Steiger & Lind, 1980; Browne & Cudeck, 1992), and the Bayesian Information Criterion or BIC (Schwartz, 1978). The four-factor model was found to have optimal values for both indices and was therefore retained. Factor loadings, communalities, and complexities for each component are presented in Table 4.

TABLE 4.
Factor Analysis results

EU-SPI Components	F1	F2	F3	F4	Communality	Complexity
Nutrition and Basic Medical Care	1.03	0.01	-0.08	-0.01	0.94	1.01
Water and Sanitation	0.39	0.03	0.11	0.44	0.63	2.10
Shelter	0.22	0.58	0.31	-0.11	0.86	1.95
Access to Basic Knowledge	-0.15	0.16	-0.02	0.82	0.70	1.15
Access to ICT	0.29	0.13	0.57	0.08	0.89	1.64
Health and Wellness	0.99	-0.11	0.06	-0.11	0.89	1.06
Environmental Quality	-0.26	0.06	0.94	-0.17	0.56	1.23
Personal Rights	0.10	0.27	0.58	0.13	0.88	1.61
Personal Freedom and Choice	-0.14	0.88	-0.02	0.30	0.92	1.29
Tolerance and Inclusion	0.38	0.19	0.56	-0.11	0.93	2.15
Access to Advanced Education	0.03	-0.29	0.89	0.21	0.74	1.34

Higher loadings represent greater importance of a component within a given factor. Communality and complexity are metrics that aid in understanding how well the model explains each component. Communality is the proportion of a component's variance explained by the model. As such, *Nutrition and Basic Medical Care* is almost fully explained while only just above half *Environmental Quality* is captured. Complexity is a measure that indicates the extent to which a component is associated with multiple factors. *Nutrition and Basic Medical Care* shows low complexity, indicating it is largely explained by a single factor (F1).

Among the components identified as most important by RFA and LDA, *Shelter* is particularly important in the second factor, exhibiting a high communality (0.86). The model also explains around 90% of other key components, including *Access to ICT*, *Personal Freedom and Choice*, and *Tolerance and Inclusion*.

Focusing on factor loadings, we find that *Nutrition and Basic Medical Care* (1.03) and *Health and Wellness* (0.99) primarily define the first factor, therefore named as Health. Factor 2, dominated by *Personal Freedom and Choice* and *Shelter*, is designated as Freedom, Employment and Housing (FEH). Notably, over half of the indicators within *Personal Freedom and Choice* relate to the labour market, despite the label not reflecting this. The third factor is most importantly composed of *Environmental Quality* and *Access to Advanced Education*, with minor contributions from *Access to ICT*, *Personal Rights*, and *Tolerance and Inclusion*. Although not suggested by its title, *Personal Rights* largely reflects institutional trust. This diverse grouping of components, most not relevant in previous factors, is termed Eco-social and Institutional Quality (ESI Quality). Finally, the fourth factor, defined mainly by *Access to Basic Knowledge*, is labelled Basic Education.

Table 5 shows the variance explained by each factor. The first row is directly correlated to table 4, as it is the sum of the squared loadings per factor. The second and third rows show the proportion of variance explained individually and cumulatively, while the fourth and fifth refer to the variance explained within the model. ESI Quality (F3) accounts for the largest share of variance in the original data, followed by Freedom, Employment and Housing (F2). The four factors combined explain 81% of the variance.

TABLE 5.
Variance explained by each factor

	F3	F2	F1	F4
SS loadings	3.236	2.783	1.697	1.231
Proportion Var	0.294	0.253	0.154	0.112
Cumulative Var	0.294	0.547	0.701	0.813
Proportion Explained	0.362	0.311	0.190	0.138
Cumulative Proportion	0.362	0.673	0.862	1.000

Table 6 presents the correlation matrix between factors. Eco-social and Institutional Quality (F3) and Health (F1) have the highest correlation while the lowest is between Health (F1) and Basic Education (F4), indicating that higher levels of Health are correlated with higher levels of ESI Quality.

TABLE 6.
Correlation matrix between factors

	F1	F2	F3	F4
F1	1.000	0.491	0.733	0.408
F2	0.491	1.000	0.672	0.432
F3	0.733	0.672	1.000	0.487
F4	0.408	0.432	0.487	1.000

Figure 4 displays the scatter plots of these correlations, with colours indicating cluster membership (red for Cluster 1, blue for Cluster 2). Most correlations exhibit a linear pattern where higher values of both factors are related to Cluster 1 membership. The relationship between Health and Basic Education, however, departs from this. Lower values in both are related to Cluster 2, but the graph shows a fork where Cluster 2 regions improve in either only Health or Basic Education and only regions in Cluster 1 excel in both areas simultaneously.

FIGURE 4.
Graphical correlation matrix between factors

To know whether these factors can accurately predict cluster membership, hierarchical clustering was computed using the Ward method as before, this time with the four factors as predictor variables instead of the EU-SPI components. The dendrogram was cut at two clusters and a confusion matrix was constructed to compare both models. Only 12 regions (5.19% of the total) were differently classified. As shown in Table 7, all Spanish regions originally in Cluster 1 -except Madrid, Basque Country and Asturias- were reassigned to Cluster 2. Conversely, all Portugues regions except Lisbon changed to Cluster 1, thereby unifying the country. Malta also moved from Cluster 2 to Cluster 1 and Bratislava shifted the opposite direction.

Overall, the original clusters showed to be robust, and the factors were validated as good predicting variables for cluster membership. Misclassifications were concentrated in the Iberian Peninsula. If these changes were considered, national boundaries would remain almost intact, except for three regions in Spain and one region in the Czech Republic (Severozápad).

Table 7.
Regions that change cluster membership

NUTS Code	Region Name	Original Cluster	Factor Cluster
ES11	Galicia	1	2
ES12	Principado de Asturias	1	2
ES13	Cantabria	1	2
ES23	La Rioja	1	2
ES24	Aragón	1	2
ES41	Castilla y León	1	2
MT00	Malta	2	1
PT11	Norte	2	1
PT15	Algarve	2	1
PT16	Centro	2	1
PT18	Alentejo	2	1
SK01	Bratislavský kraj	1	2

A Random Forest with the four factors as independent variables was computed to assess their importance in classifying observations. The results, shown in Table 8, indicate that Eco-social and Institutional Quality (F3) -driven primarily by *Environmental Quality* and *Access to Advanced Education*- was the strongest overall classifier, as can be seen by its highest value in Mean Decrease in Accuracy. Health (F1) performed best for Cluster 1 but worst for Cluster 2, pointing out notable inter-cluster differnces. Mainly, it highlights that Cluster 1 exhibits consistently high values for this factor, whereas Cluster 2 shows greater variability.

Table 8.
Random Forest results with Factors as predictor variables

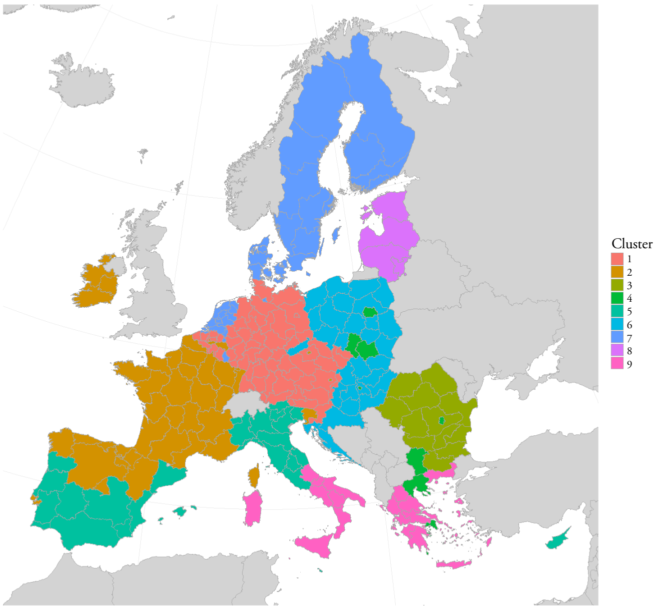
Factors	Mean Decrease Accuracy	Cluster 1	Cluster 2
F1	30.47503	30.48155	14.57364
F2	32.91022	26.95012	20.16020
F3	34.70631	25.45501	25.69503
F4	20.70812	16.13208	16.37120

3.2. More clusters

To understand how the two clusters operate internally, the original dendrogram was further divided at the 9 and 21 cluster levels (Appendices 5 and 6). Twenty-one clusters has the second highest Silhouette Index after the two-cluster model, and nine clusters represent the first local maximum, offering insight into the initial subdivision of the main clusters. Given that the two clusters ought to represent the core (Cluster 1) and the periphery (Cluster 2), these finer divisions may revel distinct sub-cores and sub-peripheries. It is also important to note that all subclusters derived from Cluster 1 remain more similar to each other than to those emerging from Cluster 2 and vice-versa.

The results of subdividing the clusters into 9 can be seen in Figure 5. The former Cluster 1 is divided into three clusters (1, 2 and 7) while the former Cluster 2 splits into six clusters (3, 4, 5, 6, 8, and 9). The national borders largely preserved for 2 clusters are now broken, mostly in southern Europe. Norther Spanish regions (except Catalonia) merge with France, while eastern Spain, southern Spain and Portugal cluster with northern Italy. At the same time, southern Italy groups with most of Greece. On the other hand, France, Germany, and central Europe largely keep their borders intact. Most notably, the Nordic countries form a single, cohesive cluster.

FIGURE 5.
NUTS-2 Map at 9 Clusters



Additionally, Table 9 reports the mean per capita GDP and EU-SPI level for each cluster. The highest SPI cluster (7) comprising the Nordic countries, Hamburg, the Netherlands and Luxemburg, represents the top performers of the former Cluster 1. Its SPI exceeds that of the second-ranked cluster (2) -which includes France, Ireland, Northern Spain, Lisbon, Viena and western Slovenia- by twelve percent, and its per capita GDP by seventeen per cent. Along with the third SPI ranked (1) (most of Germany, Austria, Czechia, Belgium and eastern Slovenia), these three clusters represent the former Cluster 1. The core thus subdivides into three groups: the most advanced (the Nordic countries), an intermediate group (France, Ireland and northern Spain), and a lower-performing core (Germany and central Europe). Its worth noting, however, that the SPI difference between the intermediate and the lower-performing core is 1% in favour of the latter, while GDP per capita is 5% higher in the former. Thus, Cluster 2 (France, Ireland and northern Spain) is is getting more social progress out of their per capita GDP than Cluster 1 (Germany and central Europe). More over, Cluster 1 scores higher in *Personal Freedom and Choice* but lower in *Access to Advanced Education* or *Environmental Quality*, suggesting a more industrial and labour-oriented model of social progress, in contrast to the education and environment-focused model of Cluster 2. This distinctions reveal that within the 'core', different models of social progress may be found, being reminded of Esping-Andersen's (1990) concept of the "Three Worlds of Welfare Capitalism".

Regarding the former cluster 2, the periphery, the fourth SPI cluster (8) comprises the three Baltic countries that are well above their per capita GDP ranking (sixth). The fifth SPI cluster (5) - also the fifth in GDP- includes northern and central Italy, eastern and southern Spain, continental Portugal (but Lisbon), Cyprus and Malta. While both clusters show similar SPI levels (three-point difference), Cluster 8 excels in education-related components while Cluster 5 performs better in health-related ones. The sixth SPI cluster (4) comprises regions that are the location of Eastern European capitals or major cities. Their SPI ranking is worse than their per capita GDP standing (fourth). They differ from other clusters in the periphery with a more industrial and labour-oriented profile, presenting lower values of Health and *Environmental Quality*. The eighth (in both SPI and pc GDP) cluster (9) includes Greece (but Athens and Thessaloniki regions) and southern Italy, Sardinia and Sicily – the classical Mezzogiorno. Similar to cluster 5, this cluster emphasises better values for Health but low values for education-related components. The worst cluster in SPI and pc GDP (3) is made of Bulgaria and Romania, excluding their capital regions.

 ISSN: 1695-7253 e-ISSN: 2340-2717

TABLE 9.
Mean GDPpc and EU-SPI values by clusters at 9 clusters

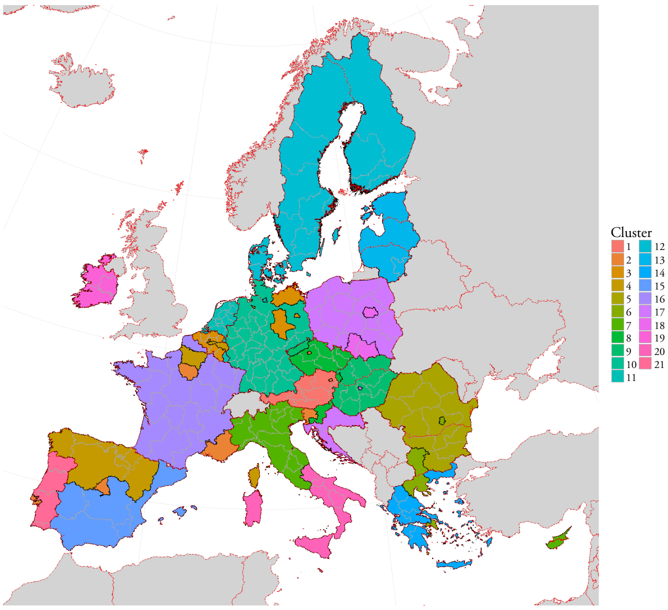
Cluster	Mean GDPpc	Mean EU-SPI
1	33,496.72	71.01177
2	31,917.07	71.71985
3	15,500.00	47.42561
4	32,262.50	59.55308
5	27,128.57	63.22684
6	18,533.33	59.25026
7	37,493.55	80.43914
8	26,700.00	66.22437
9	17,015.79	55.83056

In summary, European regions differ markedly in their capacity to translate economic resources into social progress. Capital regions consistently outperform their national counterparts, particularly in less developed countries. Longstanding targets of EU regional policy, such as the *Mezzogiorno*, continue to underperform economically and socially, whereas many Eastern European countries -like the Baltic countries or the former DDR- exceed expectations relative to their economic standing.

From this, distinct models of social progress also emerge. Within the core, some regions follow an industrially oriented path, while others emphasize education and environmental sustainability. The periphery displays even greater diversity, with clusters prioritizing health (clusters 5 and 9), education (8), or industrial and economic activity (4 and 6).

The results of further subdividing into 21 clusters can be seen in Figure 6. For clarity, cluster boundaries are outlined in black and national boundaries in red. Here, the clusters generally align with national borders, with few exceptions observed between Romania and Bulgaria, the Baltic states and the Nordic countries.

FIGURE 6.
NUTS-2 Map at 21 Clusters



To better understand the clusters, Table 10 reports the mean per capita GDP EU-SPI level for each cluster. The wealthiest cluster is Cluster 19, comprised exclusively by Ireland, while the poorest cluster is Cluster 14, comprised entirely by Greek regions. By EU-SPI, the highest mean value is in Cluster 12 (Nordic regions) amd the lowest in cluster 5 (Romania and Bulgaria).

Table 10.
Mean GDPpc and EU-SPI values by cluster at 21 clusters

Cluster	Mean GDPpc	Mean EU-SPI
1	36,025.00	73.88641
2	42,900.00	70.68821
3	31,650.00	69.34763
4	27,425.00	69.81758
5	15,500.00	47.42561
6	29,475.00	55.43730
7	32,428.57	62.01472
8	26,725.00	67.87469
9	17,481.82	57.59260
10	35,196.97	71.68054
11	40,500.00	78.07988
12	35,017.65	82.38205
13	26,700.00	66.22437
14	15,036.36	55.38756
15	21,355.56	64.39801
16	26,155.56	72.85288
17	19,256.25	60.38989
18	35,050.00	63.66885
19	55,166.67	75.28185
20	19,737.50	56.43970
21	22,680.00	64.51266

Of the 21 clusters, seven are one-country clusters. Most notably, Cluster 10 comprises most of Germany, Cluster 15 all of Austria except Vienna, and Cluster 16 almost all France. On the other end, Cluster 2 spans seven countries including the regions for Brussels, Lisbon, Ljubljana, Madrid, Marseilles, Paris, Prague and Vienna. Berlin, however, is in Cluster 3 with several Belgian and other German regions. This pattern highlights the distinctive character of capital regions, which generally differ from their national counterparts. Even more, for the capitals in cluster 2, it suggests that they are more similar between each other than to regions within their country. In southern Europe, it is worth noting that Spain and Italy uniquely divide their country in half into two clusters. In contrast, some countries -such as the Nordic countries, the Netherlands, Ireland, Croatia and Lithuania- are quite internally homogeneous according to their SPI clustering, even at 21 clusters.

3.3. Clustering per capita GDP and EU-SPI

Lastly, we assess the influence of per capita GDP on the SPI clusters. As the EU-SPI aims to provide an alternative to per capita GDP, examining whether SPI cluster performance diverges from GDP levels serves to validate this objective. Conversely, if per capita GDP significantly explains SPI cluster membership, it would mean that the clusters reflect not only social but also economic factors. GDP will be measured as per capita GDP at purchasing power parity (PPP). While per capita GDP is estimated at the NUTS-2 level, the PPP index is only calculated at the national level (NUTS-1) by Eurostat.

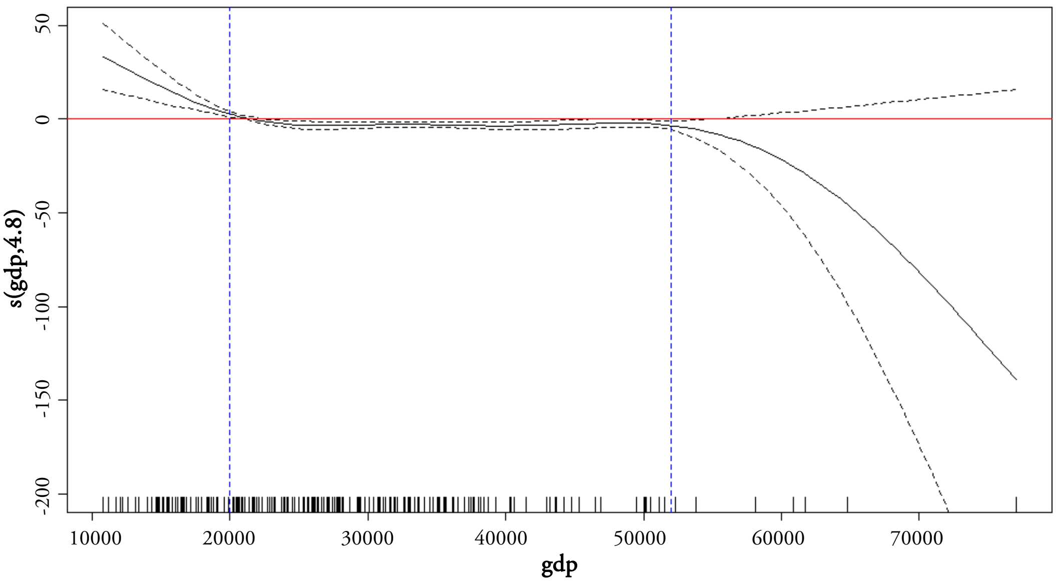
To understand the relationship between per capita GDP and cluster membership, several regression models were computed. They all take the per capita GDP as a predictor and cluster membership as the dependent variable, coded dichotomously (core and periphery). The models aimed to estimate the probability of being of belonging to Cluster 2 (periphery) relative to Cluster 1 (core).

Four models were tested: linear, logarithmic, quadratic and Generalized Additive Models (GAMs). GAMs, being non-linear and non-parametric, allow for more flexible relationships between GDP and cluster membership. Model performance was assessed using the BIC index, which balances model fit and complexity, an especially important consideration for GAMs.

Among the linear models, the logarithmic specification achieved the lowest BIC (216.45). For the GAMs, the number of basis functions (allowing varying degrees of non-linearity) tested were three, four, six, nine and ten. The model with six basis functions produced the lowest BIC (194.5), outperforming the logarithmic model. Consequently, the GAM model with a maximum of six basis functions was selected.

Figure 7 displays the results of the model. The y-axis represents the probability of being in Cluster 2 (periphery) with Cluster 1 (core) as the reference. The x-axis represents the per capita GDP. The solid black line is the expected probabilities from the GAM model, and the dashed lines represent the 95% confidence interval.

Figure 7.
GAM expected probabilities of being in Cluster 2 by GDP level

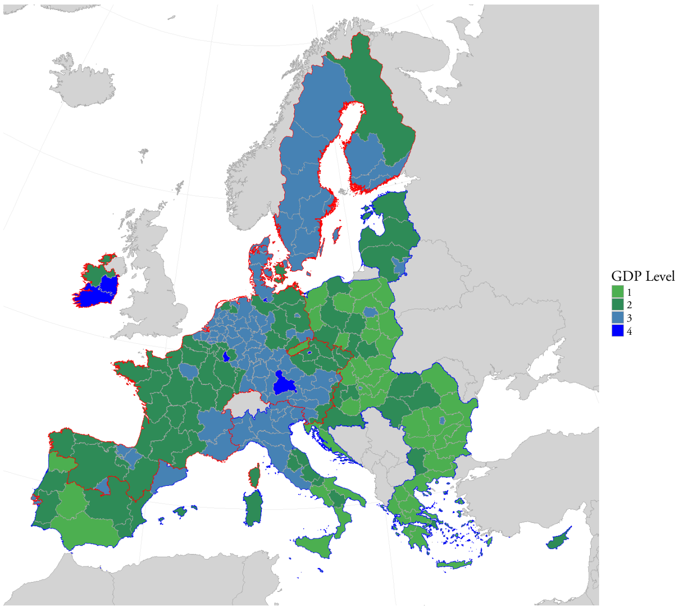


There are three distinct regions in the graph. Regions with per capita GDP below 20,000€ are associated with Cluster 2, whereas regions above 52,000€ are associated with Cluster 1. Between these thresholds, comprising 175 regions, per capita GDP does not have any effect on cluster membership. At the extremes, 49 regions fall within the lowest range and seven within the highest.

Figure 8 illustrates these effects by dividing the map into four per capita GDP levels. The first level (light green) includes regions below 20,000€, corresponding to those associated with Cluster 2 (periphery) based solely on the GDP. The fourth level (dark blue) comprises regions above 52,000€ associated with Cluster 1 (core). Levels 2 and 3 are not linked to either cluster, differing only in wether their per capita

GDP falls below or above the European mean. For comparison, cluster boundaries are overlaid: red for Cluster 1 and blue for Cluster 2.

FIGURE 8.
NUTS-2 Map with per capita GDP levels and cluster borders



This four-level classification provides additional insights. By distinguishing regions below or above the European mean, it depicts what the EU's economic core (blue) and periphery (green) would look like if clustering was based only on GDP. It is clear that the clusters computed using the EU-SPI differ greatly in some parts of the EU. For instance, most French regions fall below the EU per capita GDP yet belong to Cluster 1, indicating strong territorial cohesion and effective public policy. At the other end we find Italy. All the northern regions, though economically above EU average, fall into SPI Cluster 2, suggesting limited translation of economic wealth into social progress. The permanence of internal economic cleavages stresses that the northern failures do not profit the southern regions. Other success stories appear in other cases: former DDR regions achive SPI levels comparable to the former FDR despite lower GDP, Czech regions outperform their GDP in SPI terms, and several northern Spanish regions belong to the top SPI cluster despite below-average GDP. Just the contrary happens with the Catalan case. Are the northern Italy regions and Catalonia failures to translate economic prosperity into SPI performance a pure inefficiency loss or are they benefiting other regions?

After all, GDP exhibits mixed effects on social progress throughout Europe. While it matters for the lowest and highest income regions, for most areas economic prosperity alone neither guarantees nor constrains social progress.

4. DISCUSSION

The clustering analysis partially confirms the first hypothesis. Distinct geographical patterns of social progress exist within the EU, broadly dividing the continent into two macro-regions. However, these patterns are not strictly defined by national borders. Rather, the two clusters delineate wider transnational regions that better capture Europe's social progress disparities and within them, most country borders remain largely preserved. This outcome strongly supports the relevance of a core-periphery framework, often questioned but historically robust, and specially of the relevance of the most recent works on that matter such as Magone, Laffan and Schweiger (2016). The classification of Italy as a peripheral country is both striking and plausible, while Spain's internal division -half core, half periphery- represents an unexpected finding. Such results likely reflect the major enduring effects of major economic shocks, like the Great Recession and the Sovereign Debt Crisis, and point to the need for a reassessment of EU regional cohesion policy, especially in regions like southern Italy where its effects have fallen short of expectations.

The second hypothesis is also partially supported. Economic and political factors cannot alone fully explain cluster membership or social progress levels. Geographical proximate regions tend to be more similar to each other, particularly within the same country, with the notable exception of capital regions, which are consistently more similar to one another than to their national counterparts. This indicates that geography does play a role on cluster formation, though whether this stems from political factors or other unobserved variables remains unclear.

Additionally, economic factors appear to matter only at the extremes, affecting only the richest or poorest regions in the case of GDP, but not on the majority of regions. Thus, the second hypothesis can not be fully discarded, as social factors clearly play a role in cluster formation, and economic, geographical and political conditions can significantly shape social progress depending on regional context.

Furthermore, some key variables proved particularly important in separating the two clusters and clarifying the nature of these macro-regions. Among the EU-SPI components, *Shelter* emerged as the most influential, while among the latent variables, Eco-social and Institutional Quality was the most significant. These findings highlight the critical role of these policy areas -and particularly *Shelter*, also under ESI Quality- for policymakers. A deeper understanding of these and their potential for improvement could be a valuable instrument to enhance social progress across the EU periphery.

However, the determinants of cluster membership and social progress vary across clusters and regions. Insights from the partial dependence plots further reveal that each component exhibits critical thresholds where minimal changes can produce disproportionately large effects on cluster membership and therefore on social progress. Highlighting that, to maximise social progress, policymakers must consider the particular socioeconomic and political context of the region.

The subdivision into nine clusters revealed distinct models that drive social progress across Europe. Each cluster appears to prioritise different variables of social progress, with varying degrees of success within the EU core and periphery. This diversity demonstrates that social progress can be enhanced through multiple pathways and that no single, uniform model guarantees success.

5. Conclusions

The clustering approach to the EU-SPI reveals that EU regions are best understood as part of a unified system, with a core-periphery divide. This division likely evolved through successive stages, with the Eastern enlargements, the Great Recession and the Sovereign Debt Crisis -particularly affecting the south- serving as major turning points. These changes have profoundly transformed the EU's regional map. The EU-SPI, offering a novel, non-economic measure of cohesion, captures these transformations and provides a new lens through which to interpret Europe's territorial landscape.

The EU-SPI emerges as a valuable complement to traditional economic indicators. While a more in-depth analysis of the results presented here is required before applying these insights to public policy making or expenditure management, there has already been a first attempt on adjusting the allocation of cohesion funds using the EU-SPI (Picazo-Tadeo, et al., 2024), illustrating the potential of this index and these results. A closer examination of component and factor-level data and partial dependence plots can be a starting step to detect potential tipping points between clusters, drive data-based decisions in public policy, and inform both place-based and place-neutral policy strategies (Barca, et al., 2012).

We expect to have provided a new, more realistic view, on the spatial inequalities within the EU and the real challenges to its territorial cohesion. Perceiving the EU at its NUTS-2 level as a core-periphery system could be a major change in our collective understanding of the main challenges ahead for EU policy-makers.

Indeed, locating well-defined periphery regions is essential to recognize those most at risk of development traps (Diemer, et al., 2022). Equally important is understanding regions where social progress does not match economic performance. The paradox of high-performing in SPI yet discontent regions, such as many regions in France or Germany, raises broader questions about expectations and perceptions of not mattering (Rodriguez-Pose, 2018), pointing out that place-based expectations may be built according to per capita GDP, making the sense of not mattering prevail no matter how generous cohesion policies are unfold. We expect, with out contribution, to matter even if the outcomes are uncomfortable or challenging.

Funding

This article was funded by a Collaboration Grant from the Ministry of Education and Pompeu Fabra University awarded to the author Albert Busca.

References

Alhija, F. A. N. (2010). Factor Analysis: An Overview and Some Contemporary Advances. *International Encyclopedia of Education,* 162-170. https://doi.org/10.1016/b978-0-08-044894-7.01328-2

Annoni, P., & Bolsi, P. (2020). The regional dimension of social progress in Europe - presenting the new EU Social Progress Index. European Commission, Directorate-General for Regional and Urban Policy, *Publications Office of the European Union*. https://doi.org/10. 2776/288544

Annoni, P., & Scioni, M. (2022). The unbalance penalisation method for metrics of social progress. *Social Indicators Research*, *162*, 1093-1115. https://doi.org/10.1007/s11205-021-02876-4

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Perez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, *46*(1), 243-256. https://doi.org/10.1016/j.patcog.2012.07.021

Barca, F., McCann, P., & Rodríguez-Pose, A. (2012), The case for regional development intervention: place-based versus place-neutral approaches. *Journal of Regional Science*, *52*(1), 134-152.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *160*(901), 268-282. Retrieved October 15, 2024, from http://www.jstor.org/stable/96803

Beltran-Esteve, M., Peiró-Palomino, J., Picazo-Tadeo, A. J., & Ríos, V. (2023). Is the European Social Progress Index robust? Implications for the design of European Union regional Cohesion policy. *Regional Studies*, *57*(11), 2285-2306. https://doi.org/10.1080/00343404. 2022.2159022

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5-32 https://doi.org/10.1023/A:1010933404324

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research, 21*(2), 230-258. https://doi.org/10.1177/0049124192021002005

Burlina, C., & Rodríguez-Pose, A. (2024). Inequality, poverty, deprivation and the uneven spread of COVID-19 in Europe. *Regional Studies*, *58*(2), 263-284.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1-27. https://doi.org/10.1080/03610927408827101

Chouikhi, H., Charrad, M., & Ghazzali, N. (2015). A comparison study of clustering validity indices. *2015 Global Summit on Computer & Information Technology (GSCIT)*, 1-4. https://doi.org/10.1109/GSCIT.2015.7353330

Del Campo, C., Monteiro, C., & Oliveira, J. (2008). The European regional policy and the socioeconomic diversity of European regions: A multivariate analysis. *European Journal of Operational Research*, *187*(2), 600-612.

Diemer, A., Iammarino, S., Rodríguez-Pose, A., & Storper, M. (2022). The Regional Development Trap in Europe. *Economic Geography*, *98*(5), 487-509.

Dijkstra, L., Poelman, H., & Rodríeguez-Pose, A. (2020). The geography of EU discontent. *Regional Studies*, *54*(6), 737-753.

Dubes, R., & Jain, A. K. (1979). Validity studies in clustering methodologies. *Pattern Recognition*, *11*(4), 235-254. https://doi.org/https://doi.org/10.1016/0031-3203(79)90034-7

Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Princeton U.P.

Genuer, R., Poggi, J., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225-2236. https://doi.org/10.1016/J.PATREC.2010.03.014

Greve, B. (2016). How to Measure Social Progress?. *Social policy & Administration*, *51*(7), 1002-1022.

Huberty, C. (1975). Discriminant Analysis. *Review of Educational Research, 45*(4), 543-598. https://doi.org/10.3102/00346543045004543

Iammarino, S., Rodríguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: evidence, theory and policy implications. *Journal of Economic Geography*, *19*(2), 273-298.

Kaiser, H. F. (1970). A second generation little iffy. *Psychometrika*, *35*, 401-415. https://doi.org/10.1007/BF02291817

Lacmanovic, S., & Tijanic, L. (2025). Social Progress in the European Union: A Comparative Analysis at the National and Regional Levels. *Sustainability, 17*, 1652.

Magone, J., Laffan, B., and Schweiger, C. (eds.) (2016). *Core-periphery relations in the European Union. Power and conflict in a dualist political economy*. Routledge.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159-179. https://doi.org/10.1007/ BF02294245

Monfort, M., Cuestas, J.C., & Ordóñez, J. (2013). Real convergence in Europe: A cluster analysis. *Economic Modelling*, *33*, 689-694.

Peiró-Palomino, J., Picazo-Tadeo, A.J., & Ríos, V. (2023). Social progress around the world: trends and convergence. *Oxford Economic Papers*, *75*(2), 281-306.

Peiró-Palomino, J., Gianmoena, L., Picazo-Tadeo, A.J., & Ríos, V. (2024). Social trust and the advanced aspects of social progress. Evidence for the European regions. *European Journal of Political Economy*, *83*, 102547.

Picazo-Tadeo, A.J, Gianmoena, L., Peiró-Palomino, J., & Ríos, V. (2024), Building a Social Progress-Adjusted Indicator of GDP per Capita for the European Union's Regions. *Social Indicators Research*, *175*, 317-345.

Porter, M.E., Stern, S., & Artavia, R. (2013). *Social Progress Index 2013*. Social Progress Imperative.

Rodríguez-Pose, A. (2018). The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society*, *11*, 189-209.

Rodríguez-Pose, A., Bartalucci, F., Lozano-Gracia, N., & Dávalos, M. (2024). OIvercoming left-behindedness. Moving beyond the efficiency versus equity debate in territorial development. *Regional Science Policy & Practice*, *16*(12), 100144.

Rodríguez-Pose, A., Dijkstra, L., & Poelman, H. (2024). The Geography of EU Discontent and the Regional Development Trap. *Economic Geography*, *100*(3), 213-245.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7

Schwarz, G. (1978). Estimating the Dimension of a Model. *Ann. Statist.* *6*(2), 461-464. https://doi.org/10.1214/aos/1176344136

Siddique, A., Waseem, A., & Mamoon, D. (2017). Did we find alternate to GDP to measure national progress? Analysis of Harvard University's social progress index. *Turkish Economic Review*, *4*, 352-368. https://api.semanticscholar.org/CorpusID:56253185

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *TAXON*, *11*(2), 33-40. https://doi.org/10.2307/1217208

Steiger, J. H. (2016). Notes on the Steiger-Lind (1980) Handout. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 777-781. https://doi.org/10.1080/10705511.2016.1217487

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236-244. https://doi.org/10.1080/01621459.1963.10500845

ORCID

Albert Busca https://orcid.org/0009-0009-6099-6974

Albert Carreras https://orcid.org/0000-0003-0015-2445

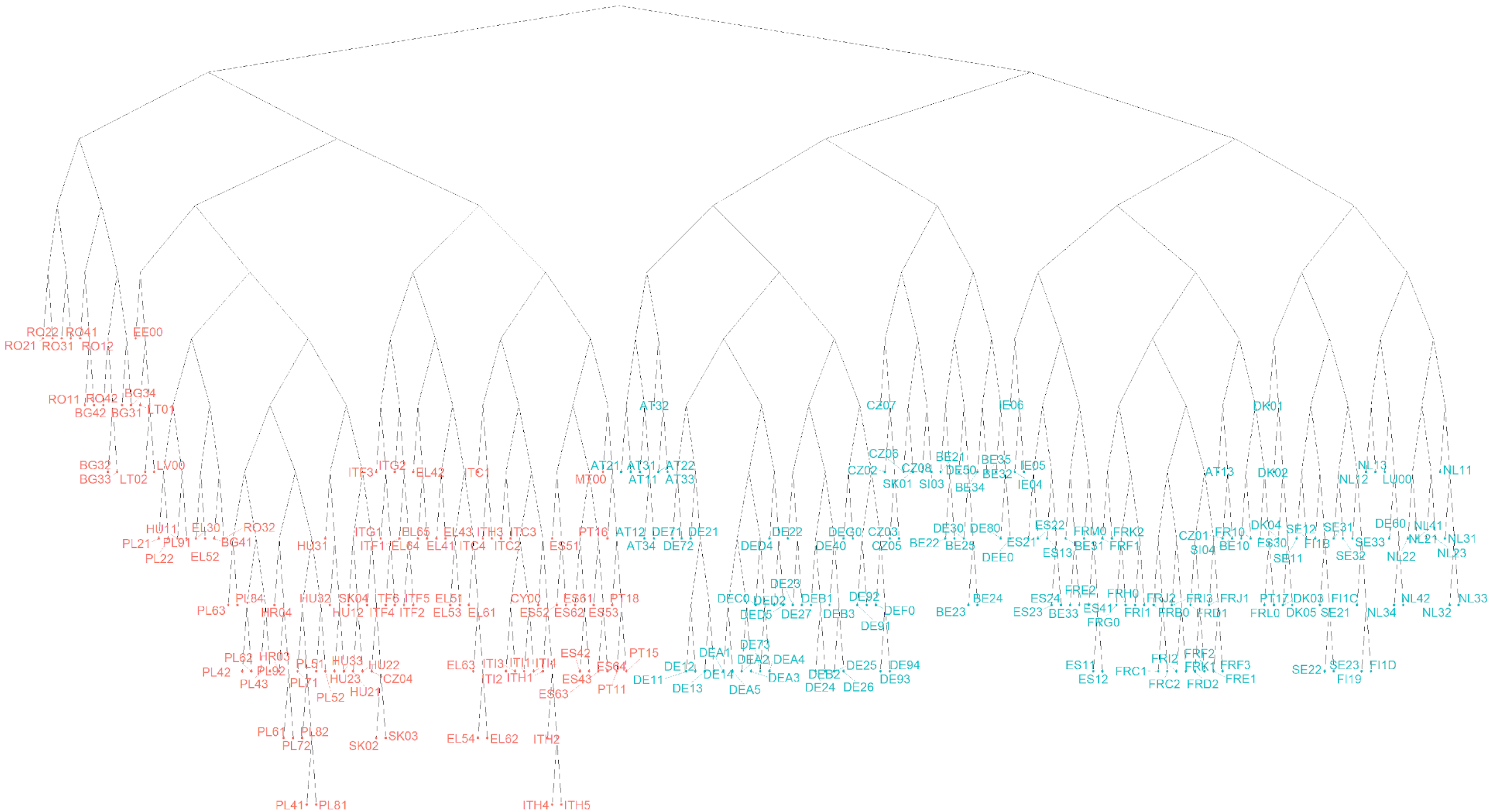
Appendices

Annex 1. Dimensions, components and indicators within the EU-SPI

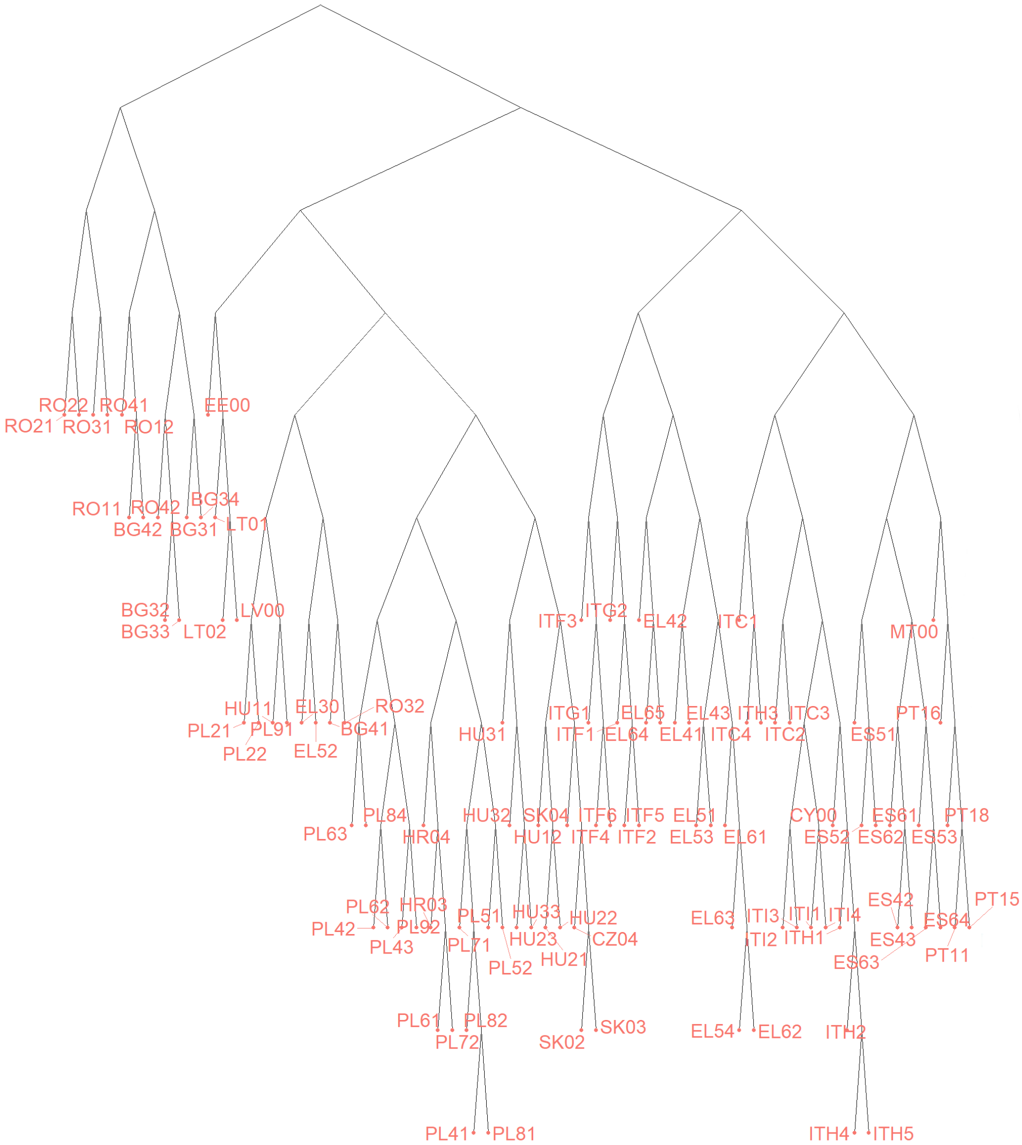
Dimension	Component	Indicator
Basic Human Needs	Nutrition and Basic Medical Care	Premature mortality (<65)
	Nutrition and Basic Medical Care	Infant mortality
	Nutrition and Basic Medical Care	Unmet medical needs
	Nutrition and Basic Medical Care	Insufficient food
	Water and Sanitation	Satisfaction with water quality
	Water and Sanitation	Lack of toilet in dwelling
	Water and Sanitation	Uncollected sewage
	Water and Sanitation	Sewage treatment
	Shelter	Burdensome cost of housing
	Shelter	Housing quality-dampness
	Shelter	Overcrowding
	Shelter	Lack of adequate heating
	Personal Security	Crime
	Personal Security	Safety at night
	Personal Security	Money stolen
	Personal Security	Assaulted/Mugged

Dimension	Component	Indicator
Foundations of Wellbeing	Access to basic knowledge	Upper-secondary enrolment rate (age 14-18)
	Access to basic knowledge	Lower-secondary completion only
	Access to basic knowledge	Early school leavers
	Access to ICT	Internet at home
	Access to ICT	Broadband at home
	Access to ICT	Online interaction with public authorities
	Access to ICT	Internet access
	Health and Wellness	Life expectancy
	Health and Wellness	Subjective health status
	Health and Wellness	Standardised cancer death rate
	Health and Wellness	Standardised heart diseas death rate
	Health and Wellness	Leisure activities
	Health and Wellness	Traffic deaths
	Environmental Quality	Air pollution NO2
	Environmental Quality	Air pollution Ozone
	Environmental Quality	Air pollution pm2.5
	Environmental Quality	Air pollution pm10
Opportunity	Personal Rights	Trust in the national government
	Personal Rights	Trust in the legal system
	Personal Rights	Trust in the police
	Personal Rights	Active citizenship
	Personal Rights	Female participation in regional assemblies
	Personal Rights	Institution quality index
	Personal Freedom and Choice	Freedom over life choices
	Personal Freedom and Choice	Job opportunities
	Personal Freedom and Choice	Involuntary part-time/temporary employment
	Personal Freedom and Choice	Young people, not in education, employment or training (NEET)
	Personal Freedom and Choice	Institutions corruption index
	Tolerance and Inclusion	Institution impartiality Index
	Tolerance and Inclusion	Tolerance towards immigrants
	Tolerance and Inclusion	Tolerance towards minorities
	Tolerance and Inclusion	Tolerance towards homosexuals
	Tolerance and Inclusion	Making friends
	Tolerance and Inclusion	Volunteering
	Tolerance and Inclusion	Gender employment gap
	Access to Advanced Education	Tertiary education attainment
	Access to Advanced Education	Tertiary enrolment
	Access to Advanced Education	Lifelong learning
	Access to Advanced Education	Lifelong learning - female

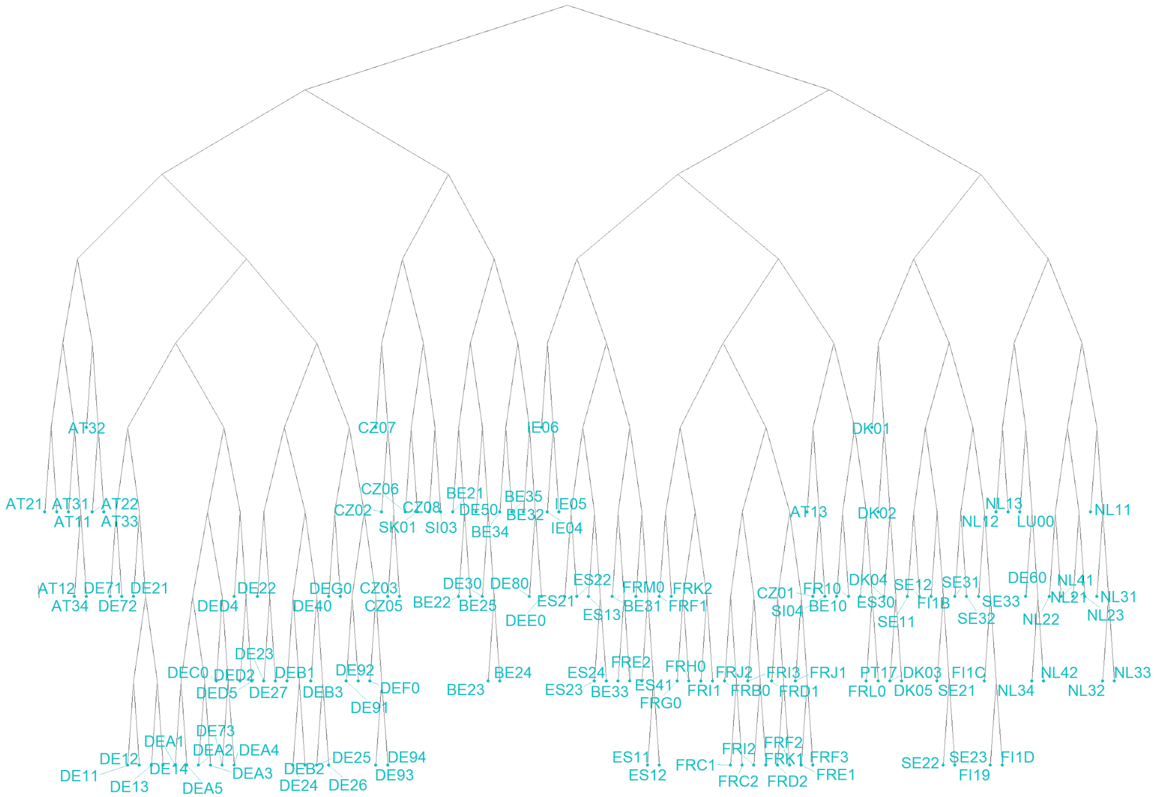
Appendix 2. Dendrogram of the clustering model coloured by cluster membership at 2 clusters



Appendix 3. Dendrogram of the clustering model coloured by cluster membership at 2 clusters, only cluster 1 is shown

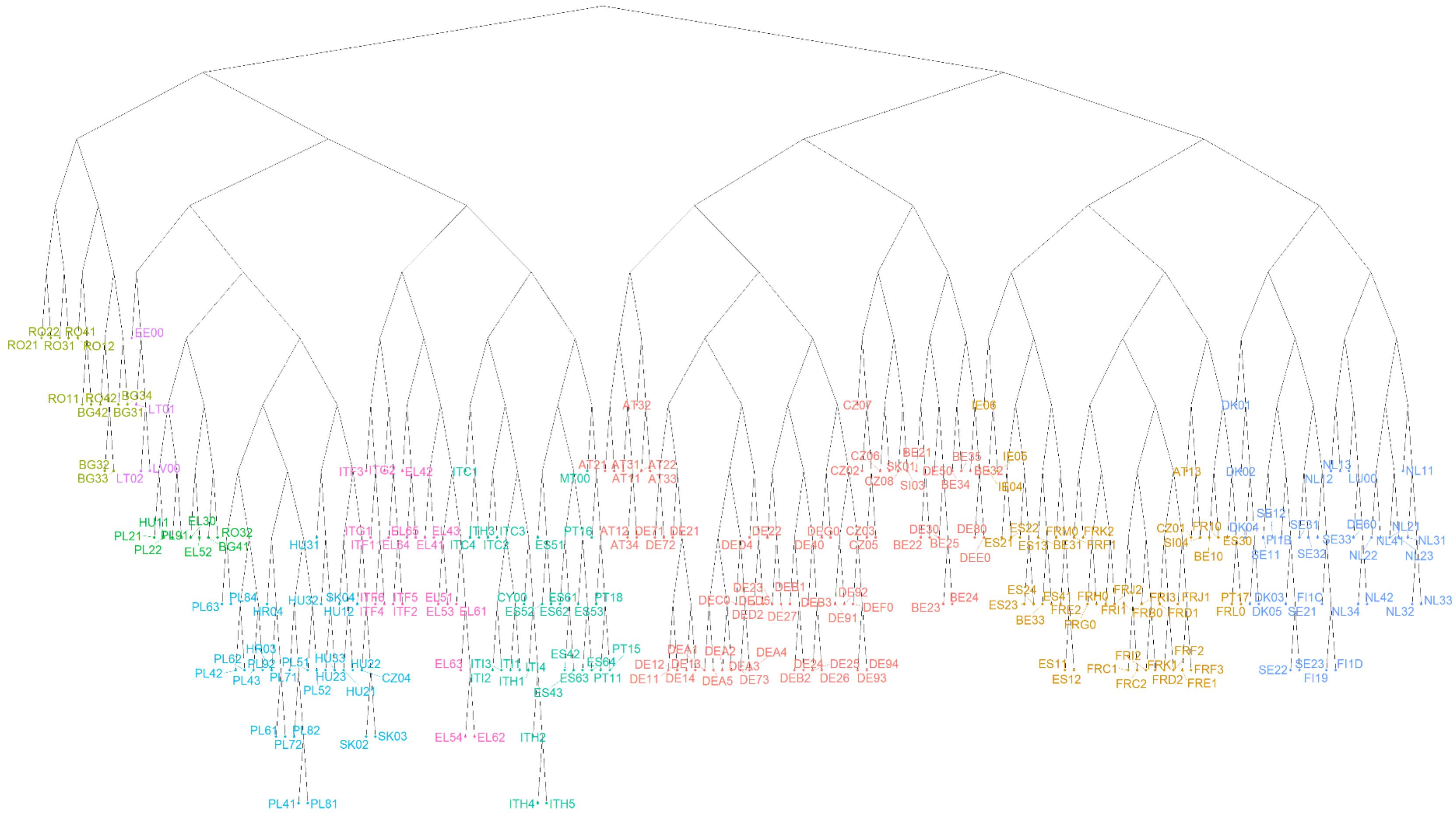


Appendix 4. Dendrogram of the clustering model coloured by cluster membership at 2 clusters, only cluster 2 is shown



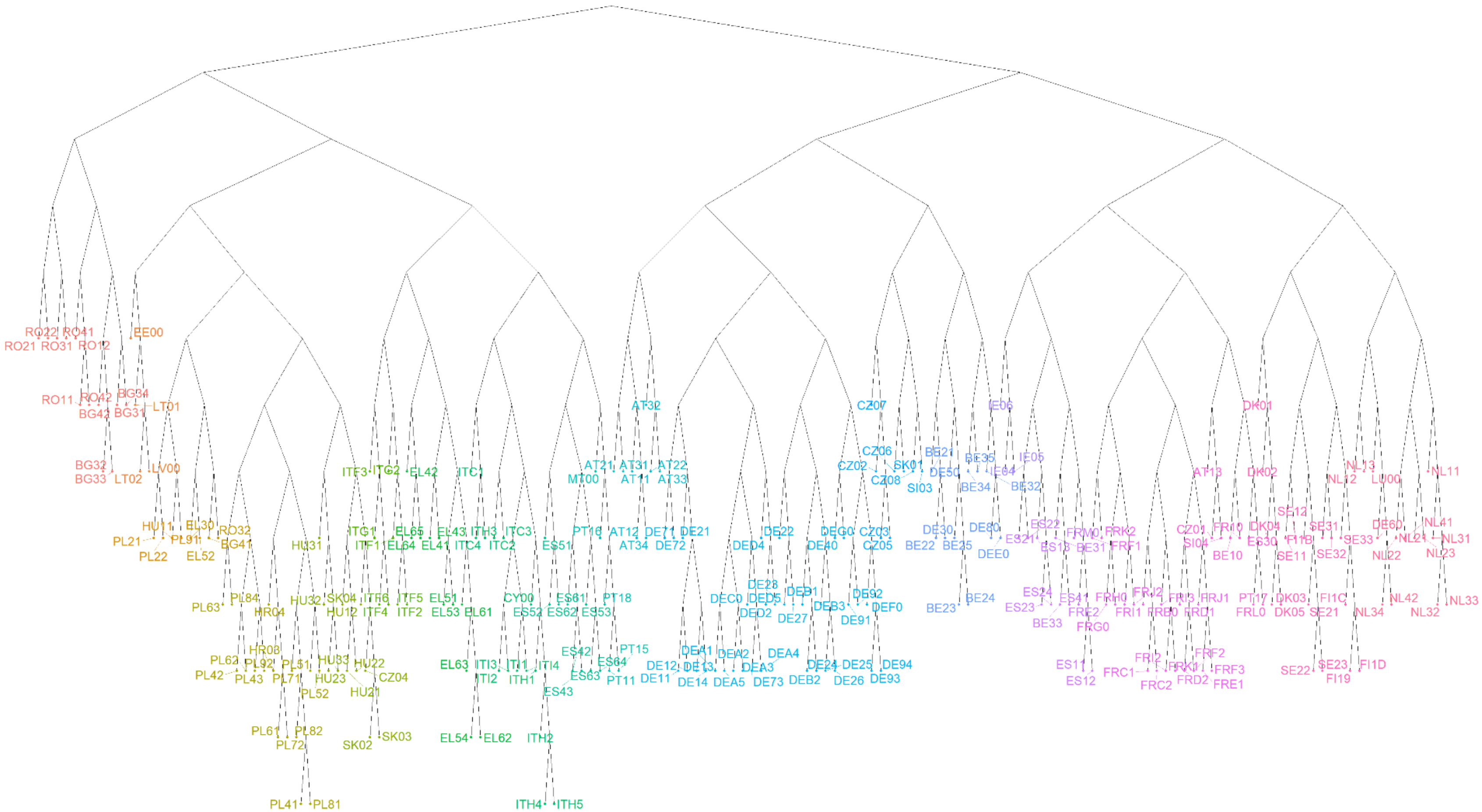
 ISSN: 1695-7253 e-ISSN: 2340-2717

Appendix 5. Dendrogram of the clustering model coloured by cluster membership at 9 clusters



ISSN: 1695-7253 e-ISSN: 2340-2717

Appendix 6. Dendrogram of the clustering model coloured by cluster membership at 21 clusters



 ISSN: 1695-7253 e-ISSN: 2340-2717

Appendix 7. NUTS-2 ID to region name conversion

NUTS ID	Region name*
AT11	Burgenland (1, 1, 1)
AT12	Niederösterreich (1, 1, 1)
AT13	Wien (1, 2, 2)
AT21	Kärnten (1, 1, 1)
AT22	Steiermark (1, 1, 1)
AT31	Oberösterreich (1, 1, 1)
AT32	Salzburg (1, 1, 1)
AT33	Tirol (1, 1, 1)
AT34	Vorarlberg (1, 1, 1)
BE10	Rég. de Bruxelles-Cap./Brussels Hfst. Gew. (1, 2, 2)
BE21	Antwerpen (1, 1, 3)
BE22	Limburg (1, 1, 3)
BE23	Oost-Vlaanderen (1, 1, 3)
BE24	Vlaams-Brabant (1, 1, 3)
BE25	West-Vlaanderen (1, 1, 3)
BE31	Brabant Wallon (1, 2, 4)
BE32	Hainaut (1, 1, 3)
BE33	Liège (1, 2, 4)
BE34	Luxembourg (1, 1, 3)
BE35	Namur (1, 1, 3)
BG31	Severozapaden (2, 3, 5)
BG32	Severen tsentralen (2, 3, 5)
BG33	Severoiztochen (2, 3, 5)
BG34	Yugoiztochen (2, 3, 5)
BG41	Yugozapaden (2, 4, 6)
BG42	Yuzhen tsentralen (2, 3, 5)
CY00	Kýpros (2, 5, 7)
CZ01	Praha (1, 2, 2)
CZ02	Střední Čechy (1, 1, 8)
CZ03	Jihozápad (1, 1, 8)
CZ04	Severozápad (2, 6, 9)
CZ05	Severovýchod (1, 1, 8)
CZ06	Jihovýchod (1, 1, 8)
CZ07	Střední Morava (1, 1, 8)
CZ08	Moravskoslezsko (1, 1, 8)
DE11	Stuttgart (1, 1, 10)
DE12	Karlsruhe (1, 1, 10)

NUTS ID	Region name*
DE13	Freiburg (1, 1, 10)
DE14	Tübingen (1, 1, 10)
DE21	Oberbayern (1, 1, 10)
DE22	Niederbayern (1, 1, 10)
DE23	Oberpfalz (1, 1, 10)
DE24	Oberfranken (1, 1, 10)
DE25	Mittelfranken (1, 1, 10)
DE26	Unterfranken (1, 1, 10)
DE27	Schwaben (1, 1, 10)
DE30	Berlin (1, 1, 3)
DE40	Brandenburg (1, 1, 10)
DE50	Bremen (1, 1, 3)
DE60	Hamburg (1, 7, 11)
DE71	Darmstadt (1, 1, 10)
DE72	Gießen (1, 1, 10)
DE73	Kassel (1, 1, 10)
DE80	Mecklenburg-Vorpommern (1, 1, 3)
DE91	Braunschweig (1, 1, 10)
DE92	Hannover (1, 1, 10)
DE93	Lüneburg (1, 1, 10)
DE94	Weser-Ems (1, 1, 10)
DEA1	Düsseldorf (1, 1, 10)
DEA2	Köln (1, 1, 10)
DEA3	Münster (1, 1, 10)
DEA4	Detmold (1, 1, 10)
DEA5	Arnsberg (1, 1, 10)
DEB1	Koblenz (1, 1, 10)
DEB2	Trier (1, 1, 10)
DEB3	Rheinhessen-Pfalz (1, 1, 10)
DEC0	Saarland (1, 1, 10)
DED2	Dresden (1, 1, 10)
DED4	Chemnitz (1, 1, 10)
DED5	Leipzig (1, 1, 10)
DEE0	Sachsen-Anhalt (1, 1, 3)
DEF0	Schleswig-Holstein (1, 1, 10)
DEG0	Thüringen (1, 1, 10)
DK01	Hovedstaden (1, 7, 12)
DK02	Sjælland (1, 7, 12)

NUTS ID	Region name*
DK03	Syddanmark (1, 7, 12)
DK04	Midtjylland (1, 7, 12)
DK05	Nordjylland (1, 7, 12)
EE00	Eesti (2, 8, 13)
EL30	Attiki (2, 4, 6)
EL41	Voreio Aigaio (2, 9, 14)
EL42	Notio Aigaio (2, 9, 14)
EL43	Kriti (2, 9, 14)
EL51	Anatoliki Makedonia, Thraki (2, 9, 14)
EL52	Kentriki Makedonia (2, 4, 6)
EL53	Dytiki Makedonia (2, 9, 14)
EL54	Ipeiros (2, 9, 14)
EL61	Thessalia (2, 9, 14)
EL62	Ionia Nisia (2, 9, 14)
EL63	Dytiki Ellada (2, 9, 14)
EL64	Sterea Ellada (2, 9, 14)
EL65	Peloponnisos (2, 9, 14)
ES11	Galicia (1, 2, 4)
ES12	Principado de Asturias (1, 2, 4)
ES13	Cantabria (1, 2, 4)
ES21	País Vasco (1, 2, 4)
ES22	Comunidad Foral de Navarra (1, 2, 4)
ES23	La Rioja (1, 2, 4)
ES24	Aragón (1, 2, 4)
ES30	Comunidad de Madrid (1, 2, 2)
ES41	Castilla y León (1, 2, 4)
ES42	Castilla-La Mancha (2, 5, 15)
ES43	Extremadura (2, 5, 15)
ES51	Cataluña (2, 5, 15)
ES52	Comunidad Valenciana (2, 5, 15)
ES53	Illes Balears (2, 5, 15)
ES61	Andalucía (2, 5, 15)
ES62	Región de Murcia (2, 5, 15)
ES63	Ciudad Autónoma de Ceuta (2, 5, 15)
ES64	Ciudad Autónoma de Melilla (2, 5, 15)

NUTS ID	Region name*
FI19	Länsi-Suomi (1, 7, 12)
FI1B	Helsinki-Uusimaa (1, 7, 12)
FI1C	Etelä-Suomi (1, 7, 12)
FI1D	Pohjois- ja Itä-Suomi (1, 7, 12)
FR10	Île de France (1, 2, 2)
FRB0	Centre - Val de Loire (1, 2, 16)
FRC1	Bourgogne (1, 2, 16)
FRC2	Franche-Comté (1, 2, 16)
FRD1	Basse-Normandie (1, 2, 16)
FRD2	Haute-Normandie (1, 2, 16)
FRE1	Nord-Pas de Calais (1, 2, 16)
FRE2	Picardie (1, 2, 4)
FRF1	Alsace (1, 2, 16)
FRF2	Champagne-Ardenne (1, 2, 16)
FRF3	Lorraine (1, 2, 16)
FRG0	Pays de la Loire (1, 2, 16)
FRH0	Bretagne (1, 2, 16)
FRI1	Aquitaine (1, 2, 16)
FRI2	Limousin (1, 2, 16)
FRI3	Poitou-Charentes (1, 2, 16)
FRJ1	Languedoc-Roussillon (1, 2, 16)
FRJ2	Midi-Pyrénées (1, 2, 16)
FRK1	Auvergne (1, 2, 16)
FRK2	Rhône-Alpes (1, 2, 16)
FRL0	Provence-Alpes-Côte d'Azur (1, 2, 2)
FRM0	Corse (1, 2, 4)
HR03	Jadranska Hrvatska (2, 6, 17)
HR04	Kontinentalna Hrvatska (2, 6, 17)
HU11	Budapest (2, 4, 18)
HU12	Pest (2, 6, 9)
HU21	Közép-Dunántúl (2, 6, 9)
HU22	Nyugat-Dunántúl (2, 6, 9)

NUTS ID	Region name*
HU23	Dél-Dunántúl (2, 6, 9)
HU31	Észak-Magyarország (2, 6, 9)
HU32	Észak-Alföld (2, 6, 9)
HU33	Dél-Alföld (2, 6, 9)
IE04	Northern and Western (1, 2, 19)
IE05	Southern (1, 2, 19)
IE06	Eastern and Midland (1, 2, 19)
ITC1	Piemonte (2, 5, 7)
ITC2	Valle d'Aosta/Vallée d'Aoste (2, 5, 7)
ITC3	Liguria (2, 5, 7)
ITC4	Lombardia (2, 5, 7)
ITF1	Abruzzo (2, 9, 20)
ITF2	Molise (2, 9, 20)
ITF3	Campania (2, 9, 20)
ITF4	Puglia (2, 9, 20)
ITF5	Basilicata (2, 9, 20)
ITF6	Calabria (2, 9, 20)
ITG1	Sicilia (2, 9, 20)
ITG2	Sardegna (2, 9, 20)
ITH1	Prov. Autonoma di Bolzano/Bozen (2, 5, 7)
ITH2	Provincia Autonoma di Trento (2, 5, 7)
ITH3	Veneto (2, 5, 7)
ITH4	Friuli-Venezia Giulia (2, 5, 7)
ITH5	Emilia-Romagna (2, 5, 7)
ITI1	Toscana (2, 5, 7)
ITI2	Umbria (2, 5, 7)
ITI3	Marche (2, 5, 7)
ITI4	Lazio (2, 5, 7)
LT01	Sostinės regionas (2, 8, 13)
LT02	Vidurio ir vakarų Lietuvos regionas (2, 8, 13)
LU00	Luxembourg (1, 7, 11)
LV00	Latvija (2, 8, 13)
MT00	Malta (2, 5, 21)
NL11	Groningen (1, 7, 11)

NUTS ID	Region name*
NL12	Friesland (1, 7, 11)
NL13	Drenthe (1, 7, 11)
NL21	Overijssel (1, 7, 11)
NL22	Gelderland (1, 7, 11)
NL23	Flevoland (1, 7, 11)
NL31	Utrecht (1, 7, 11)
NL32	Noord-Holland (1, 7, 11)
NL33	Zuid-Holland (1, 7, 11)
NL34	Zeeland (1, 7, 11)
NL41	Noord-Brabant (1, 7, 11)
NL42	Limburg (1, 7, 11)
PL21	Małopolskie (2, 4, 18)
PL22	Śląskie (2, 4, 18)
PL41	Wielkopolskie (2, 6, 17)
PL42	Zachodniopomorskie (2, 6, 17)
PL43	Lubuskie (2, 6, 17)
PL51	Dolnośląskie (2, 6, 17)
PL52	Opolskie (2, 6, 17)
PL61	Kujawsko-pomorskie (2, 6, 17)
PL62	Warmińsko-mazurskie (2, 6, 17)
PL63	Pomorskie (2, 6, 17)
PL71	Łódzkie (2, 6, 17)
PL72	Świętokrzyskie (2, 6, 17)
PL81	Lubelskie (2, 6, 17)
PL82	Podkarpackie (2, 6, 17)
PL84	Podlaskie (2, 6, 17)
PL91	Warszawski stołeczny (2, 4, 18)
PL92	Mazowiecki regionalny (2, 6, 17)
PT11	Norte (2, 5, 21)
PT15	Algarve (2, 5, 21)
PT16	Centro (2, 5, 21)
PT17	Área Metr. de Lisboa (1, 2, 2)
PT18	Alentejo (2, 5, 21)
RO11	Nord-Vest (2, 3, 5)
RO12	Centru (2, 3, 5)

ISSN: 1695-7253 e-ISSN: 2340-2717

NUTS ID	Region name*
RO21	Nord-Est (2, 3, 5)
RO22	Sud-Est (2, 3, 5)
RO31	Sud - Muntenia (2, 3, 5)
RO32	Bucureşti - Ilfov (2, 4, 6)
RO41	Sud-Vest Oltenia (2, 3, 5)
RO42	Vest (2, 3, 5)
SE11	Stockholm (1, 7, 12)
SE12	Östra Mellansverige (1, 7, 12)
SE21	Småland med öarna (1, 7, 12)
SE22	Sydsverige (1, 7, 12)
SE23	Västsverige (1, 7, 12)

NUTS ID	Region name*
SE31	Norra Mellansverige (1, 7, 12)
SE32	Mellersta Norrland (1, 7, 12)
SE33	Övre Norrland (1, 7, 12)
SI03	Vzhodna Slovenija (1, 1, 8)
SI04	Zahodna Slovenija (1, 2, 2)
SK01	Bratislavský kraj (1, 1, 8)
SK02	Západné Slovensko (2, 6, 9)
SK03	Stredné Slovensko (2, 6, 9)
SK04	Východné Slovensko (2, 6, 9)

*Each region name is followed by three numbers that represent cluster membership when there are 2 clusters, when there are 9 clusters and when there are 21 clusters.