Regional economic disparities in Europe: Time-series Clustering of NUTS 3 regions

Ana María López-Villuendas[a*], Cristina del Campo[a]

[a] Departamento de Economía Financiera y Actuarial y Estadística, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 – Madrid, Spain

[*] Corresponding author.

E-mail address: anamlo09@ucm.es

ABSTRACT

The aim of this research is to identify the regional economic disparities in the level of economic wealth and its dynamics in the NUTS 3 regions in EU28 over the period from 2000 to 2017. By performing a time-series clustering analysis at NUTS 3 level, we expect to uncover the economic disparities that might have been hidden in the aggregate NUTS 2 regions. Our results indicate that at a finer spatial scale (NUTS 3 level) disparities flourish, particularly in the period after the global crisis of 2008, in which different recovery rates are observed. In general, NUTS 2 regions tend to spatially cluster at the national level and, although NUTS 3 regions show slightly this tendency as well, the spatial effect is not as strong as it is for NUTS 2 level, revealing specific behaviours of the local economies and markets that remain hidden at the aggregate NUTS 2 level.

Keywords: Time-series Clustering, European NUTS 3 Regions, Cohesion Policy, regional inequalities

1. Introduction

Regional disparities in the European Union (EU) have led to an intense debate in the last decades due to its importance in ensuring the well-balanced and harmonious development of the regions. Reduction of disparities is one of the main priorities of the Cohesion Policy, since their presence may jeopardise the enhance of competitiveness, integration and cohesion of the EU. Article 158 of the Treaty of the European Union states "in particular, the Community aims to reduce the disparities between the levels of development of the different regions". To that end, the Cohesion Policy dedicates a large quantity of funds, targeted specially to lagging regions. However, despite the labour of the Cohesion Policy, regional disparities have been increasing throughout the last years fostered by the enlargement of the European Union with the accession of new countries and the global financial crisis (Heidenreich and Wunder 2008; Meliciani 2015; Mora and Vaya 2004). Regional inequalities substantially increased as the ratio between the richest and the poorest regions was about one to five for EU15 in 2000, but one to nine for EU25 in 2004 and one to 13 for EU27 in 2007. As highlighted in European Commission (2003, 2004), the enlargement would produce an increase of about 30 percent of the total European area and an increase of more than 25 percent of the population, but GDP would increase by only 5 percent. There are several authors who have compared the regional disparities in the economic performance before and after the enlargement and confirmed this increase in the level of regional disparities within the EU derived from that enlargement (e.g. Ezcurra 2019; Heidenreich and Wunder 2008).

The study of regional disparities is essential in order to highlight the deficiencies and problematic issues in the regions and to identify the potential areas of development. In the literature different approaches and methodologies for measuring disparities can be found (Kutscherauer et al. 2010). One of them is cluster analysis, which can be employed to classify and measure the disparities since this method enables to group the regions with similar characteristics in homogeneous groups or clusters.

The evolution of disparities in the European regions are usually measured at NUTS 2 level (e.g. Ezcurra 2019; Monfort 2008; Tvrdoň 2012), as the Cohesion Policy is defined and the allocation of the funds is carried out at this level. However, the allocation of the funds within NUTS 2 regions may be uneven throughout the NUTS 3 regions since the eligibility for receiving funds is defined at NUTS 2 level regardless of the economic performance of their NUTS 3 regions (Gagliardi and Percoco 2017). By performing the analysis at a finer geographical scale, i.e., the NUTS 3 level, some local effects that are not evident at NUTS 2 level might be revealed (Fratesi and Wishlade 2017). This level enables a more efficient identification of the cohesion programme's impact as NUTS 2 regions are too broad regional aggregates, particularly in the southern countries, which may cover the specific patterns of their NUTS 3 regions. Additionally, there are some countries such as Slovakia (Leška 2015), Czech Republic (Illner 2010) or Hungary (Tvrdoň and Skokan 2011), whose NUTS 3 regions are the higher-level territorial self-governing units. Actually, for these countries, the NUTS 2 regions were artificially created for the allocation of the Structural Funds as a combination of one to three NUTS 3 regions.

As stated above, the number of disparities at NUTS 3 level is rising and policies such as the Cohesion Policy are implemented at NUTS 2 level regardless of the performance at the

NUTS 3 level. Due to the huge amount of funds dedicated to the Cohesion Policy, it is necessary to determine whether the current regional level at which the funds are allocated is masking the economic inequalities of their NUTS 3 regions. However, only the evolution of the NUTS 3 economic disparities in last decades and the whole EU has been analysed in literature (Butkus et al. 2018; Goecke and Hüther 2016), while we also incorporate the patterns of that evolution in our study.

The purpose of this research then, is to uncover the NUTS 3 economic patterns which may be hidden in their aggregates NUTS 2 level by performing a model-based clustering, which will account for the distribution of the data and will allow the inclusion of the time dimension in the clustering, thus getting an insight of the patterns of the economic disparities through time. In order to do that, we will identify the regional economic disparities in the level of economic wealth and its dynamics in all 1348 NUTS 3 regions in EU28 throughout the recent time period 2000-2017 and contrast them to the ones obtained at NUTS 2 level.

The contribution to the literature will be threefold. On one hand we will identify the economic disparities at both NUTS 2 and NUTS 3 levels, and by contrasting them, we will determine whether the NUTS 3 regions share the same behaviours as their aggregates NUTS 2 regions and therefore policies such as the Cohesion Policy are equally effective for all territories. On the other hand, we will also obtain groups of regions sharing similar economic characteristics throughout all the covered period which can be used as a starting point in future empirical studies focussed on exploring the driving forces behind the regional economic disparities. Finally, as we found that the NUTS 2 level is masking the heterogeneity of their NUTS 3 regions, we made an approach to characterise the reasons behind the revealed NUTS 3 economic disparities.

The article is organized as follows. Section 2 reviews the literature on regional disparities. Section 3 describes the methodology and the data used for the analysis while section 4 presents and discusses the obtained results. Finally, section 5 concludes the paper.

2. Background Literature

Ever since the early time of the European integration process, the issue of regional inequality has always been an important concern for the EU. The Cohesion Policy aims to reduce disparities between EU regions in order to achieve balanced economic, social and territorial development.

The analysis of regional disparities in the European Union is a recurrent theme in research studies since it serves as an indicator of integration and cohesion of the European Union and it enables to assess the performance of the Cohesion Policy.

Economic disparities are the most addressed ones in inequality literature and are usually measured by several inequality indicators processed by different statistical and mathematical methods.

Many studies addressing economic disparities in the EU have examined the effect of regional inequalities on the development of the regions by performing growth regressions (de Dominicis 2014; Lessmann 2014; Panzera and Postiglione 2021; Petrakos, Rodríguez-

Pose, and Rovolis 2003). Some authors have focussed on determining possible factors of regional disparities within EU countries (e.g. Ezcurra 2019; Kyriacou and Roca-Sagalés 2012) while the gross of the literature aims at measuring the magnitude and evolution of regional inequalities. An uniformed criteria has not been reached by researchers when measuring regional disparities and, in fact, different approaches based on statistical or mathematical methods can be found in the literature (Kutscherauer et al. 2010).

Regional economic disparities are usually calculated by some of the numerous measures of inequality existing in the literature, based on the dispersion of the GDP indicator, such as the coefficient of variation (e.g. Butkus et al. 2018; Ezcurra 2019; Goecke and Hüther 2016; Kokocińska and Puziak 2020; Monfort 2008; Postiglione, Cartone, and Panzera 2020; Smetkowski and Wójcik 2012; Tvrdoň 2012; Tvrdoň and Skokan 2011), the Gini index (e.g. Butkus et al. 2018; Ezcurra 2019; Monfort 2008; Mora and Vaya 2004; Tvrdoň and Skokan 2011), the Theil index (e.g. Butkus et al. 2018; Ezcurra 2019; Monfort 2008; Tvrdoň 2012), the mean logarithmic deviation (e.g. Butkus et al. 2018; Ezcurra 2019; Heidenreich and Wunder 2008; Monfort 2008) or the Atkinson index, among others (e.g. Monfort 2008). Other measures include the dynamics of the distribution of the data, such as the kernel density function (e.g. Mora and Vaya 2004; Kokocińska and Puziak 2020), Markov chain analysis (e.g. le Gallo 2004), non-parametric estimation of density functions, cumulative density functions and Salter graphs (e.g. Geppert and Stephan 2008; Monfort 2008) or the spatial component (e.g. Ertur and Koch 2006; le Gallo 2004; le Gallo and Ertur 2005; Smetkowski and Wójcik 2012). Finally, other authors have considered multivariate statistical methods, adding more indicators, such as cluster analysis and factor analysis (e.g. del Campo, Monteiro, and Soares 2008; Pavone et al. 2021; Soares, Lourenço, and Ferreira 2003) or multicriteria decision-making methods (e.g. Poledníková 2014).

One widely shared result of many of these studies is that there has been a decrease in between-state inequalities while within-state inequalities have increased (e.g. Butkus et al. 2018; European Commission 2003; Ezcurra 2019; Heidenreich and Wunder 2008; Meliciani 2015; Monfort 2008; Puga 2002). For example, Heidenreich and Wunder (2008), measuring regional disparities with the decomposed form of mean logarithmic deviation, stated that within-states regional inequalities in the enlarged EU increased by 15 percent over the period from 1995 to 2003, while the between-state inequalities decreased by 45 percent. They determined that most of the regional economic inequalities in the EU are within-states (67 percent at the NUTS 3 level, 84 percent at the NUTS 2 level). More recently, Butkus et al. (2018), analysing economic disparities with the decomposed form of Theil index over the period 1995-2014, stated that disparities between countries are diminishing, accounting for 60.4 percent of total disparities in the EU in 2014 compared to 76.4 percent in 1995, while within-state inequalities are increasing. 23.8 percent of total inequalities were attributed to inequalities between NUTS 2 regions within the country (13.4 percent in 1995) and 15.8 percent to inequalities between NUTS 3 regions within NUTS 2 regions (10.2 percent in 1995).

Despite the publication of numerous studies on the evolution of regional disparities, the territory they usually address is the NUTS 2 level, paying less attention to NUTS 3. At this level, the studies focus generally on groups of countries (e.g. Smetkowski and Wójcik (2012) or Tvrdoň and Skokan (2011) in eastern Europe) or old members (Geppert and Stephan 2008; Postiglione, Cartone, and Panzera 2020). Although there are some contributions addressing the whole European Union. For example, Heidenreich and Wunder (2008), studying the evolution of regional inequalities over 1214 NUTS 3 regions in the EU25 from 1995 to 2003 using the Gini coefficient and the mean logarithmic deviation, stated that the disparities at NUTS 3 level increased by 15 percent over that period. Goecke and Hüther (2016), using the coefficient of variation as the measure of regional inequality over 1289 NUTS 3 regions in EU28, determined that the disparities declined from 2000 to 2009 but increased until 2011. More recently, Butkus et al. (2018), studying economic inequalities with the coefficient of variation over the period 1995-2014 in 1342 NUTS 3 regions in EU28, obtained that disparities at NUTS 3 level increased by 8.3 percent over the studied period. In 2014, disparities at the NUTS 3 level were bigger by 70 percent compared with the national level, while the initial difference was 41 percent. On the other hand, disparities at the NUTS 2 level increased (1.2 percent higher in 2014 than in 1995) and were 28.3 percent higher than disparities at the country level.

Measures based on the dispersion of GDP, such as the coefficient of variation, Gini index, Theil index, etc, provide relevant information regarding economic disparities, but may be considered as insufficient. Kokocińska and Puziak (2020), Quah (1993a, 1993b), or Wójcik (2016), among others, argued that dispersion indicators do not provide any information on the behaviour of the GDP distribution. By performing a model-based clustering, the grouping of the regions is carried out according to their GDP distribution. In particular, observations belonging to the same cluster, i.e., regions which similar GDP characteristics in the studied period, are generated by the same distribution. Additionally, unlike other clustering algorithms, model-based clustering allows to deal with time-series, thus including time as an essential characteristic of the analysis. Other advantage with respect to other clustering algorithms is that it selects the optimal number of clusters and the appropriate model according to which the data is distributed.

Since one of the most important descriptors of regional economic differences is the region's performance and is mainly measured by GDP, GDP per capita has been chosen as the indicator for measuring the economic disparities. Hence economic inequalities will be analysed for levels of GDP per capita, and although the level of GDP finally reflects the cumulative rate of economic growth for the covered period, we will examine the rate of GDP growth as well, in order to account for the economic dynamism.

3. Methodology

3.1. Clustering

An uniformed criteria has not been reached by researchers when measuring regional disparities. In fact, different approaches based on statistical or mathematical methods can be found in the literature (Kutscherauer et al. 2010). Cluster analysis constitutes one of the multivariate methodologies to determine disparities by classifying regions with similar characteristics into clusters which are homogeneous within and heterogeneous between. Concretely, time-series clustering enables to identify similarities in the economic time trends of the different territories.

However, time-series present peculiar nuances, such as high dimensionality, high feature correlation or large amount of noise, so that conventional clustering algorithms do not work properly (Lin et al. 2004). In the literature, three ways of addressing clustering time-series are presented (Aghabozorgi, Seyed Shirkhorshidi, and Ying Wah 2015): feature-based, shape-based and model-based approaches.

In the feature-based approach, the raw time-series are converted into a feature vector of lower dimension, therefore reducing the time-series dimensionality using a representation method. Later, a conventional clustering algorithm is applied to the extracted feature vectors. This approach is very common due to the high dimensionality of time-series since it reduces the effect of noise and has a low complexity of distance calculation. However, reduction of dimensionality can cause overlooking of data (Lai, Chung, and Tseng 2010).

In the shape-based approach, clusters are generated based on similarity in shape (the time of occurrence of patterns is not important to find similar time-series in shape). This approach works directly with the raw time-series data and usually employs conventional clustering methods, whose distance measure has been adapted for time-series. Distances such as Dynamic Time Warping (Chu et al. 2002) can calculate the similarity in shape between time-series. This approach is not usually employed in existing works since its high computational demand makes its use in large time-series datasets expensive.

Finally, in model-based methods, raw time-series are transformed into model parameters and then a conventional clustering algorithm is applied to the extracted model parameters. This approach considers that data is generated by a model or by a mixture of underlying probability distributions, each of which represents a different cluster. By fitting the assumed mixture model to observed data, observations can be classified as belonging to the cluster under which they have the greatest probability of occurring (Warren Liao 2005). One advantage of model-based clustering is that it provides a precise framework for assessing the resulting partitions of the data and especially for choosing a relevant number of clusters (Biernacki, Celeux, and Govaert 2000).

In our study we choose to perform a model-based clustering to group the NUTS 3 European regions which have similar time tendencies according to their level of GDP per capita. This type of clustering has been proved to be useful in economics since it accounts for unobserved heterogeneity in the data (Grün 2019). Its main benefit, with respect to the conventional clustering algorithms such as the hierarchical clustering or k-means, is the selection of the number of clusters and the appropriate model according to which the data is distributed.

The most popular model-based approach is the Gaussian mixture models (GMM) in which each component or cluster is modelled by a Gaussian distribution parametrized by its mean and covariance matrix. Even if an underlying mixture component is non-normal, it can be approximated by several Gaussian components (Fraley and Raftery 1998). Model-based clustering provides then a soft assignment where each time-series has a probability of belonging to each cluster.

Given data $\overline{y}$ with independent multivariate observations $y_1, \ldots, y_n$ the likelihood for a mixture model with G components is:

$$L_{mix}(\theta_1, \ldots, \theta_G; \tau_1, \ldots, \tau_G \vee \overline{y}) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(y_i \vee \theta_k) \quad (1)$$

where f_k and θ_k are, respectively, the density and parameters of the *k*th component in the mixture and τ_k is the probability that an observation belongs to the *k*th component. f_k is the multivariate normal (Gaussian) density ϕ_k, parametrized by its mean μ_k and covariance matrix Σ_k:

$$\phi_k(y_i|\mu_k, \Sigma_k) = \frac{exp\{-1/2\,(y_i - \mu_k)^T \sum_k^{-1}(y_i - \mu_k)\}}{\sqrt{det\,(2\pi\Sigma_k)}} \qquad (2)$$

The covariance matrix of Eq. 2 determines the geometric features of the clusters, that is, their shape, volume and orientation, and is computed by Expectation Maximization (EM) algorithms.

In our study, model-based clustering using Gaussian mixed models has been performed with the statistical software R version 3.6.1 and the package *mclust* (Scrucca et al. 2016). *mclust* uses an identifier for each possible parametrization of the covariance matrix. The first identifier refers to volume, the second to shape and the third to orientation and each have three possibilities identified as E for "equal", V for "variable" and I for "coordinate axes". There are a total of 14 combinations of volume, shape and orientation or models (see Appendix Table A1 for a detailed description of the 14 combinations).

The most appropriate models were those allowing for the most homogeneous grouping of time-series regarding their patterns of variation, selected among those with the lowest Bayesian Information Criterion (BIC). Apart from selecting the parametrization of the model, BIC selects the number of clusters as well (Fraley and Raftery 2002).

3.2. Data description

The main variable in our analysis is the Gross Domestic Product in Purchasing Power Standars (GDP PPS) per capita, standard measure of the degree of prosperity, collected from the Cambridge Econometrics' European Regional Database that is available from the new Annual Regional Database of the European Commission's Directorate General for Regional and Urban Policy (ARDECO 2020).

The geographical units considered in this study are all the 1348 NUTS 3 regions in EU-28 according to the NUTS 2016 classification. Nomenclature of Territorial Units for Statistics (NUTS) is a hierarchical classification on regional levels from 1 to 3 that subdivides each Member State into a number of regions of level NUTS 1 (e.g. Baden-Württemberg in Germany or Isole in Italy). Each of them is then subdivided in regions of level NUTS 2 (e.g. Stuttgart and Sicilia, respectively) and these are divided, at their turn, in regions of level NUTS 3 (e.g. Böblingen and Palermo, respectively). NUTS 1 level identifies larger socio-economic regions, NUTS 2 level indicates basic regions and NUTS 3 the smaller regions. The NUTS classification was created in order to enable the collection of harmonised regional statistics in the EU and all statistics transmitted by the Member States to the Commission should use the NUTS classification[1] (Regulation EC No. 1059/2003 of the European Parliament and of the Council, 2003).

The NUTS 3 level is an ideal unit of analysis because throughout most EU countries the boundaries of those regions were often derived from geography and industrial history. Moreover, the finer the disaggregation of the data, the more precisely EU interventions and national economic policies can be studied, implemented and evaluated. Our sample

[1] When an amendment to the NUTS classification occurs, the concerned Member State replaces the time-series accordingly in Eurostat, which is the main source of the ARDECO database.

includes data of all 1348 regions at NUTS 3 level in the 28 EU countries over the period 2000-2017 (UK is included in the analysis as it had not left the EU yet).

The descriptive statistics in Table 1 display the asymmetries between the NUTS 3 EU regions in terms of GDP PPS per capita (GDPpc). The dispersion in this variable varies from a ratio of 1:89 for year 2000 to a ratio of 1:54 for year 2008. This fact, together with the rest of descriptive statistics proves the presence of outliers, as shown in Figure 1. The two extreme values correspond for all years to the same two regions, Candem and City of London, and Westminster both in UK, which have values of GDPpc much higher than the rest of the regions. However, these regions have not been removed from the analysis since model-based clustering is not significantly sensitive to outliers (Fraley and Raftery 2002).

4. Results and discussion

Figure 2 and 3 show the patterns and the mapping resulted from the model-based clustering over the GDPpc values of the NUTS 2 European regions in the period 2000-2017. The best model according to Bayesian Information Criterion (BIC) is a VEE model with 4 components or clusters (See Appendix Table A2 for a specification of the best solutions).

The first cluster (in magenta in Figure 2 and 3) is composed of regions in Spain, France, Italy, Greece, Portugal, Germany, UK, Finland, Belgium, Netherlands and Czech Republic. It has values close to the mean but lower.

The second cluster (in dark green in Figure 2 and 3) is constituted by some regions of UK, Italy, Czech Republic, Greece, Cyprus, Finland, Denmark, Sweden and Estonia, among others, and some capital regions as Lisbon in Portugal, Madrid in Spain or London in UK. It is characterized by a GDP per capita pattern superior to the mean.

The third one (in blue in Figure 2 and 3) has a value of GDP per capita much superior to the rest of the patterns and it is formed mainly by regions of Ireland.

Finally, cluster number four (in yellow in Figure 2 and 3) is mostly formed by regions of the eastern countries such as Hungary, Latvia, Lithuania, Poland, Romania, Bulgaria, Slovakia and Croatia.

In general, as shown in Figure 2, all the patterns are characterized by a rise in GDPpc values until year 2008, followed by a decrease in 2009, reflecting the effect of the global financial crisis of 2008, decrease that is sharper when the level of GDPpc is higher, and a rise again until 2017.

However, when we perform the model-based clustering with the same indicator at a finer level of disaggregation, the NUTS 3 level, we obtain that the number of clusters rises up to 11. This suggests that by performing the analysis at a more detailed spatial scale, more disparities flourish, since regions in a specific cluster share similar patterns of GDPpc but are dissimilar to regions not belonging to that cluster. To corroborate this statement we calculate two inequality measures, the coefficient of variation and the Gini coefficient, at NUTS 2 and NUTS 3 level for the period from 2000 to 2017 and compare them to verify whether there is more dispersion in the economic disparities between the two levels. Both measures, whose formulas appear below, are based on dispersion of GDP per capita.

$$CV = \frac{\frac{1}{N}\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2}}{\overline{y}} \quad (3)$$

$$Gini = \frac{1}{2N^2\overline{y}}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|y_i - y_j\right| \quad (4)$$

where y_i is the GDP per capita of the *i* region, $\overline{y}$ is the mean and *N* is the number of regions. We use the two inequality measures in its unweighted version according to Gluschenko (2018), who claims that the weighted version might lead to inconsistent outcomes.

As we see in Figures 4 and 5, the level of disparities of NUTS 3 regions is greater than the level for NUTS 2 regions in all years and for both measures. This is in line with studies at NUTS 3 level exploring the evolution of disparities in terms of the dispersion of GDP, such as Butkus et al. (2018), who claimed that regional disparities become more evident when considering smaller territorial units.

The patterns and mapping of the 11-cluster solution for NUTS 3 regions are represented in Figures 6 and 7. In general terms, some changes arise by clustering at NUTS 3 level with respect to the previous clustering on NUTS 2 regions. The cluster which comprised the eastern countries at NUTS 2 level (cluster number 4 in Figure 2 and 3), has been divided into two clusters (cluster number 9 and 10 in Figure 6 and 7), whose patterns of GDPpc are quite similar except for its behaviour during the financial crisis (growth rates are higher for cluster 10 during the crisis). Cluster number 1 in Figure 2 and 3 has been disaggregated in several clusters at the NUTS 3 level analysis. For example, most regions of Spain and some regions of Greece now form a separate cluster (cluster number 7 in Figure 6 and 7), which presents a slow recovery from the financial crisis. Regions of South Italy, Portugal, France and some regions in UK constitute another cluster which also have a slow recovery from the crisis and a lower level of GDPpc between years 2005-2009 (cluster 11 in dark blue in Figure 6 and 7). North Italy regions represent now a separated cluster as well (cluster 6 in grey in Figure 6 and 7), which has superior levels of GDPpc than the rest of the previous clusters (clusters 7 and 11 in Figure 6 and 7).

Additionally, we see that countries in the geographic centre of EU, such as Austria, Germany, Belgium, Netherlands or Czech Republic, are very heterogeneous since they contain regions belonging to several clusters. These NUTS 3 regions are quite small geographical areas, suggesting that the geographical size of the regions also plays a role in disparities. Cluster number 5 (in magenta in Figure 6 and 7) has economic values much higher than the rest of the clusters, since it contains the outliers mentioned in section 3.2.

Finally, it can be observed that the spatial effect is not as strong as it was for NUTS 2 clustering, in which regions belonging to the same cluster were usually spatially near (see Figure 3). This suggests that at a finer level of disaggregation, the local economic and market features arise and regions may share similar economic characteristics with remote regions rather than with contiguous ones. Although determining the forces behind the clustering is beyond the scope of this paper, we have performed an initial characterization of the 11 clusters obtained for the level of GDPpc at NUTS 3 level according to the variables used in Aumayr (2006) (in Table 2) who performed a multivariate clustering of NUTS 3 regions for year 2002 identifying 11 groupings[2]: agricultural rural peripheral

[2] Aumayr (2006) included an additional variable in the cluster analysis, the accessibility of population, which represents the amount of people weighted by a function of distance measured as the average travel time from one region to every other region and made a distinction between urban and non-urban regions in the clustering. We have used the variable “inhabitants” as a proxy for the potential size of the urban market (See Boix et al. 2013 and Lazzeretti et al. 2009).

regions, industrialized rural peripheral regions, highly intensive touristic regions, touristic regions, industrial regions, industrialized regional centres with surroundings, tertiary regional centres with surroundings, central agglomerations, industrial cities, tertiary cities and metropolitan areas.

As seen in table 3, we can characterise the regions belonging to cluster number 1 as the regions which can benefit more from the urbanization economies having the greatest potential for the urban market. Almost all regions are in central Europe with the exception of one region in the southern and six in the eastern periphery. Cluster number 4 regions, which present higher GDPpc levels than the rest, are the ones with the highest share in the tertiary sector. This cluster has the higher proportion of urban regions and it does not have totally rural regions[3]. Cluster number 5 regions, which have the highest levels of GDPpc, are the most industrialized and the ones which can benefit more from the agglomeration economies. All regions are metropolitan and are allocated in Central Europe. Cluster number 6 regions, situated in the southern periphery of EU, are the ones more specialized in tourism. Cluster number 7 regions are the least industrialized and have the lower potential for the urban market. All regions are in the southern periphery with the exception of one in the centre. Cluster number 8 regions, which have high values of GDPpc, are the least agricultural regions, not being entirely rural regions. All of them are in Central Europe except for one region in Greece. Cluster number 9 regions are the ones with less specialization in tourism, very few of them are urban, and all of them are in the eastern periphery except for four regions situated in the centre. Cluster number 10 regions, which

[3] The rural, urban and metropolitan classification of the regions has been taken from Eurostat NUTS 3 typologies (https://ec.europa.eu/eurostat/web/nuts/tercet-territorial-typologies).

have lowest levels of GDPpc, are the more agricultural and least specialized in the tertiary sector and agglomerated regions. Also it is the cluster with the lowest rate of urban regions. These regions are located in the eastern periphery (except for one in Denmark, another in France and two in Belgium). Finally, clusters number 2, 3 and 11 are characterized by regions with lower shares of the primary and secondary sectors, slightly higher shares in the tertiary sector and more specialized in tourism than the rest of the clusters.

This characterization shows that factors such as the structure (urban/rural), the sectorization, the agglomeration or the specialization of the regions may be considered as potential underlying reasons for explaining the economic inequalities that arise at the NUTS 3 level.

4.1. Economic growth analysis

An interesting step when analysing economic patterns is to study the (spatial) evolution of the economic growth (Postiglione, Cartone, and Panzera 2020). To this end, we will explore the patterns of growth rates of GDPpc for the whole 2000-2017[4] period performing a model-based clustering as well.

At NUTS 2 level, we obtain that the optimal number of clusters is 10 according to BIC. The patterns of this solution are represented in Figure 8 and mapping in Figure 9.

[4] Derived from the original 2000-2017 period, the growth period we analyse is from 2001 to 2017.

First, we notice in Figure 9 that regions are closely spatially clustered at national level. All regions of Spain (except Balearic Islands) represent an isolated cluster (cluster number 7 in Figure 8 and 9), nearly all regions in Italy constitute a separate cluster as well (cluster number 4 in Figure 8 and 9) and the same happens for Greece (cluster number 6 in Figure 8 and 9). The last one presents a pattern after the crisis different to the rest, since the growth rates are negative until 2013. Cluster number 9 is constituted by most regions of Poland (in purple in Figure 8 and 9) and is characterised by a high growth rate of GDPpc, although these regions belong to the cluster with the lowest level of GDPpc, as seen in Figure 2. Actually, mainstream economic growth theories claim that the lower the initial GDP per capita the higher the growth will be (Barro and Sala-i-Martin, 1992). Regions belonging to eastern countries, except Poland, together with Irish regions and Cyprus form cluster number 3 (in Figure 8 and 9). In this case it is worth highlighting the remoteness of Ireland and Cyprus with respect to the eastern countries, and even so, they share similar trends in growth rates of GDPpc. Concretely, Irish regions had high values of GDPpc (Figure 2 and 3) and present high growth rates of GDPpc whereas eastern countries had the lowest values of GDPpc. Regions of Austria and Belgium comprise cluster number 1 (in red in Figure 8 and 9) whereas almost all regions in Germany correspond to cluster number 5 (in magenta in Figure 8 and 9). Part of French regions belong to cluster number 8 (in black in Figure 8 and 9) while the rest, together with regions belonging to Denmark, Portugal, Netherlands, Czech Republic, Sweden, Finland and Slovenia, are included in cluster number 2 (in dark green in Figure 8 and 9). Finally, most regions of UK correspond to cluster 10 (in green in Figure 8 and 9), which is characterized by the highest fall of GDPpc in the 2008 crisis.

All these clusters present different growth rate patterns (Figure 8) but in general, all of them present a fall in the growth rates in 2003, an abrupt drop in 2009 coinciding with the financial crisis followed by a rise in growth rates in 2010. After some fluctuations, the growth rates reach a positive peak in 2015 and fall again in 2016.

By performing the model-based clustering for GDPpc growth rates at NUTS 3 level we obtain 15 clusters. Again, we see that at this level the number of clusters increases with respect to the NUTS 2 clustering. Although the spatial effect is not as strong as it is for NUTS 2 level, the regions also show a tendency to spatially cluster at the national level, except for some cases, such as Germany, whose regions belong to several clusters while at NUTS 2 level almost all regions belonged to the same cluster (cluster number 5 in Figure 8 and 9). This indicates that at finer level of disaggregation, the specialization of the regions becomes more visible and is not masked in the aggregate NUTS 2 level. In this case it is also worth noticing that the size of the NUTS 3 regions is quite small, enhancing this specialization effect. Performing the analysis at NUTS 3 level also reveals, for example, that regions of Latvia, Lithuania and Estonia belong to the same cluster on their own (cluster number 6 in Figure 10 and 11), whereas at NUTS 2 level they shared similar growth rates with Irish, eastern regions and Cyprus. Their pattern in the years before the crisis seems to be reversed with respect to other cluster patterns, as when they present a rise in growth, the rest present a decrease and vice versa. Most regions of Greece, Spain, Italy, UK and Poland are classified in individual clusters, similarly to NUTS 2 level (clusters 9, 14, 11, 7 and 13 in Figure 11, respectively). However, now the regions of Portugal and Romania form individual clusters (cluster number 15 and 12 in Figure 11, respectively). Regions of Austria, Belgium and some from Germany, belong to cluster number 2 (in dark green in Figure 10 and 11). Some eastern regions belong to the same cluster as Irish regions (number 10 in Figure 10 and 11) while the others, together with some regions of the Baltic countries, the north of UK and Cyprus, form a cluster (number 3 in Figure 10 and 11). Finally, it should be noticed that the patterns of growth of GDPpc at NUTS 3 level exhibit the same key features as at NUTS 2 level.

4.2. Pre- and post-2008 financial crisis analysis

Studies of the evolution of disparities highlight that after years of reduction in regional disparities, the financial crisis of 2008 led to an increase of inequalities (Butkus et al. 2018; Postiglione, Cartone, and Panzera 2020). Aiming to observe the effect of the financial crisis on the regional economic growth we split the growth period 2001-2017 into two time spans, one from 2001 to 2008 (pre-crisis period) and the other from 2009 to 2017 (post-crisis period).

Clustering the NUTS 2 regions for the pre-crisis period, gives us a grouping of 8 components. Similarly to the entire period, many regions are spatially clustered together at national level. Some regions are individually clustered at national level as it is the case of Spanish regions (cluster number 8 in Figure 12 and 13), French regions (cluster number 4 in Figure 12 and 13), Italian regions (cluster number 5 in Figure 12 and 13) and UK regions (cluster number 1 in Figure 12 and 13). Central EU countries, Portugal and part of the Baltic countries constitute cluster number 2 (in dark green in Figure 12 and 13). Some eastern countries, some regions of Greece and Ireland form cluster 7 (in dark red in Figure

12 and 13). Regions of Poland, Slovenia, Cyprus and most of Finland form cluster number 3 (in blue in Figure 12 and 13). Finally, Romania, Slovakia, Latvia, Lithuania and Estonia belong to cluster number 6 (in grey in Figure 12 and 13), which has the highest growth rates of GDPpc.

By performing the clustering at NUTS 3 level (Figures 14 and 15), we obtain the same number of components (eight) as for NUTS 2 clustering, indicating that in the years preceding the crisis, the overall disparities are not greater at NUTS 3 level than at NUTS 2. However, as shown in Figure 15, we see that the spatial relationship seems not to be as strong as it was for NUTS 2 level, i.e., the growth rates of GDPpc of NUTS 3 regions seem not to be so influenced by the values of the contiguous regions, but by regions that are not close in space, evidencing the presence of within-countries or within-NUTS 2 regions disparities. According to Butkus et al. (2018), disparities in the EU during 2000–2009 decreased mainly because of the reduction of disparities between countries, but the part of disparities that can be attributed to within-country and within-region disparities increased. Since the contrast of the spatial effect is so clear as we have obtained the same number of clusters at both NUTS 2 and NUTS 3 levels of analysis, we will test for spatial autocorrelation in order to validate what was perceived. To that end, we have calculated the Global Moran's I index, which indicates the presence of spatial autocorrelation. In order to be comparable at NUTS 2 and NUTS 3 level, we calculate the index in both cases at territorial NUTS 3 level. That is, in the case of the growth rates given at the NUTS 2 level, NUTS 3 regions get the value of the growth rates of GDPpc of the NUTS 2 region they belong to.

Global Moran's I index is formulated as:

$$I_t = \frac{n}{s_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} z_{i,t} w_{ij} z_{j,t}}{\sum_{i=1}^{n} {z_{i,t}}^2} \quad (5)$$

where $z_{i,t}$ is the deviation from the mean of the growth rate of GDPpc ($z_{i,t} = x_i - \overline{x}$ with $\overline{x}$ the mean of the variable x) in region *i* and period *t*. *n* is the number of NUTS 3 regions and w_{ij} is the *ij* element of the spatial weight matrix, which indicates the neighbours of each region. It is constructed by the K nearest neighbour's criterion, with k = 7. The outermost regions (Canary Islands, Guadeloupe, Martinique, Guyane, Mayotte, Reunion, Azores and Madeira) have been removed from the calculation since the K nearest neighbour's criterion is not suitable in these cases. $s_0 = \sum_i \sum_j w_{ij}$ is the sum of all weights. The inference is based on a random permutation approach with 999 permutations.

Table 4 displays the Moran's I index calculated for the annual growth rates of GDP per capita for each year of the pre-crisis period 2001-2008 for NUTS 3 and NUTS 2 regions. The values reveal a positive significant spatial autocorrelation for all years (p-values smaller than 0.05), i.e., regions with similar values of GDPpc growth rates tend to cluster together in space. Figure 16 displays the comparation between the evolution of the Index for growth rates at NUTS 2 and NUTS 3 level. We observe that the spatial effect is stronger for the NUTS 2 regions as we expected by visual inspection of Figures 13 and 15, indicating that NUTS 2 regions with similar GDPpc growth rates tend to be more spatially clustered than NUTS 3 regions with similar GDPpc growth rates. This evidences that at NUTS 2 level the behaviour of the NUTS 2 regions is more similar to that of its neighbours. However, at NUTS 3 level, the specialization becomes more visible and is not masked in the aggregate NUTS 2 level, revealing differences of the NUTS 3 regions with respect to their neighbours.

In the post-crisis period, at NUTS 2 level we obtain a grouping of 9 clusters, only one more than for the pre-crisis period. The differences with the pre-crisis period are basically that now regions of Italy and Finland belong to the same cluster (number 5 in magenta in Figure 17 and 18), German regions form now a separate cluster from the rest of the central countries (number 3 in blue in Figure 17 and 18), Greek regions form a separate cluster (number 8 in black in Figure 17 and 18), and the same happens for Polish regions (cluster number 9 in purple in Figure 17 and 18).

At NUTS 3 level, the clustering groups the regions in 15 clusters with similar tendencies of GDP per capita growth. This result indicates two things: that we obtain more clusters at NUTS 3 level than at NUTS 2, and that we obtain more clusters at NUTS 3 level in the post-crisis period than in the pre-crisis one. The NUTS 3 level seems to reveal the different behaviours and capacities of the regions when facing the financial crisis and the different recovery rates, behaviours that at NUTS 2 level are hidden. For example, Spain was a single cluster at NUTS 2 level, but when performing the cluster at NUTS 3 level, it has disaggregated in different clusters with different recovery patterns. Or all regions of Estonia, which belonged to a single cluster when the analysis was performed at NUTS 2 level, at NUTS 3 level one region has been revealed to belong to a different cluster with lower values of growth than the rest of the country regions.

5. Conclusions

This study has explored the regional economic disparities in the level of economic wealth and growth in the NUTS 3 regions in EU28 over the period from 2000 to 2017 by applying time-series clustering.

Our results suggest that by performing the analysis of both the level and growth rates of GDPpc at a more detailed spatial scale (NUTS 3 level) more disparities flourish. In general, NUTS 2 regions tend to spatially cluster at the national level and, although NUTS 3 regions show a slight tendency to cluster together in space at national level as well, the spatial effect is not as strong as it is for NUTS 2 level, indicating that NUTS 2 regions with similar economic characteristics tend to be more spatially clustered than NUTS 3 regions. This suggest that at a finer level of disaggregation, the local economic and market features arise and regions may share similar economic characteristics with remote regions rather than with contiguous regions. Hence it seems that a process of expanding stratification, or differentiation, in the GDPpc growth rates is occurring, implying that the current Cohesion Policy at NUTS 2 level might no longer be an effective antidote for the increasing disparities within the EU.

Aiming to observe the effect of the financial crisis on the regional economic growth we split the period 2000-2017 into two time spans, the pre-crisis and post-crisis periods. For the pre-crisis period, we have obtained the same number of components for the NUTS 3 and NUTS 2 clustering. However, again, we noticed that the spatial relationship at NUTS 3 level is not as strong as it was for NUTS 2 level. As for the post-crisis period, we obtained more clusters at NUTS 3 level than at NUTS 2, and more clusters at NUTS 3 level in the post-crisis period than in the pre-crisis one. The NUTS 3 level seems to reveal the different behaviours and capacities of the regions when facing the financial crisis and the different recovery rates that remain concealed at the NUTS 2 level.

Post-crisis policies were applied both at European and national level, and the clustering at NUTS 2 level seems to reflect the result, since most regions are clustered together in space at national level. However, when the clustering is performed at NUTS 3 level, we discovered that countries are very heterogeneous with regions belonging to several clusters which are not clustered together in space, revealing specific behaviours of the local economies and markets that are masked in the aggregate NUTS 2 level. Hence some less developed NUTS 3 regions might have been left behind in the general recovery with subdued GDP growth rates compared with those seen in the few years before the crisis.

All these findings evidence that at a more detailed spatial scale disparities flourish because of the local characteristics and specialization of the regions. Certainly, part of this effect arises because of the well-known Modifiable Areal Unit Problem (MAUP) that deals with the indeterminacy of any statistical measure as a response to changes in the level of aggregation of data (GDPpc and NUTS 2 and 3 level in this case) provoking a higher dispersion in the lower NUTS 3 level. However, this fact only reinforces our results that the NUTS 2 level is masking the heterogeneity of the EU regional economic disparities. The main consequence is that some relatively rich NUTS 3 regions may receive Structural Funds because their NUTS 2 region is eligible, while other less prosperous NUTS 3 regions may not receive Structural Funds because they are within an ineligible NUTS 2 region. Hence the Cohesion Policy, aimed at reducing regional disparities and defined at NUTS 2 level, should consider that NUTS 3 regions do not share the same behaviours as their aggregate NUTS 2 regions and therefore it may not have the expected effect in all the territories.

Although as evidenced, factors such as the structure (urban/rural), the sectorization, the agglomeration or the specialization of the regions may be considered as potential underlying reasons for explaining the economic inequalities that arise at the NUTS 3 level, the obtained groupings of regions which share the same economic characteristics throughout all the covered period can be used as a starting point in future empirical studies. These studies can focus on exploring deeper the driving forces behind the economic regional disparities and the unlinking of the economic wealth to geographic location, since the topic would be of great interest to scholars and policy makers alike.

Finally, we have to highlight that there are some limitations in our research that should be considered when studying the economic disparities using the GDP per capita as the economic indicator. At NUTS 3 level, some part of the GDP per capita may correspond to commuters and may cause some inaccuracies in this indicator when measuring the economic wealth and growth.

6. References

Aumayr, Christine (2006). “European Region Types: A Cluster Analysis of European NUTS 3 Regions.” 46th Congress of the European Regional Science Association: "Enlargement, Southern Europe and the Mediterranean", August 30th - September

3rd, 2006, Volos, Greece, European Regional Science Association (ERSA), Louvain-la-Neuve.

Aghabozorgi, Saeed, Ali Seyed Shirkhorshidi, and Teh Ying Wah. 2015. “Time-Series Clustering - A Decade Review.” *Information Systems* 53: 16-38. doi:10.1016/j.is.2015.04.007.

[dataset] ARDECO, 2020. European Regional Database: GDP PPS per head of population (1980-2017). https://knowledge4policy.ec.europa.eu/territorial/ardeco-online_en on February 2021.

Barro, Robert J., and Xavier Sala-i-Martin. 1992. “Convergence”. *Journal of Political Economy* 100(2): 223–251. doi:10.1086/261816

Biernacki, Christophe, Gilles Celeux, and Gérard Govaert. 2000. “Assessing a mixture model for clustering with the integrated completed likelihood”. *IEEE transactions on pattern analysis and machine intelligence* 22(7): 719-725. doi:10.1109/34.865189

Boix, Rafael, Blanca De-Miguel-Molina, and Jose-Luis Hervas-Oliver. 2013. “Creative service business and regional performance: Evidence for the European regions”. *Service Business* 7(3): 381–398. Doi:10.1007/s11628-012-0165-7

Butkus, Mindaugas, Diana Cibulskiene, Alma Maciulyte-Sniukiene, and Kristina Matuzeviciute. 2018. “What Is the Evolution of Convergence in the EU? Decomposing EU Disparities up to NUTS 3 Level.” *Sustainability (Switzerland)* 10 (5). doi:10.3390/su10051552.

Chu, Selina, Eamonn Keogh, David Hart, and Michael Pazzani. 2002. “Iterative deepening dynamic time warping for time series.” In *Proceedings of the Second SIAM international conference on data mining 195-212*. Society for Industrial and Applied Mathematics. doi: 10.1137/1.9781611972726.12

De Dominicis, Laura. 2014. “Inequality and Growth in European Regions: Towards a Place-Based Approach.” *Spatial Economic Analysis* 9 (2): 120-41. doi:10.1080/17421772.2014.891157.

del Campo, Cristina, Carlos Manuel Ferreira Monteiro, and Joao Oliveira Soares. 2008. “The European regional policy and the socio-economic diversity of European regions: A multivariate analysis”. *European Journal of Operational Research* 187: 600-612. doi:10.1016/j.ejor.2007.03.024.

Ertur, Cem, and Wilfried Koch. 2006. "Regional disparities in the European Union and the enlargement process: an exploratory spatial data analysis, 1995–2000." *The Annals of Regional Science* 40: 723-765. doi:10.1007/s00168-006-0062-x

European Commission. 2003. Second progress report on economic and social cohesion. Official Publication Office, Luxembourg.

European Commission. 2004. Third report on economic and social cohesion. Official Publication Office, Luxembourg.

Ezcurra, Roberto. 2019. “Regional disparities and within-country inequality in the European Union” *Revista de Economía Mundial* 51: 139-62. doi:10.33776/rem.v0i51.3907

Fraley, Chris, and Adrian E. Raftery. 1998. “How many clusters? Which clustering method? Answers via model-based cluster analysis.” *The Computer Journal* 41(8): 578–588. doi:10.1093/comjnl/41.8.578

Fraley, Chris, and Adrian E. Raftery. 2002. “Model-Based Clustering, Discriminant Analysis, and Density Estimation.” *Journal of the American Statistical Association* 97 (458): 611-31. doi:10.1198/016214502760047131.

Fratesi, Ugo, and Fiona G. Wishlade. 2017. “The Impact of European Cohesion Policy in Different Contexts” *Regional Studies* 51(6): 817-821. doi:10.1080/00343404.2017.1326673.

Gagliardi, Luisa, and Marco Percoco. 2017. “The Impact of European Cohesion Policy in Urban and Rural Regions.” *Regional Studies* 51 (6): 857-68. doi:10.1080/00343404.2016.1179384.

Gluschenko, Konstantin. 2018. “Measuring Regional Inequality: To Weight or Not to Weight?” *Spatial Economic Analysis* 13 (1): 36-59. doi:10.1080/17421772.2017.1343491.

Geppert, Kurt, and Andreas Stephan. 2008. “Regional Disparities in the European Union: Convergence and Agglomeration.” *Papers in Regional Science* 87 (2): 193-217. doi:10.1111/j.1435-5957.2007.00161.x.

Goecke, Henry, and Michael Hüther. 2016. “Regional Convergence in Europe.” *Intereconomics* 51 (3): 165-71. doi:10.1007/s10272-016-0595-x.

Grün, Bettina. 2019. Model-based clustering. CRC Press, Boca Raton 163-198

Heidenreich, Martin, and Christoph Wunder. 2008. “Patterns of Regional Inequality in the Enlarged Europe.” *European Sociological Review* 24 (1): 19-36. doi:10.1093/esr/jcm031.

Illner, Michal. 2010. “Top-Down or Bottom-Up? Coping with Territorial Fragmentation in the Czech Republic.” In *Balderstein, H. and L.E. Rose (eds) Territorial Choice 214-233.* Palgrave Macmillan, London UK. doi:10.1057/9780230289826_12

Kokocińska, Małgorzata, and Marcin Puziak. 2020. “Evolution of Regional Income Disparities in Large European Economies” *Revista de economía mundial* 54: 129-150. doi:10.33776/rem.v0i54.3826

Kutscherauer, Alois, Hana Fachinelli, Miroslav Hučka, Karel Skokan, Jan Sucháček, Petr Tománek, and Pavel Tuleja. 2010. “Regional disparities: Disparities in the regional development, their concept, identification and assessment.” *Ostrava: Technical University of Ostrava*.

Kyriacou, Andreas P., and Oriol Roca-Sagalés. 2012. “The Impact of EU Structural Funds on Regional Disparities within Member States.” *Environment and Planning C: Politics and Space* 30: 267-281. doi:10.1068/c11140r.

Lai, Cheng-Ping, Pau-Choo Chung, and Vincent S. Tseng. 2010. “A novel two-level clustering method for time series data analysis.” *Expert Systems with Applications* 37(9): 6319–6326. doi:10.1016/j.eswa.2010.02.089.

Lazzeretti, Luciana, Rafael Boix, and Francesco Capone. 2009. “Why do creative industries cluster? An analysis of the determinants of clustering of creative industries”. Summer Conference 2009, January, 1–34. Available at http://www2.druid.dk/conferences/viewpaper.php?id=5746&cf=32

Le Gallo, Julie, and Cem Ertur. 2005. “Exploratory Spatial Data Analysis of the Distribution of Regional per Capita GDP in Europe, 1980-1995.” *Papers in Regional Science* 82 (2): 175–201. doi:10.1111/j.1435-5597.2003.tb00010.x.

Le Gallo, Julie. 2004. “Space-Time Analysis of GDP Disparities among European Regions: A Markov Chains Approach.” *International Regional Science Review* 27 (2): 138–63. doi:10.1177/0160017603262402.

Lessmann, Christian. 2014. “Spatial Inequality and Development – Is there an inverted U Relationship?” *Journal of Development Economics* 106: 35-51. doi:10.1016/j.jdeveco.2013.08.011

Leška, Dušan. 2015. “Formation of regional level of self-government in the process of transformation in Slovakia.” *Journal for universal excellence* 4(3): A1-A11. Available at https://www.fos-unm.si/media/pdf/ruo/2015%203/8.pdf

Lin, Jessica, Michail Vlachos, Eamonn Keogh, and Dimitrios Gunopulos. 2004. “Iterative incremental clustering of time series.” In *International Conference on Extending Database Technology 106-122.* Springer, Berlin, Heidelberg.

Meliciani, Valentina. 2015. *Regional Disparities in the Enlarged European Union: Geography, innovation and structural change*. London, UK: Routledge.

Monfort, Philippe. 2008. “Convergence of EU Regions - Measures and Evolution.” Directorate-General for Regional Policy Working Paper, 1, 1-20. Available at https://ec.europa.eu/regional_policy/sources/docgener/work/200801_convergence.pdf.

Mora, Toni, and Esther Vaya. 2004. “The Enlargement of the European Union and the Spatial Distribution of Economic Activity” *Eastern European Economics* 42 (5): 6-35. doi:10.1080/00128775.2004.11041086

Panzera, Domenica, and Paolo Postiglione. 2021. “The Impact of Regional Inequality on Economic Growth: A Spatial Econometric Approach.” *Regional Studies*, forthcoming. doi:10.1080/00343404.2021.1910228.

Pavone, Pasquale, Francesco Pagliacci, Margherita Russo, and Simone Righi. 2021. “Multidimensional Clustering of EU Regions: A Contribution to Orient Public Policies in Reducing Regional Disparities.” *Social Indicators Research* 156 (2): 739-59. doi:10.1007/s11205-020-02324-9.

Petrakos, George, Andrés Rodríguez-Pose, and Antonis Rovolis. 2003. “Growth, Integration and Regional Inequality in Europe.” *Environment and Planning A*, 37: 1837–1855. doi:10.1068/a37348

Poledníková, Eva. 2014. “Multicriteria analysis of regional disparities in the context of the EU cohesion.” in: XVII. mezinárodní kolokvium o regionálních vědách, Hustopeče, 18-20.6.2014, Sborník příspěvků, 54-61. doi:10.5817/CZ.MUNI.P210-6840-2014-5

Postiglione, Paolo, Alfredo Cartone, and Domenica Panzera. 2020. “Economic Convergence in EU NUTS 3 Regions: A Spatial Econometric Perspective.” *Sustainability* 12(17): 6717. doi:10.3390/su12176717

Puga, Diego. 2002. “European regional policies in light of recent location theories.” *Journal of Economic Geography* 2: 373-406. doi:10.1093/jeg/2.4.373

Quah, Danny. 1993a. “Empirical cross-section dynamics in economic growth”, *European Economic Review* 37(2-3): 426-434. doi:10.1016/0014-2921(93)90031-5

Quah, Danny. 1993b. “Galton's fallacy and tests of the convergence hypothesis” *Scandinavian Journal of Economics* 95: 427-443. doi:10.2307/3440905

Scrucca, Luca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. 2016. “Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.” *R Journal* 8 (1): 289-317. doi:10.32614/rj-2016-021.

Smetkowski, Maciej, and Piotr Wójcik. 2012. “Regional Convergence in Central and Eastern European Countries: A Multidimensional Approach.” *European Planning Studies* 20: 923-939. doi:10.1080/09654313.2012.673560

Soares, Joao Oliveira, Maria Manuela Lourenço Marquês, and Carlos Manuel Ferreira Monteiro. 2003. “A Multivariate Methodology to Uncover Regional Disparities: A Contribution to Improve European Union and Governmental Decisions” *European Journal of Operational Research* 145: 121-135. doi:10.1016/S0377-2217(02)00146-7

Tvrdoň, Michal, and Karel Skokan. 2011. “Regional disparities and the ways of their measurement: the case of the visegrad four countries” *Technological and Economic Development of Economy* 17 (3): 501-518. doi:10.3846/20294913.2011.603485. doi:10.3846/20294913.2011.603485

Tvrdoň, Michal. 2012. “Cohesion Policy, Convergence and Regional Disparities: The Case of the European Union.” *WSEAS Transactions on Business and Economics* 9 (2): 89-99. Available at: https://wseas.org/multimedia/journals/economics/2012/53-827.pdf

Warren Liao, T. 2005. “Clustering of Time Series Data - A Survey.” *Pattern Recognition* 38 (11): 1857-74. doi:10.1016/j.patcog.2005.01.025.

Wójcik, Piotr. 2016. “Parallel convergence of income and educational achievements on a local level in Poland - Joint distribution dynamics”, *56th Congress of the European Regional Science Association: "Cities & Regions: Smart, Sustainable, Inclusive?"*, 23-26 August 2016, Vienna, Austria, European Regional Science Association (ERSA), Louvain-la-Neuve. Available at: http://hdl.handle.net/10419/174676

7. Tables and Figures

Table 1: GDP PPS per capita descriptive statistics by year (2000-2017)

Year	Min	Max	Mean	SD	Skewness	Kurtosis
2000	2424.90	216065.64	19006.23	11363.35	6.70	98.03
2001	2941.08	219482.70	19689.56	11662.01	6.61	94.82
2002	3118.54	227189.30	20272.89	11888.53	6.82	99.76
2003	3196.35	242891.94	20565.48	12135.57	7.47	115.72
2004	3791.34	263918.02	21532.59	12749.72	7.94	127.82
2005	3654.25	284668.66	22209.95	13327.37	8.46	141.12
2006	4515.43	299139.35	23329.51	13858.58	8.48	142.78
2007	4964.68	329885.39	24528.71	14849.87	9.15	160.06
2008	5895.72	317621.44	24477.84	14556.34	9.11	155.84
2009	5541.05	313248.98	22955.96	13917.09	9.84	175.86
2010	5581.77	329119.21	23985.94	14793.20	9.64	169.25
2011	5930.31	336346.31	24794.87	15425.39	9.23	156.25
2012	5902.01	343185.39	25291.76	15799.21	9.34	157.88
2013	6048.57	346897.44	25446.90	16004.78	9.46	159.98
2014	6525.19	370808.40	26388.09	16830.60	9.92	172.63
2015	6609.63	382945.15	27523.14	17445.72	9.74	167.94
2016	6587.15	402957.26	27642.40	18013.51	10.20	178.87
2017	6995.98	400425.27	28299.72	18174.35	9.97	171.06

Table 2: Variables for the characterization of the 11 clusters for the GDPpc level of NUTS 3 regions

Variable	Proxy	Source	Unit(**)
Inhabitants	This variable approaches the market potential	Eurostat	Total number of people
Population density	This variable serves as an indicator of agglomeration economies	Eurostat	Number of people divided by the area
Share of employment in the primary sector	This variable accounts for agrarian regions	ARDECO(*)	Percentage of the employment in the primary sector over the total employment
Share of GVA in the secondary sector	This variable determines industrialized regions	ARDECO(*)	Percentage of the GVA in the secondary sector over the total GVA
Share of employment in the tertiary sector	This variable determines the tertiarization of a region	ARDECO(*)	Percentage of the employment in the tertiary sector over the total employment
Beds in hotels per 1000 habitants	This variable indicates tourism specialization	Eurostat	Number of beds in hotels per thousand inhabitants

Note: (*) data of UK has been extracted from the Office for National Statistics of UK
(**) The indicators are computed as the mean of the variables over the 2000-2017 period

Table 3: Characterization of the 11 clusters for the GDPpc level of NUTS 3 regions

Indicators	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**	**Cluster 10**	**Cluster 11**
Inhabitants	**192738.91**	296458.14	337501.42	407538.64	291625.27	517115.71	**803723.06**	367604.07	402383.75	364796.35	440250.50
Population density	204.38	442.62	712.13	1534.96	**3613.88**	284.92	458.84	2208.11	103.45	**74.64**	472.30
Share employment primary sector	4.32	4.31	5.60	3.69	2.05	3.71	9.25	**1.77**	21.21	**22.81**	7.29
Share GVA secondary sector	23.30	23.48	22.08	22.02	**39.48**	24.03	**16.04**	31.60	23.95	28.38	16.65
Share employment tertiary sector	67.37	68.56	70.09	**74.98**	74.33	66.05	68.93	71.68	49.83	**45.60**	70.25
Beds in hotels per 1000 habitants	55.98	68.61	78.75	62.29	32.80	**109.21**	68.72	38.76	**23.48**	43.48	76.43

Note: Grey cells identify the indicators for which the Kruskal-Wallis test rejects the null hypothesis of similar sample means. Bold black figures identify the highest values of the means and red figures identify the lowest ones. The Kruskal-Wallis test is a non-parametric method for equality of means. It indicates the significant differences among means and it can be used to determine the between-group differences among clusters.

Table 4. Moran's I index for the annual growth rate of GDPpc and pre-crisis period. E(I) is the theoretical expected value, which equals -1/(n-1).

	Moran's I		E(I)		Standard Deviation		p-value	
Year	NUTS 3	NUTS 2	NUTS 3	NUTS 2	NUTS 3	NUTS 2	NUTS 3	NUTS 2
2001	0.2826	0.7170	-0.0008	-0.0008	0.0136	0.0141	0.001	0.001
2002	0.2711	0.6925	-0.0008	-0.0008	0.0133	0.0141	0.001	0.001
2003	0.3409	0.7310	-0.0008	-0.0008	0.0141	0.0135	0.001	0.001
2004	0.4332	0.7893	-0.0008	-0.0008	0.0137	0.0137	0.001	0.001
2005	0.2489	0.7246	-0.0008	-0.0008	0.0136	0.0139	0.001	0.001
2006	0.3331	0.6938	-0.0008	-0.0008	0.0139	0.0010	0.001	0.001
2007	0.5160	0.8122	-0.0008	-0.0008	0.0135	0.0137	0.001	0.001
2008	0.4869	0.8177	-0.0008	-0.0008	0.0139	0.0137	0.001	0.001

8. Figures

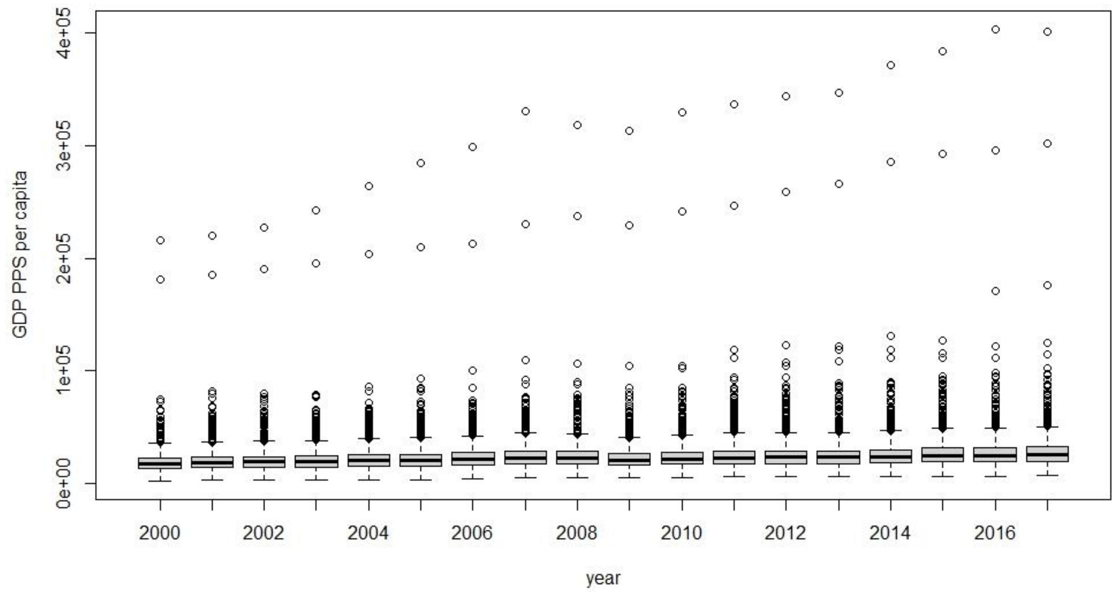

Figure 1: Boxplots of GDP per capita by year (2000-2017).

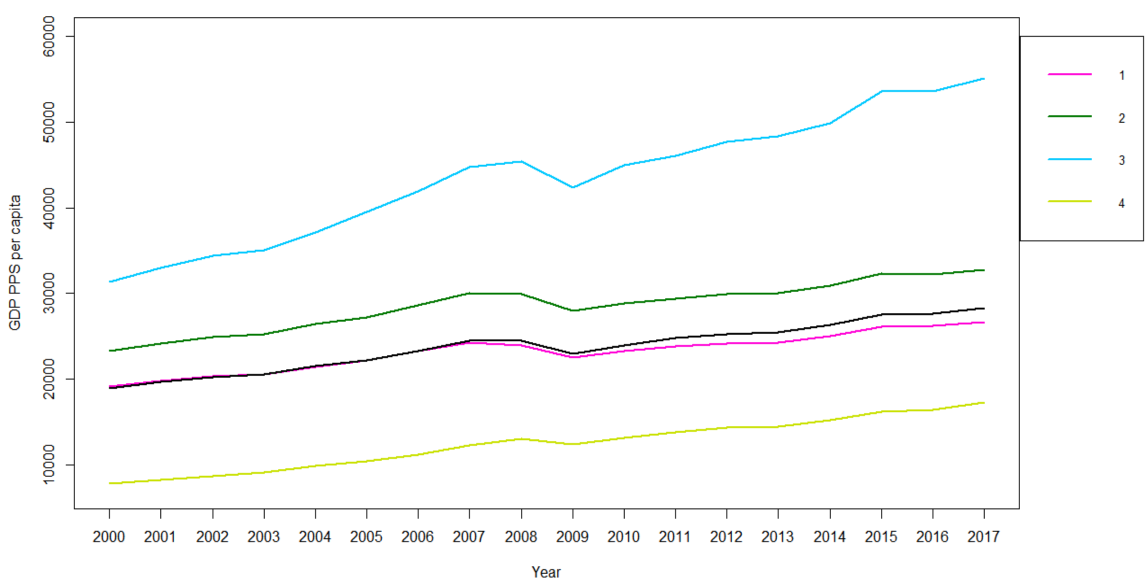

Figure 2. Patterns of 4-cluster solution for GDPpc of NUTS 2 regions for period 2000-2017. The mean values for each year are represented with a black line.

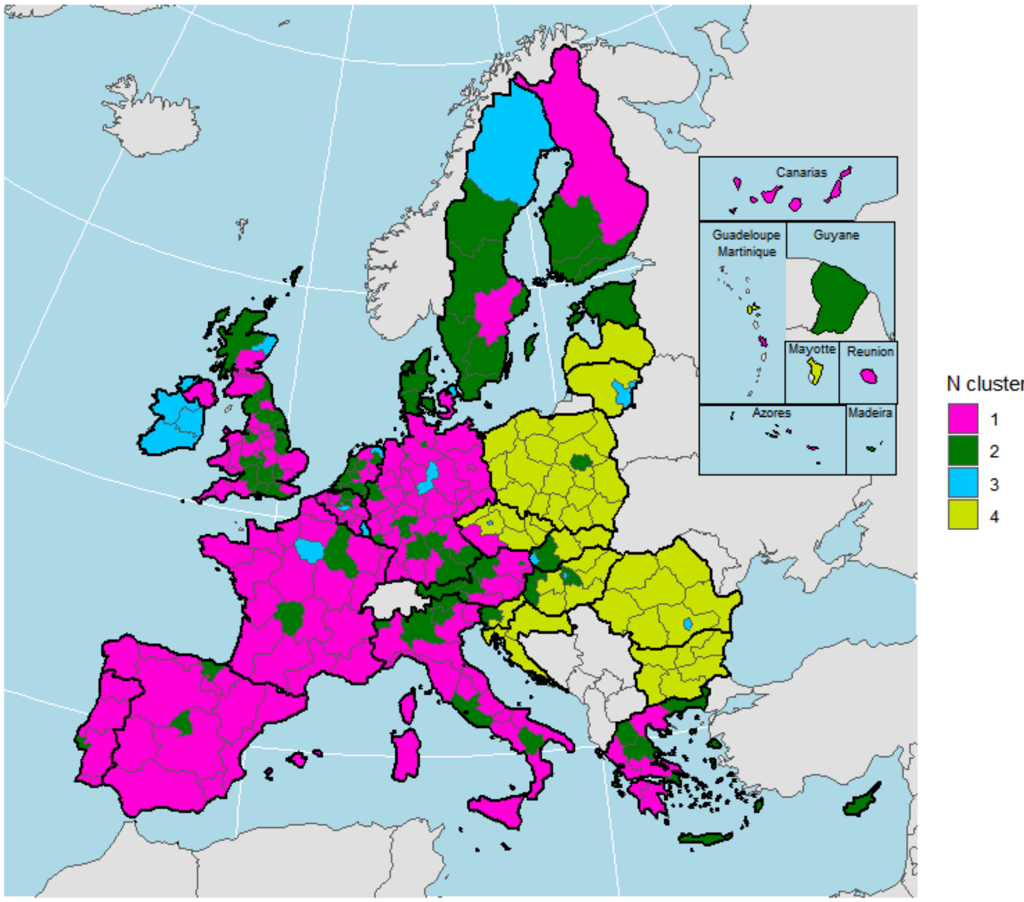

Figure 3. Mapping of 4-cluster solution for GDPpc of NUTS 2 regions for period 2000-2017.

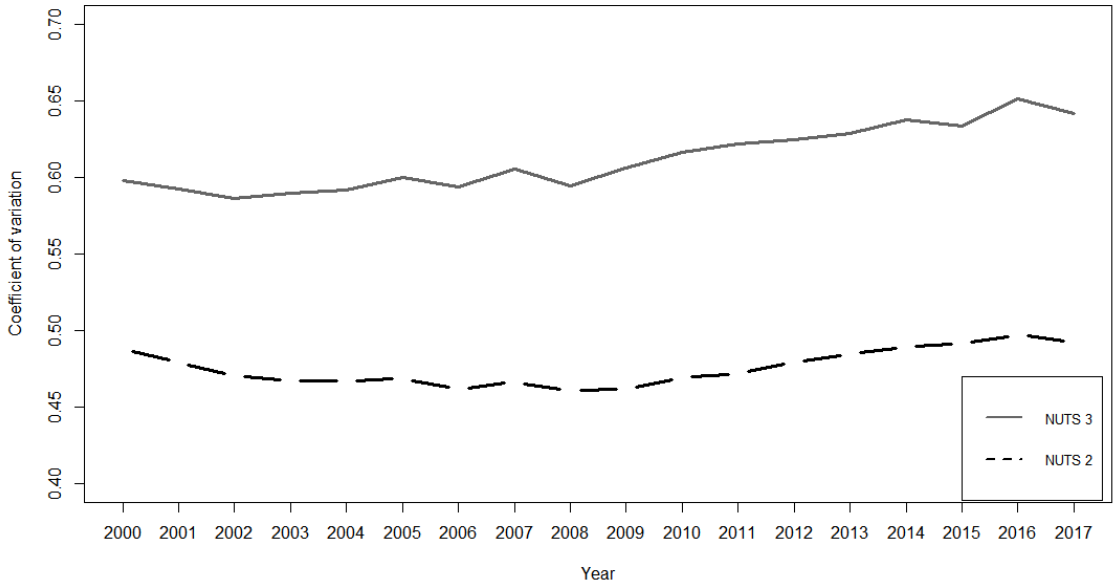

Figure 4. Coefficient of variation of GDPpc of NUTS 2 and NUTS 3 regions for period 2000-2017.

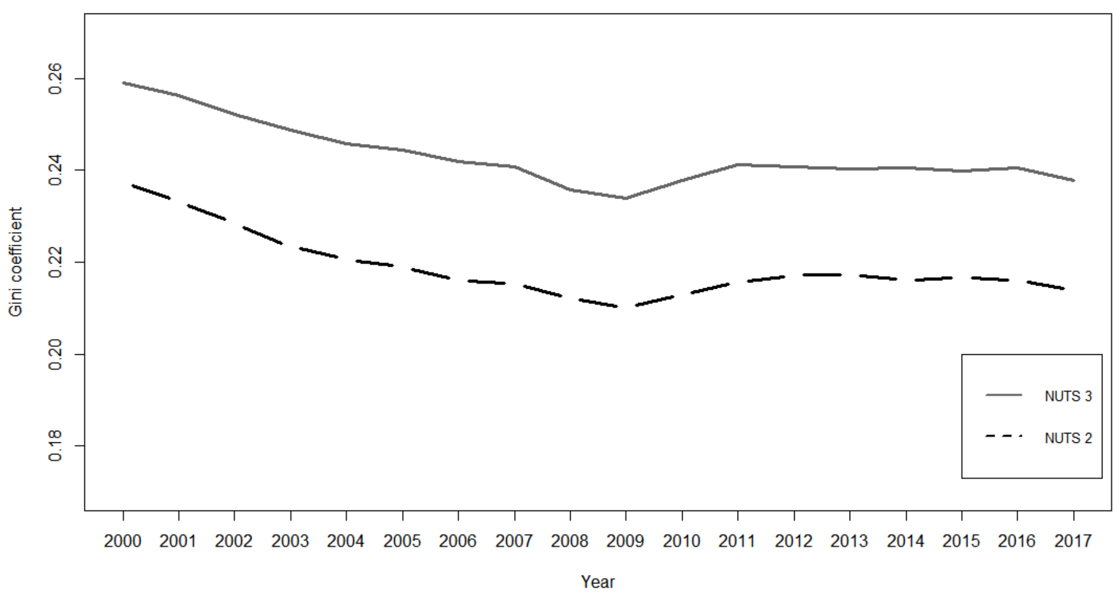

Figure 5. Gini coefficient of GDPpc of NUTS 2 and NUTS 3 regions for period 2000-2017.

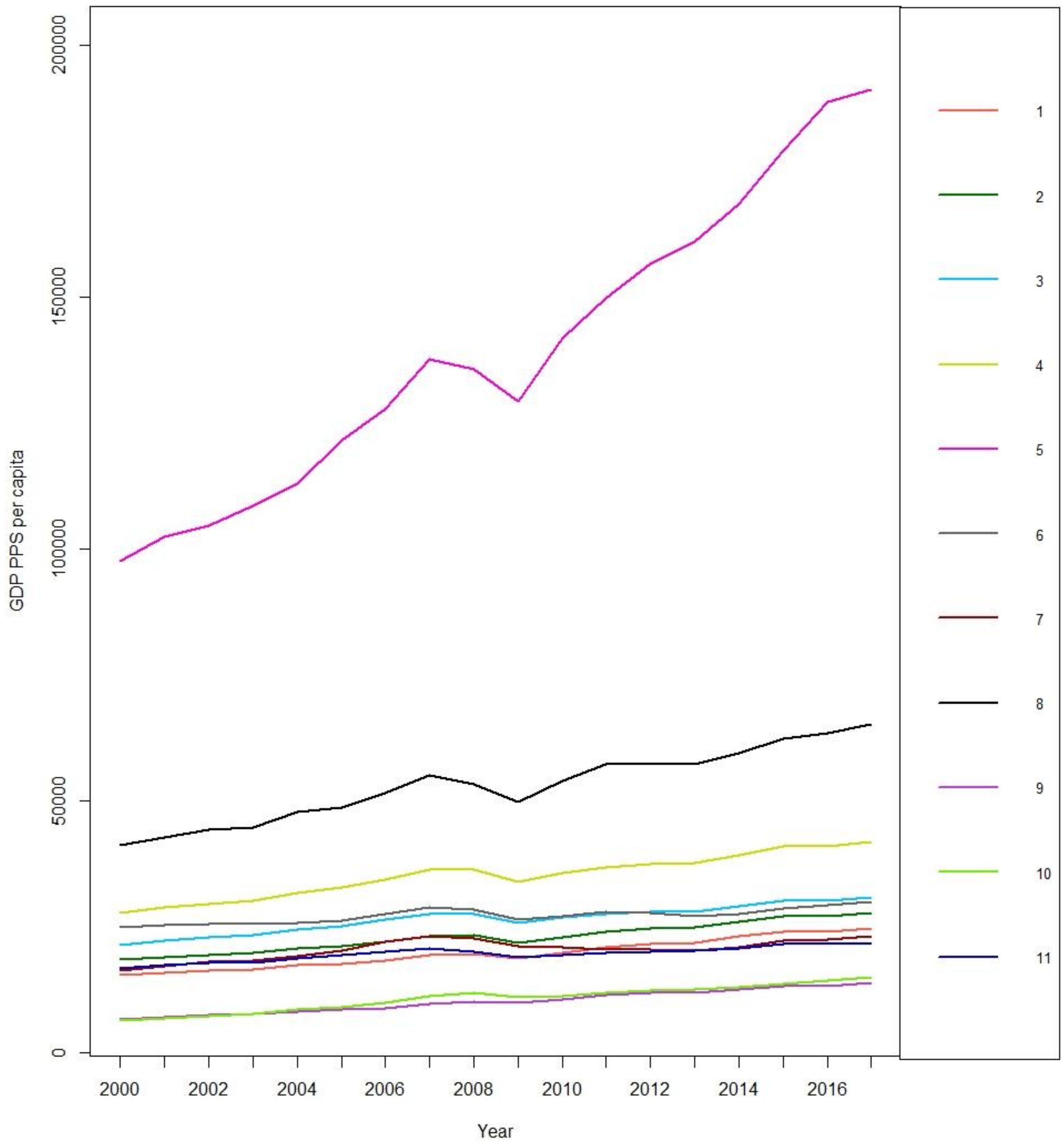

Figure 6. Patterns of 11-cluster solution for GDPpc of NUTS 3 regions for period 2000-2017.

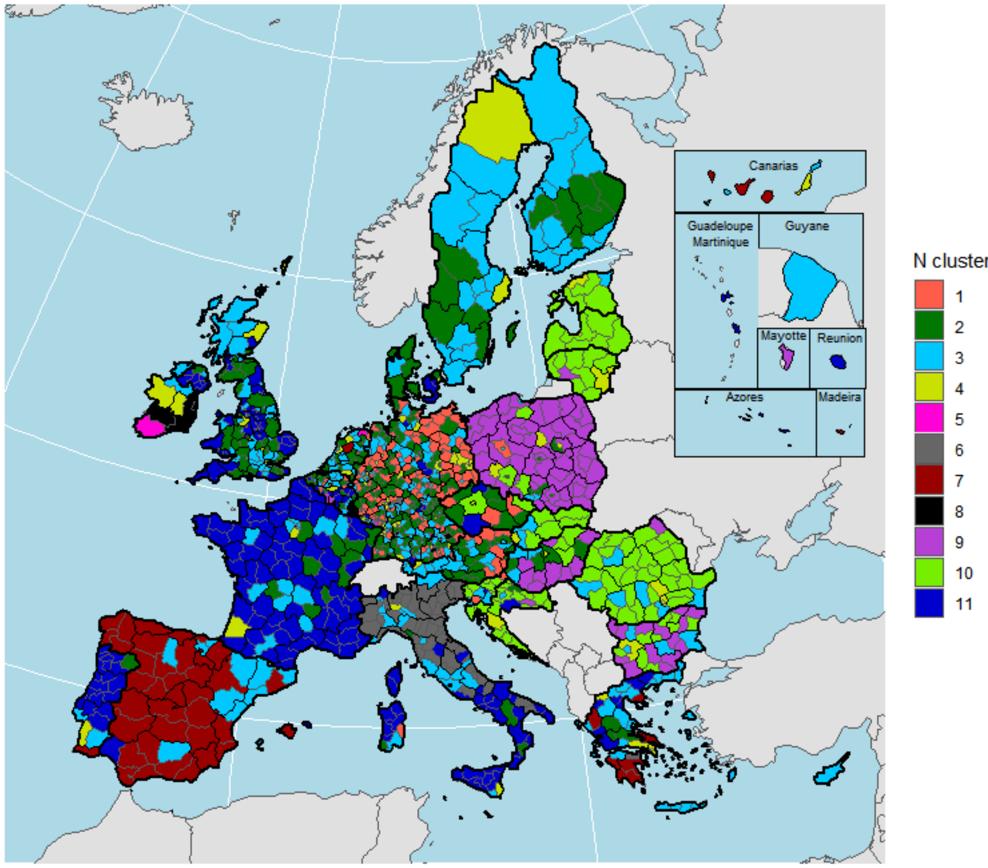

Figure 7. Mapping of 11-cluster solution for GDPpc of NUTS 3 regions and period 2000-2017.

Figure 8. Patterns of 10-cluster solution for average GDP pc growth rates of NUTS 2 regions for period 2001-2017.

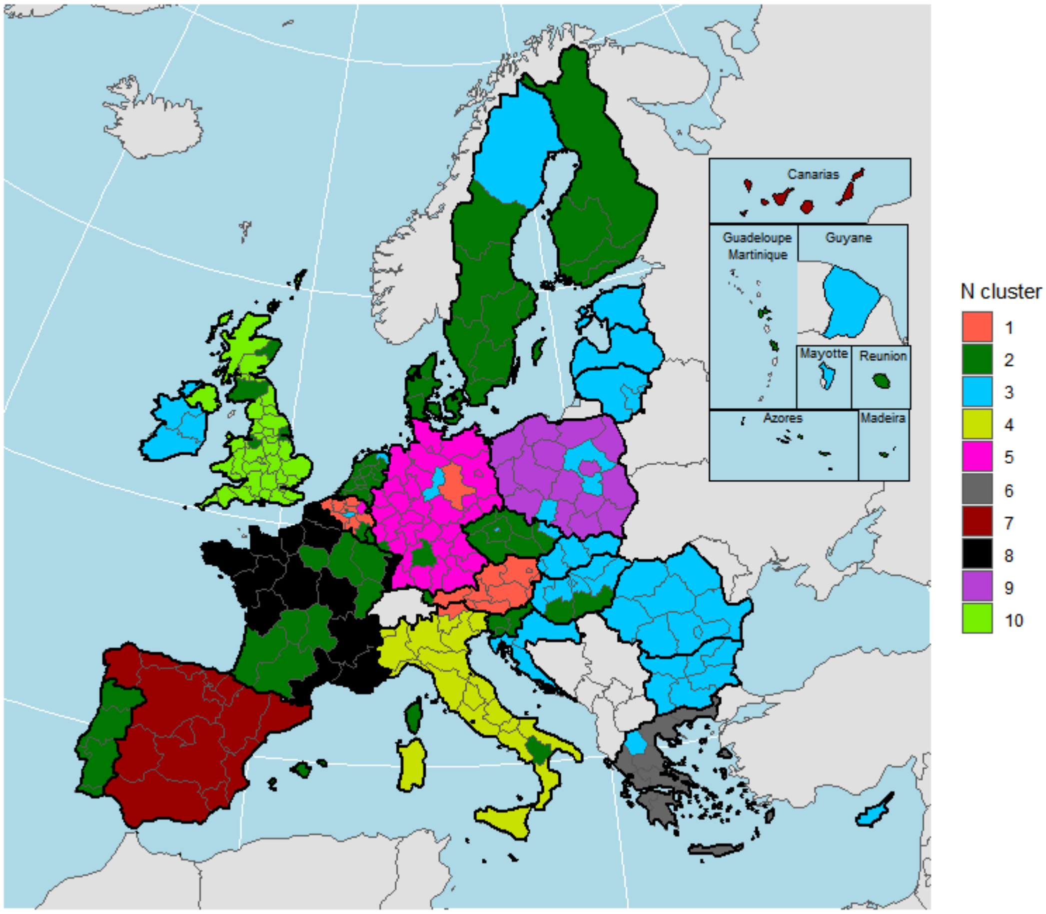

Figure 9. Mapping of 10-cluster solution for GDPpc growth rates of NUTS 2 regions for period 2001-2017.

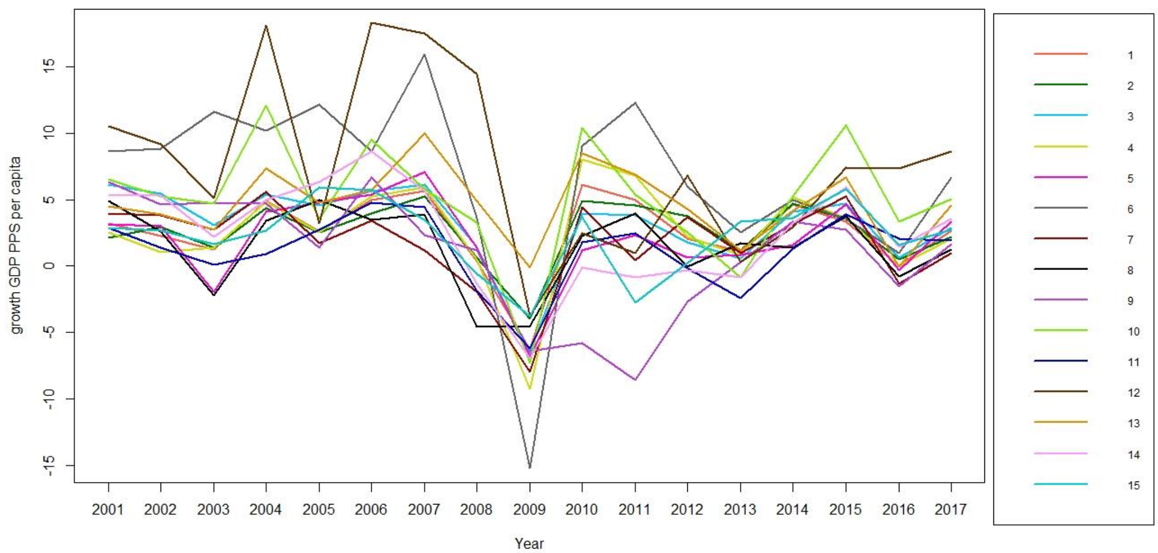

Figure 10. Patterns of 15-cluster solution for average GDP pc growth rates of NUTS 3 regions for period 2001-2017.

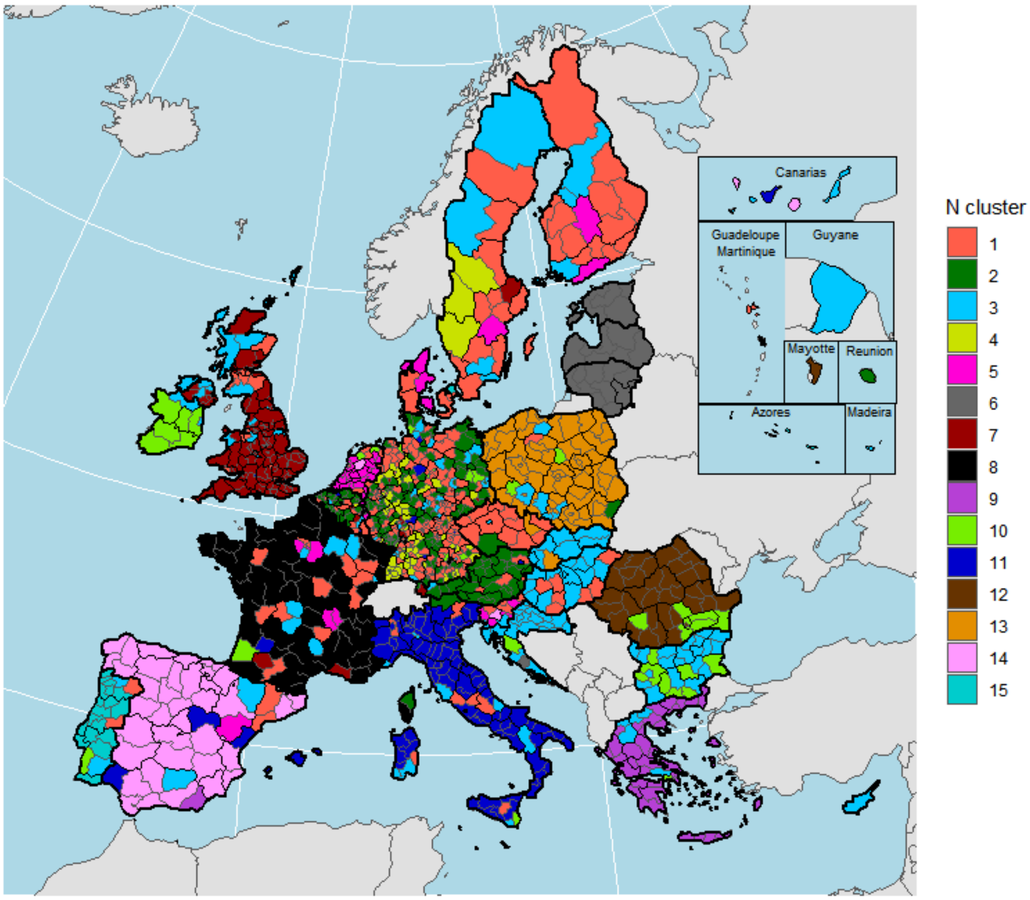

Figure 11. Mapping of 15-cluster solution for GDPpc growth rates of NUTS 3 regions for period 2001-2017.

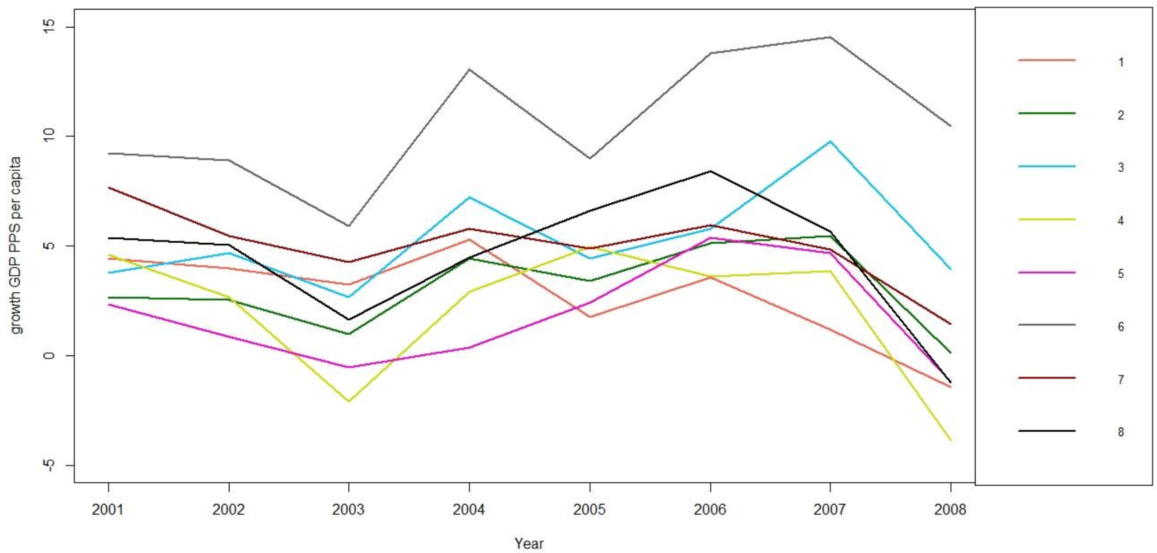

Figure 12. Patterns of 8-cluster solution for average GDP pc growth rates of NUTS 2 regions and pre-crisis period.

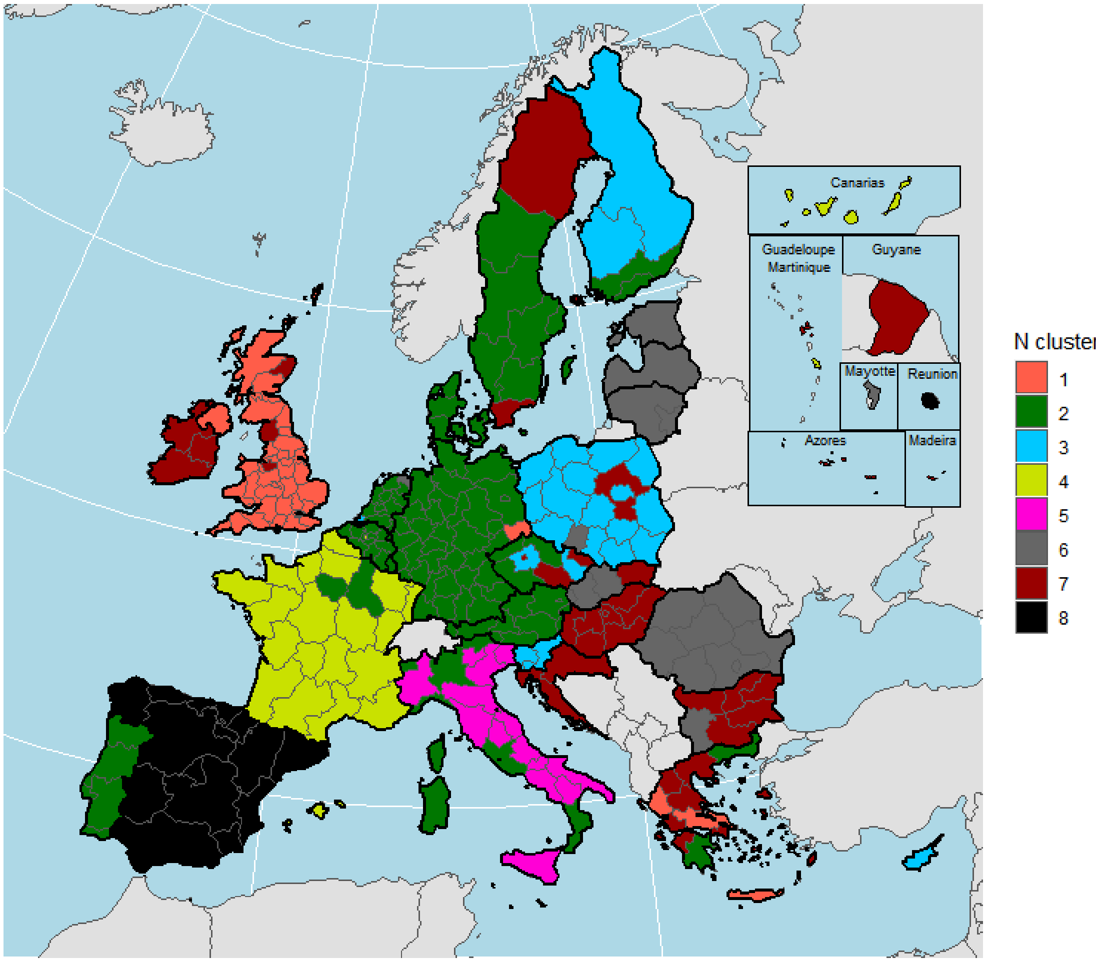

Figure 13. Mapping of 8-cluster solution for GDPpc growth rates of NUTS 2 regions and pre-crisis period.

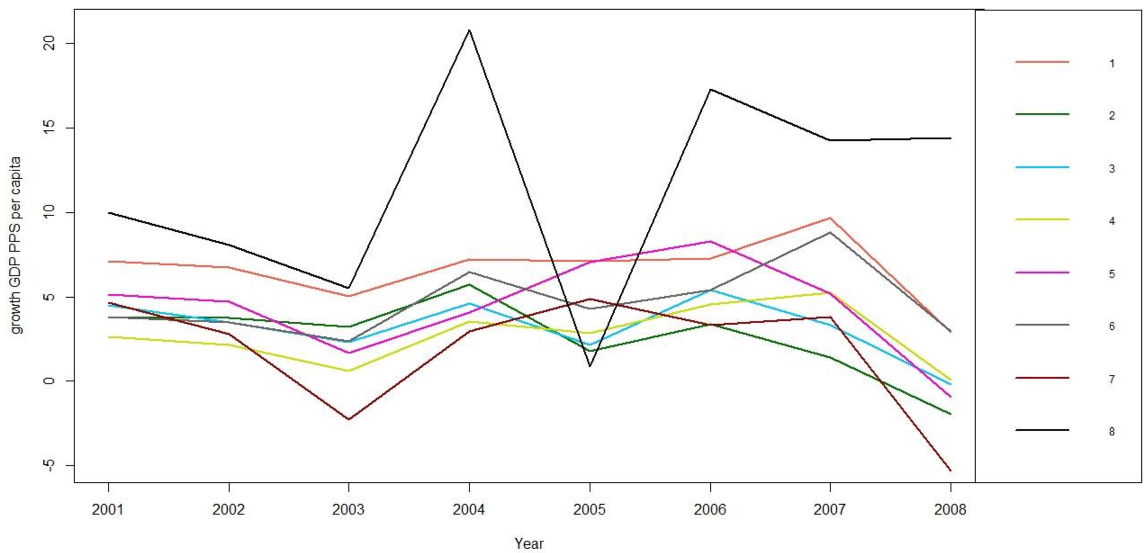

Figure 14. Patterns of 8-cluster solution for average GDP pc growth rates of NUTS 3 regions and pre-crisis period.

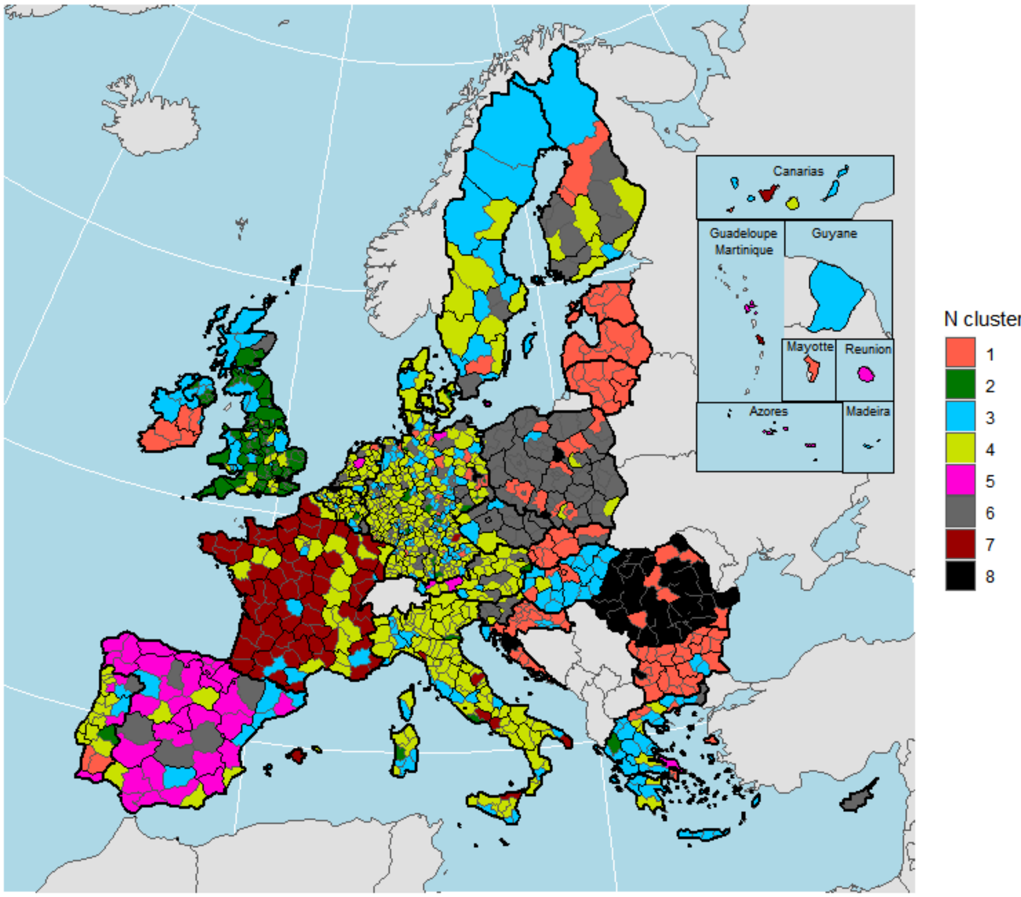

Figure 15. Mapping of 8-cluster solution for GDPpc growth rates of NUTS 3 regions and pre-crisis period.

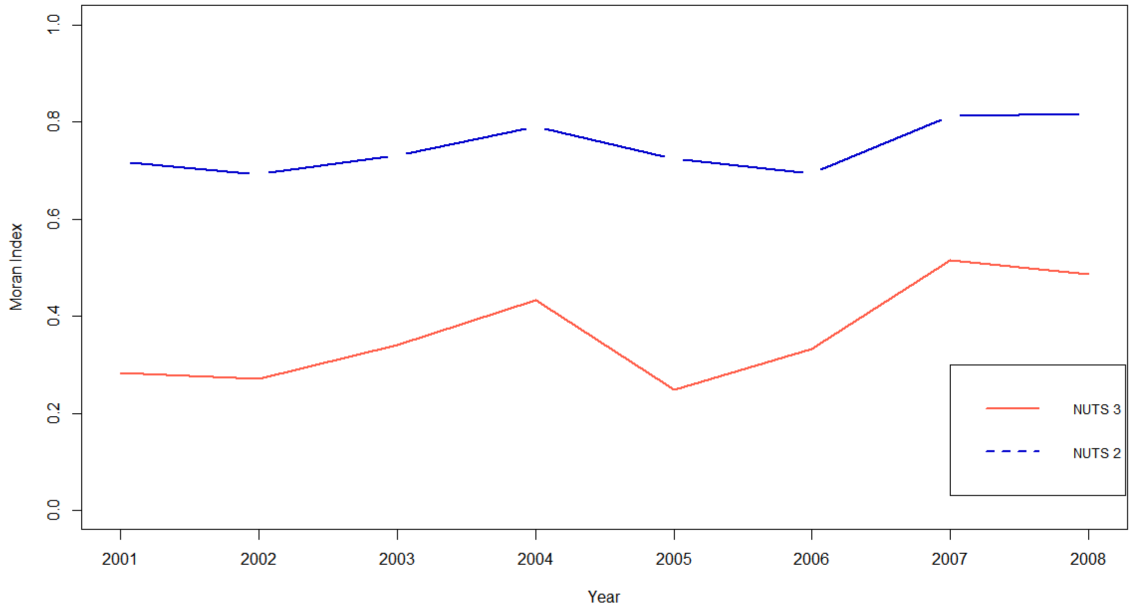

Figure 16. Moran's I index for GDPpc growth rates of NUTS 2 and NUTS 3 regions and pre-crisis period.

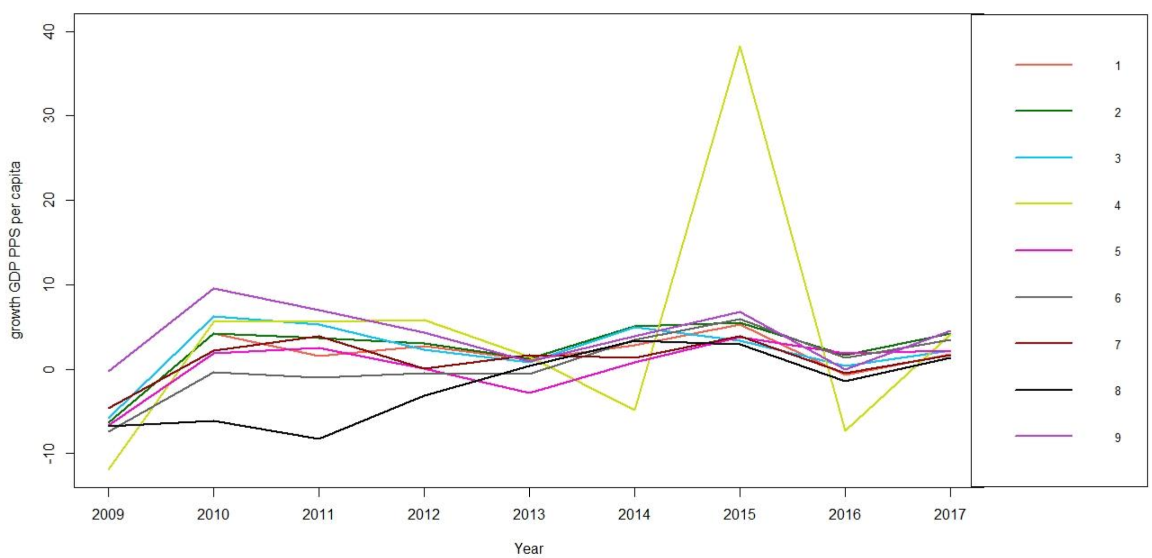

Figure 17. Patterns of 9-cluster solution for average GDP pc growth rates of NUTS 2 regions and post-crisis period.

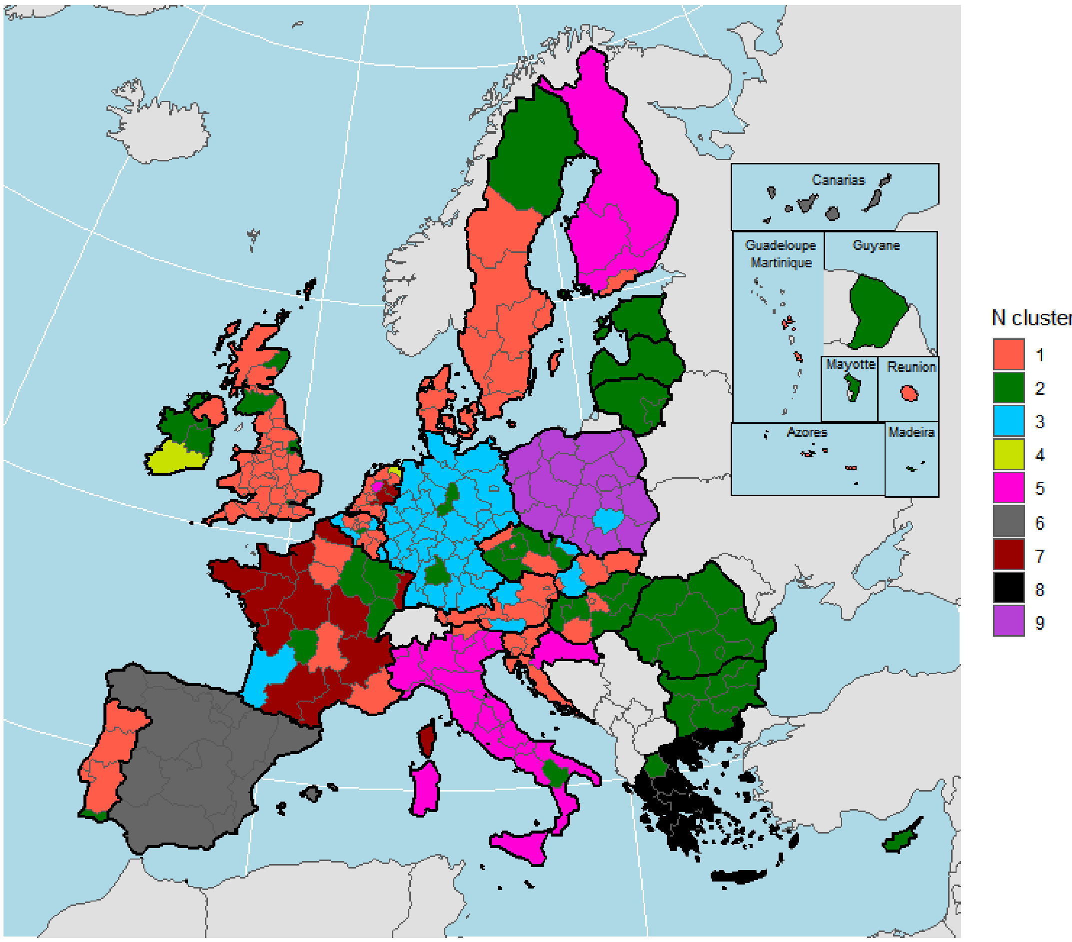

Figure 18. Mapping of 9-cluster solution for GDPpc growth rates of NUTS 2 regions and post-crisis period.

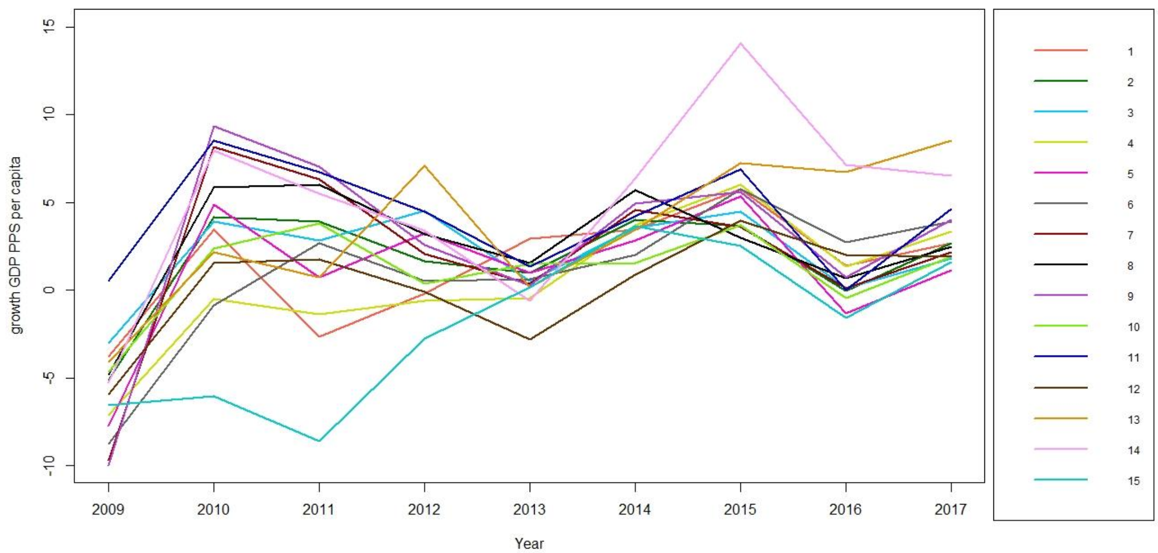

Figure 19. Patterns of 15-cluster solution for average GDP pc growth rates of NUTS 3 regions and post-crisis period.

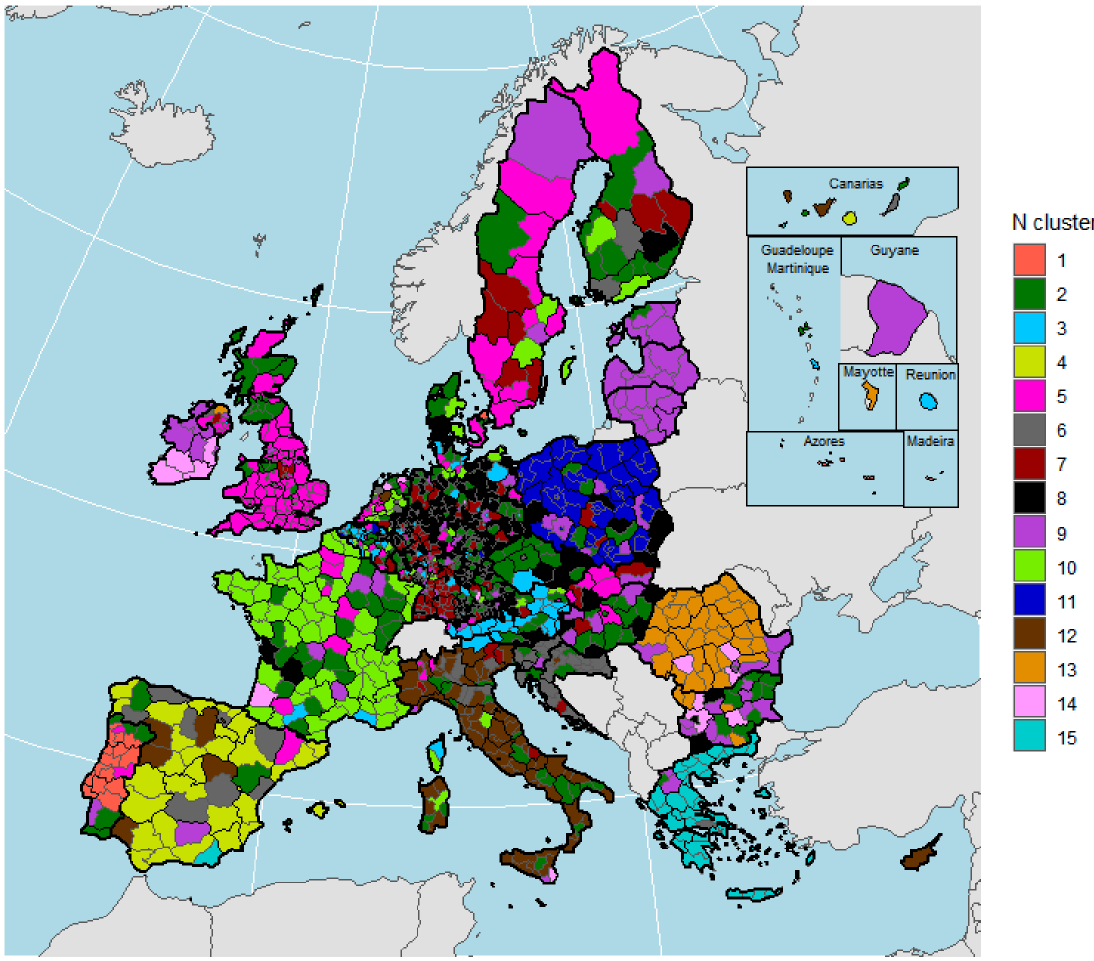

Figure 20. Mapping of 15-cluster solution for GDPpc growth rates NUTS 3 regions and post-crisis period.

Appendix A

In order to select the number of clusters, the BIC is a popular choice in the context of GMMs. In the *mclust* package, BIC is used by default for model selection and allows to obtain a matrix of BIC values for all the available models and number of components in order to determine the optimal number of clusters.

Table A1: Possible models in clustering

Model	Distribution	Volume	Shape	Orientation
EII	Spherical	Equal	Equal	-
VII	Spherical	Variable	Equal	-
EEI	Diagonal	Equal	Equal	Coordinate axes
VEI	Diagonal	Variable	Equal	Coordinate axes
EVI	Diagonal	Equal	Variable	Coordinate axes
VVI	Diagonal	Variable	Variable	Coordinate axes
EEE	Ellipsoidal	Equal	Equal	Equal
EVE	Ellipsoidal	Equal	Variable	Equal
VEE	Ellipsoidal	Variable	Equal	Equal
VVE	Ellipsoidal	Variable	Variable	Equal
EEV	Ellipsoidal	Equal	Equal	Variable
VEV	Ellipsoidal	Variable	Equal	Variable
EVV	Ellipsoidal	Equal	Variable	Variable
VVV	Ellipsoidal	Variable	Variable	Variable

Table A2: Best BIC values for each clustering solution

Clustering		1st best BIC value	2nd best BIC value	3rd best BIC value
Level of GDP of NUTS 2 regions in period 2000-2017	BIC	-80259.07	-80366.15	-80546.90
	Model	VEE	VEE	VEE
	Clusters	4	3	2
Level of GDP of NUTS 3 regions in period 2000-2017	BIC	-392951.3	-392967.93	-393227.54
	Model	VEE	VEE	VEE
	Clusters	11	10	9
Growth rates of GDP of NUTS 2 regions in period 2001-2017	BIC	-22138.4	-22155.93	-22256.11
	Model	VEI	VEI	VEI
	Clusters	10	9	8
Growth rates of GDP of NUTS 3 regions in period 2001-2017	BIC	-119064.7	-119108.74	-119132.55
	Model	VEI	VEI	VEI
	Clusters	15	14	16
Growth rates of GDP of NUTS 2 regions in pre-crisis period	BIC	-10762.27	-10779.74	-10790.81
	Model	VII	VII	VEI
	Clusters	8	9	8
Growth rates of GDP of NUTS 3 regions in pre-crisis period	BIC	-58517.89	-58534.82	-58547.77
	Model	VEI	VII	VEI
	Clusters	8	8	7
Growth rates of GDP of NUTS 2 regions in post-crisis period	BIC	-11789.31	-11814.95	-11830.12
	Model	VEI	VEI	VEI
	Clusters	9	10	13
Growth rates of GDP of NUTS 3 regions in post-crisis period	BIC	-62277.34	-62293.51	-62329.46
	Model	VEI	VEI	VEI
	Clusters	15	16	19