D2.5. Report on Spatiotemporal ESDA on GDP, Income and Educational Attainment in European regions

Kassoum Ayouba, Julie Le Gallo

Integrative Mechanisms for Addressing Spatial Justice and Territorial Inequalities in Europe

D2.5. Report on Spatiotemporal ESDA on GDP, Income and Educational Attainment in European regions

Version 1

Authors: Kassoum Ayouba, Julie Le Gallo

Grant Agreement No.: 726950

Programme call: H2020-SC6-REV-INEQUAL-2016-2017

Type of action: RIA – Research & Innovation Action

Project Start Date: 01-01-2017

Duration: 60 months

Deliverable Lead Beneficiary: UNIOVI

Dissemination Level: PU

Contact of responsible author: julie.le-gallo@agrosupdijon.fr

This project has received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreement No 726950.

Dissemination level:
- PU = Public
- CO = Confidential, only for members of the consortium (including the Commission Services)

Change control

VERSION	DATE	AUTHOR	ORGANISATION	DESCRIPTION / COMMENTS

Bibliographic information

Kassoum Ayouba, Julie Le Gallo (2019) *Report 5. Spatiotemporal ESDA on GDP, Income and Educational Attainment in European Regions*. IMAJINE WP2 Analysis of Territorial Inequalities in Europe

This report can be downloaded from our website http://imajine-project.eu/

Acknowledgements

We thank the research assistants Vincent Larmet and Abdou Nourzad. We thank Eleonora Elzeguabal, Gilles Lafferté, Julien Mischi and Lionel Védrine for useful comments.

Acronyms and Abbreviations

ESDA	Exploratory Spatial Data Analysis
ESTDA	Exploratory Space Time Data Analysis
EU	European Union
GIMA	Global Indicator of Mobility Association
LAU	Local administrative unit
LISA	Local Indicator of Spatial Association
MFA	Multiple Factor Analysis
NUTS	Nomenclature of Territorial Units for Statistics
PCA	Principal Component Analysis
WP	Work Package

Table of Contents

List of Tables

List of Figures

Introduction

With more than one third of the European Union budget devoted to Cohesion Policy, the regional policy of the European Union (EU), representing 351.8 billion euros for the 2014-2020 programming period, the effort provided by the European Union to support job creation, business competitiveness, economic growth, sustainable development and general improvement of citizens' quality of life, is considerable. Since the creation of the European Community, the six initial Member States already had the vision, set out in the founding Treaty, that the Community shall aim at reducing the disparities between the levels of development of the various regions. This gave the tone for subsequent regional policies. For 2014-2020, Cohesion Policy has set eleven thematic objectives covering various priorities such as strengthening research and R&D, support the shift towards a low-carbon economy or promoting sustainable employment, labor mobility and social inclusion. While the EU's regional policy covers all European regions, it is nevertheless mainly concentrated on less developed European countries and regions in order to help them catching up and reduce economic, social and territorial disparities that are still widely present in the EU, especially with the various enlargements.

Given these stakes, it comes at no surprise that the empirical literature devoted to the analysis of regional economic disparities in Europe is substantial and has given rise to numerous studies since the 80s. Existing studies can be broadly classified in two categories. On the one hand, confirmatory approaches to formal growth modeling are based on models set in the growth econometrics literature (Durlauf and Johnson, 2005) and focus on unconditional, conditional (the so-called β-convergence) or club convergence. On the other hand, a rather atheoretical exploratory literature departs from the representative economy assumption underlying most of regression approaches and examines instead the entire distribution of the variable of interest, typically income, using tools such as Markov chains and distribution dynamics. With the regional turn that this literature has taken from the end of the 90s, and because regions typically experience greater openness and heterogeneity than national economies, issues arising from the presence of spatial dependence and spatial heterogeneity in regional growth datasets have been largely explored (see for instance Rey and Le Gallo (2009) for a review). While confirmatory approaches use spatial econometric methods to tackle these issues, exploratory approaches have been developed to analyse the spatial and space-time mobility of income distributions (Rey *et al.*, 2011). This report, by implementing a large range of exploratory spatial data analysis (ESDA) and exploratory spatial and temporal data analysis (ESTDA) techniques, belongs to this second strand of the literature.

One common feature of these studies is that they overwhelmingly focus on a univariate measure of income, such as income per capita or Gross Domestic Product (GDP) per capita, as it is the main variable in testable predictions of growth models. Moreover, when it comes to analyzing European regional disparities, this choice is further supported by the fact that some European regional policies use thresholds of this measure to allocate specific or additional fundings, from the former Objective 1 regions, with GDP per capita less than 75% of the European average, to current transition (between 75% and 90%) and less developped (less that 75%) regions. Applications making use of ESDA applied to the distribution of GDP or income per capita in European regions include, among others, Lopez-Bazo et al. (1999), Le

Gallo and Ertur (2003), Dall'erba (2005) or Ertur and Koch (2006). Yet, other dimensions of disparities might be interesting. For instance, ESDA methods were applied on educational attainment and inequality in European regions (Rodríguez-Posé and Tselios, 2011; Chocolatá and Furková, 2017; Kalogirou, 2010), on human capital in Turkish districts (Erdem, 2016) or on social capital (Fazio and Lavecchia, 2013; Botzen, 2016) and demographic ageing (Gregory and Patuelli, 2015) in European regions. More generally, the use of other measures can be rooted in the debate pertaining to income or GDP as a very incomplete and partial measure of well-being and social welfare.

The first aim of this report is then to depart from the current state of the literature by implementing a vast range of ESTDA measures to synthetic measures covering various aspects of economic activity: economic development, education, population and employment dynamics. These synthetic measures are obtained from a multiple factor analysis based on a large range of indicators collected at the NUTS-2 level for the period 2000-2015 for the EU-28. While ESTDA measures have been applied to analyse the space-time dynamics of income distribution in US states and Chinese states (Rey and Ye, 2010), Mexican states (Gutiérreza and Rey, 2013), Canadian cities (Breau *et al.*, 2018) or other measures, such as total factor productivity (Di Liberto and Usai, 2013) in European regions, or burglary patterns in US cities (Rey and Ye, 2010), Mexican states (Gutiérreza and Rey, 2013), Canadian cities (Breau *et al.*, 2018) or other measures, such as total factor productivity (Di Liberto and Usai, 2013) in European regions, to the best of our knowledge, this is the first time that ESTDA methods are applied in such a way, i.e. by combining them with multiple factor analysis. We then extend the analysis by Del Campo *et al.* (2008) who also construct synthetic factors from a standard principal component analysis applied on a sample of European regions but their analysis remains static and they are not concerned with spatial issues. Further, as our first synthetic factor can be interpreted as economic development, we compare the results obtained for this factor to those obtained for GDP per capita and show that there are indeed substantial temporal and spatial differences.

The second aim of this report is to assess regional disparities at a much finer scale than what is commonly used in the literature, i.e. NUTS-2 or NUTS-3 levels. For that purpose, we use the data provided by the report D2.3 of Work Package 2 (WP2) of the IMAJINE Project.

The remainder of the report is organized as follows. In section 1, we briefly present a review of the papers that have applied ESDA and ESTDA on European regional data. In section 2, we present the database that we use in section 3 to perform the multiple factor analysis. In section 4, we analyze regional inequalities and their dynamics using the first component and contrast the results with those obtained with GDP per capita. Both a global and a local analysis are undertaken. In section 5, we briefly present the results obtained for three other meaningful factors. In section 6, we undertake an ESDA on the local data provided by D2.3. Section 7 concludes.

1. Literature review of ESDA and ESTDA applied to European regional data

Table 1.1 presents the studies that perform ESDA or ESTDA at the European scale. We report for each paper the authors, the sample, the variables analysed, the methods applied and the main results. We did not include the papers analysing only one country. Some papers reviewed here also include other methods, such as distributional dynamics or regression analysis but we only report in the table the main results obtained with the parts applying ESDA and ESTDA.

Early papers (López-Bazo et al., 1999; Le Gallo and Ertur, 2003, Dall'erba, 2005; Ertur and Koch, 2006; Ezcurra et al., 2007; Hierro and Maza, 2009) consider GDP per capita and GDP per worker and follow the same structure. First, they consider global spatial autocorrelation using Moran's *I* statistics and its evolution over time. Second, they perform a comparative statics analysis on Moran scatterplots and local spatial autocorrelation statistics (Local Moran statistics and Getis-Ord statistics). All the papers find positive and significant spatial autocorrelation in the distributions of GDP per capita levels and growth rates, together with spatial heterogeneity and polarization of the distribution of regional per capita GDP.

Ródriguez-Posé and Tselios (2011) are the first to apply ESDA on another indicator at the European level, i.e. educational attainment. They were followed in that path by Chocolatá and Furková (2017). Fazio and Lavecchia (2013) consider generalized trust survey data while Tselios and Stathakis (2019) use nighttime light data as a proxy of economic activity. While Ródriguez-Posé and Tselios (2011) and Tselios and Stathakis (2019) use ESDA as the previous papers, Fazio and Lavecchia (2013) also estimate LISA transition probabilities. ESTDA tools such as directional Moran scatterplots were only applied at the European scale on Total Factor productivity by Di Liberto and Usai (2013). Our report then goes beyond the current state of the literature by applying ESDA and ESTDA methods on a large range of socio-economic variables at the European scale.

Note also that, with the exception of Tselios and Stathakis (2019), all papers apply ESDA and ESTDA on data collected at aggregated spatial levels, i.e. NUTS1 or NUTS2 levels. The case of Tselios and Stathakis (2019) is particular: they work at a disaggregated spatial scale (LAU2) but with nighttime data as a proxy of economic activity. The report then also goes beyond the literature in applying ESDA on the spatially disaggregated data provided by the report D2.3 of WP2 of the IMAJINE Project.

Table 1.1: Summary of papers applying ESDA and ESTDA on European regional data

Authors	Sample	Variables	Methods	Main results
López-Bazo et al. (1999)	129/143 NUTS1/2 regions for the period 1981-1992	GDP per worker GDP per capita	Global SA (Moran's *I*) Local SA (GO and Local Moran statistics)	There are substantial differences between GDP per capita and GDP per worker with an absence of convergence in GDP per capita. The geographical location of clusters of high production regions shifts over time with the traditional core shifting southwards while there is an inability of poor regions to make significant moves up the ranking, translated by the persistence in the spatial clusters of low values in the traditional periphery.
Ertur and Le Gallo, 2003	138 NUTS2 regions for the period 1980-1995	GDP per capita GDP per capita annual growth rate	Global SA (Moran's *I*) Moran scatterplots Local SA (GO and Local Moran statistics)	Spatial autocorrelation and spatial heterogeneity characterize the distribution of regional per capita GDP. The overall picture is consistent with a North-South polarization of European regions but also considerable amount of spatial heterogeneity and spatial outliers.
Le Gallo and Ertur, 2003	138 NUTS2 regions for the period 1980-1995	GDP per capita in PPS GDP per capita annual growth rate	Global SA (Moran's *I*) Moran scatterplots Local SA (GO and Local Moran statistics)	Spatial autocorrelation and spatial heterogeneity characterize the distribution of regional per capita GDP. The overall picture is consistent with a North-South polarization of European regions but also considerable amount of spatial heterogeneity and spatial outliers.
Dall'erba, 2005	145 NUTS2 regions for the period 1989-1999	GDP per capita GDP per capita annual growth rates Structural funds	Global SA (Moran's *I*) Moran scatterplots Local SA (Local Moran statistics)	Positive and significant spatial autocorrelation in GDP per capita levels and growth rates. Evidence of two spatial clusters of rich and poor regions highlighting the persistence of a significant core-periphery pattern.
Ertur and Koch, 2006	258 NUTS2 regions for the period 1995-2000	GDP per capita in PPS GDP per capita annual growth rate	Global SA (Moran's *I*) Moran scatterplots Local SA (GO and Local Moran statistics)	Strong evidence of global and local spatial autocorrelation for per capita regional GDP. The enlargement process has led to a North-West-East polarization scheme.
Ezcurra *et al.*, 2007	71 NUTS1 regions for the period 1993-2000	Income from the European Community Household Panel	Global SA (Moran's *I* and Geary's *c*) Moran scatterplots Local SA (Local Moran statistics)	High level of positive spatial dependence in regional inequality levels. The most egalitarian income distributions are found in the Scandinavian countries, the Netherlands, Germany, Austria, and the north of Italy. In contrast, the regions with the highest degrees of income dispersion are located mainly in Ireland, the United Kingdom, Portugal, Spain, and Greece.
Hierro and Maza, 2009	196 NUTS2 regions for the period 1980-2005	GDP per capita in PPS	Global SA (Moran's *I*) Moran scatterplots	Existence of a pattern of positive spatial autocorrelation in the regional per capita income distribution.

Rodríguez-Pose and Tselios, 2011	102 NUTS2 regions for the period 1995-2000	Educational attainment	Global SA (Moran's *I*) Local SA (Local Moran statistics)	Strong correlation between levels of educational attainment and inequality across regions in Europe. Regions with similar educational conditions tend to cluster, often within national borders. a North–South and an urban–rural dimension are evident. Northern regions and large European metropolis have not only the most-educated labour force but also the lowest levels of inequality.
Chapman and Meliciani, 2012	180 NUTS0/1/2 regions for the period 1998-2005	GDP per capita	Global SA (Moran's *I*) Moran scatterplot	Strong but falling spatial autocorrelation of per capita income.
Di Liberto and Usai, 2013	199 NUTS2 regions for the period 1985-2006	Total factor productivity	Moran scatterplots Directional Moran Scatterplots	High and persistent level of TFP heterogeneity across EU regions with nonetheless constant positive spatial dependence. The cluster of successful regions becomes smaller and is turning more distant in economic and geographical terms from the cluster of less productive regions.
Fazio and Lavecchia, 2013	182 NUTS2 regions for 2002 and 2008	Generalized trust survey data	Global SA (Moran's *I* statistic) Moran scatterplots Local SA (Local Moran statistics) LISA transition probabilities	Spatial heterogeneity in the trusting attitude of the citizens of European regions. Clusters of regions with high levels and low levels emerge intranationally and across borders. Regional trust characterized by positive and increasing over time spatial autocorrelation.
Chocolatá and Furková, 2017	252 NUTS2 regions for the period 2007-2015	Educational attainment in upper secondary education	Global SA (Moran's *I* statistic) Local SA (Local Moran statistics)	Positive spatial autocorrelation and persistence of disparities in education attainment level across EU regions
Tselios and Stathakis, 2019	NUTS3 and LAU 2 regions fort the period 1992-2013	Nighttime light data as a proxy of economic activity	Local SA (Local Moran statistics)	The distribution of areas characterised by high or low economic activities is spatially dependent

Notes:
GO: Getis-Ord statistics
SA: spatial autocorrelation
PPS: purchasing power standard
TFP : Total factor productivity

2. Data description for analysis of socio-economic diversity of NUTS2 regions

Table 2.1 presents the socio-economic variables collected for the empirical analysis. They are grouped into five broad categories: demography, economy, employment, education, and health. Employment variables (by economic sector) come from the Cambridge Econometrics' European Regional Database (ERD) and the remaining ones from the Eurostat Database REGIO. The list of variables in Table 2.1 include 19 out of the 24 main regional indicators published in the third report on economic and social cohesion (European Commission, 2004)- variables with (*) in Table 2.1.[1] As in Del Campo *et al.* (2008), we exclude the remaining five variables as they do not meet the requirement necessary to undertake the empirical analysis, i.e. availability for all EU-28 countries[2] and expression as ratio to avoid scale problems. We then enrich and extend this first list using additional variables, which provide insights on other dimensions of disparities among European regions: demography variables (life expectancy, mean age of woman at childbirth, mean number of children), education variables (participation rate in education and training) and health variables (hospital beds and health personnel per 100,000 inhabitants).

Table 2.1: Regional indicators considered

Code	**Description**
Demography	
pop_dens (*)	Population density (100 inhabitants/km2)
pop_14 (*)	Percentage of the population aged less than 15 years
pop_1564 (*)	Percentage of the population between 15 and 64 years
pop_65 (*)	Percentage of the population aged more than 65
lifexp_0	Life expectancy at birth
lifexp_50	Life expectancy at 50
fert_age	Mean age of woman at childbirth
fert_rate	Mean number of children that would be born alive to a woman during her lifetime
Economy	
gdp_head (*)	GDP per head (PPS[3]) in deviation from the EU-28 average
emp_agri (*)	Agriculture, forestry and fishing employment (in % of total)
emp_indu (*)	Industry employment, excluding construction (in % of total)
emp_cons (*)	Construction employment (in % of total)
emp_trad (*)	Wholesale, retail, transport, etc. employment (in % of total)
emp_fin (*)	Financial and business services employment (in % of total)
emp_adm (*)	Non-market services employment (in % of total)

[1] We replace however the service employment variable by the following more disaggregated ones: emp_trad, emp_fin, and emp_adm. Moreover, we use an additional sectoral employment variable: emp_cons.

[2] For variables with a limited number of missing values, we made some adjustments presented in Table A1 in the appendix.

[3] GDP per capita is expressed in Purchasing Power Standards (PPS) to take into account price level variations between countries not reflected by prevailing exchange rates.

Employment	
emp_tot (*)	Total employment rate (ages 15-64 as % of pop. ages 15-64)
emp_fem (*)	Female employment rate (ages 15-64 as % of pop. ages 15-64)
emp_mal (*)	Male employment rate (ages 15-64 as % of pop. ages 15-64)
unemp_tot (*)	Total unemployment rate (%)
unemp_lt (*)	Long term unemployed (% of total unemployment)
unemp_fem (*)	Female unemployment rate (%)
unemp_you (*)	Youth unemployment rate (%)
neet_fem	Young female people neither in employment nor in education and training
neet_mal	Young male people neither in employment nor in education and training
neet_tot	Young people neither in employment nor in education and training
Education	
low_edu (*)	Active pop. with primary and lower secondary education (in %)
med_edu (*)	Active pop. with upper secondary and post-secondary non tertiary education (in %)
high_edu (*)	Active pop. with tertiary education (in %)
trng_fem	Female participation rate in education and training (last 4 weeks)
trng_mal	Male participation rate in education and training (last 4 weeks)
trng_tot	Total participation rate in education and training (last 4 weeks)
health	
bed_hos	Hospital beds per 100 000 inhabitants
per_health	Health personnel per 100 000 inhabitants

Source: Own elaboration using EuroStat and Cambridge Econometrics

Our sample includes 276 regions at the NUTS-2[4] level in 28 European countries over the period 2000-2015 (see Table A2 in the appendix for regions' distribution per EU-28 countries). However, we remove the region FRA5 – Mayotte, a remote island, as it presents numerous missing values for several variables. For a given year, we then have observations from 275 regions for each variable. We report in Table 2.2 the descriptive statistics for the considered variables from 2000 to 2015. Most display some huge asymmetries between EU-28 regions. The largest ones are observed for population density (a 1:3800 ratio between the minimum and maximum densities), and for the variables young people neither in employment nor in education and training on female, male and total population (between 1:1020 and 1:1250). Variations in the remaining demography variables are much less important compared to the population density variable: the ratios range from is 1:1.5 to 1:7.5. Regarding economy variables, GDP per capita (GDP p.c. hereafter) shows the highest dispersion (1:33) while the lowest concerns the variable wholesale, retail, transport, etc. employment (1:5). Overall, their variations are higher compared to demography variables excluding population density. While female, male and total employment variables exhibit some quite low dispersion (around 1:3), the unemployment variables' dispersion is important, specifically for female unemployment (1:50) and youth unemployment (1:40). Among education variables, the variable participation

[4] NUTS stands for Nomenclature of Territorial Units for Statistics used by Eurostat. In this nomenclature, NUTS-2 refers to Basic Administrative Units. It is the level at which eligibility to support from cohesion policy is determined.

rate in education and training shows very significant variations between regions (between 1:100 and 1:150). In comparison, the remaining variables of this group, related to the level of education display important, but less variations (between 1:10 and 1:30). Health variables' dispersion among regions is around 1:8.

Table 2.2: Descriptive statistics for variables used in the MFA

	Min	Max	Mean	Median	CV
pop_dens	0.0290	110.4030	4.4507	1.3300	258.1053
pop_14	9.9859	35.8286	16.2506	16.1227	16.3091
pop_1564	59.9493	75.2917	66.5646	66.2872	3.7152
pop_65	3.7722	28.0009	17.1848	17.1119	19.1708
lifexp_0	70.2000	84.9000	79.2019	79.8000	3.4797
lifexp_50	24.9000	36.1000	31.2996	31.8000	7.0644
fert_age	24.3000	33.5000	29.5739	29.6000	4.5424
fert_rate	0.8600	3.9400	1.5586	1.4900	19.4072
gdp_head	0.1865	6.1485	1.0000	0.9621	45.2418
emp_agri	0.0000	65.7544	6.5530	3.4977	127.2641
emp_indu	1.8152	38.9245	16.7328	15.8250	44.0677
emp_cons	1.4868	16.0431	7.1904	7.0202	28.2576
emp_trad	9.0437	50.3541	27.0972	26.6835	18.0114
emp_fin	1.7400	38.1456	12.2929	11.5496	42.1072
emp_adm	7.1711	60.2139	30.1336	30.1711	25.0354
emp_tot	37.8000	82.5000	64.4385	65.3000	12.7759
emp_fem	22.1000	80.5000	57.7791	59.1000	17.5644
emp_mal	46.3000	89.1000	71.1137	72.0000	10.5504
unemp_tot	1.2000	37.0000	9.0156	7.5000	60.7966
unemp_lt	7.2000	83.2000	40.2132	40.1000	34.6902
unemp_fem	1.0000	48.0000	9.8204	7.9000	65.7198
unemp_you	1.0000	79.2000	20.7442	18.2000	59.0285
neet_fem	0.0368	39.2891	12.8793	11.8000	51.8784
neet_mal	0.0310	38.6263	10.9271	9.8658	55.0817
neet_tot	0.0357	36.4800	12.0049	10.8000	49.1880
low_edu	2.7000	87.2000	29.3993	25.9000	54.5156
med_edu	6.9000	80.3000	46.6519	44.8000	32.3690
high_edu	3.7000	69.8000	24.0564	23.7000	38.7869
trng_fem	0.3704	48.5845	11.0283	7.5000	81.9212
trng_mal	0.4828	43.9065	9.0010	7.1000	70.6958
trng_tot	0.2963	46.1692	9.9782	7.2667	76.6901
bed_hos	161.5200	1239.2900	557.6706	547.5408	38.1760
pers_health	120.8500	976.2500	332.0476	324.6100	32.6312

3. Multiple factor analysis

Since we collected data for numerous variables (Table 2.1) informing on the regions' socio-economic conditions, we turn to data reduction techniques. Indeed, instead of analyzing the spatial pattern of each variable separately and then try to raise a global picture of regional inequalities, we extract the important information within our set of variables and express it as a collection of some (few) new orthogonal variables called factors. This could be achieved using the well-known method of principal component analysis (PCA). However, since our objective is the analysis of the dynamics of disparities from 2000 to 2015, PCA is not the appropriate tool. Indeed, if we apply as much PCAs as there are years of observations, it will likely result in factors that are not comparable over years. We therefore use an approach called Multiple Factor Analysis (MFA).[5] MFA extends PCA to analyze observations described by several variables that are characterized by a structure. This is particularly useful in our case since for each EU-28 region, observations on a given variable are grouped by time (from 2000 to 2015) and it is important to preserve this data structure. Specifically, MFA handles the multiple data tables[6] and derives an integrated picture of the observations and the relations between variables and between groups of variables with a two-step procedure. In the first step, the groups of variables are made comparable in order to avoid that the analysis is dominated by the group with the strongest structure. To this end, each group of variables is normalized by dividing all its elements by first singular value (matrix equivalent of the standard deviation). Then, the normalized data tables are concatenated into a unique data table which is submitted in the second step to PCA. As MFA boils down to a PCA on the concatenated-normalized data tables, the usual PCA outputs (coordinates, cosine, contributions, etc.) are available. Moreover, some specific-MFA outputs can also be derived to quantify the importance of each group in the common solution.

We apply MFA to extract a few principal components accounting for the major proportion of the total variance present in the dataset. Table 3.1 reports the eigenvalues (reflecting the importance of a component) of the first ten components derived from the analysis. The inertia of the first component is around 30%: this component explains up to 30% of the total variance in the original variables. The first four factors explain more than 65% of the total variance.

[5] The method has been introduced in Escofier and Pages (1983, 1994). For an extensive and comprehensive review, see Pagès (2014) and Abdi *et al*. (2018).

[6] For each region, the variables are grouped by time, from 2000 to 2015, i.e. there are 16 groups. For the first group Year-2000, variables are ordered as in Table 2.1.

Table 3.1: Multiple factor analysis - explained variance

Factor	Eigenvalue	% Variance	Cumulative % variance
1	15.05	30.64	30.64
2	8.58	17.46	48.10
3	5.11	10.41	58.51
4	3.64	7.40	65.91
5	3.01	6.12	72.04
6	1.82	3.71	75.74
7	1.57	3.20	78.94
8	1.41	2.87	81.81
9	0.98	2.00	83.81
10	0.94	1.91	85.72

Source: Own elaboration using EuroStat and Cambridge Econometrics

Table 3.2 contains the correlations between the first four factors and the original variables. In fact, since each variable appears sixteen times, as much correlation coefficients can be computed with the factors. However, since the correlation coefficients between factors and the yearly versions of each of the variables have a stable sign, we present, for ease of reading, the average correlation with factors for each variable. The most relevant correlations are shown in bold in Table 3.2. We also display in Table 3.2 the squared cosines of each variable to check for the quality of its projection on the four factors. All this information allows to interpret these factors.

Table 3.2: Multiple factor analysis – correlations and squared cosines

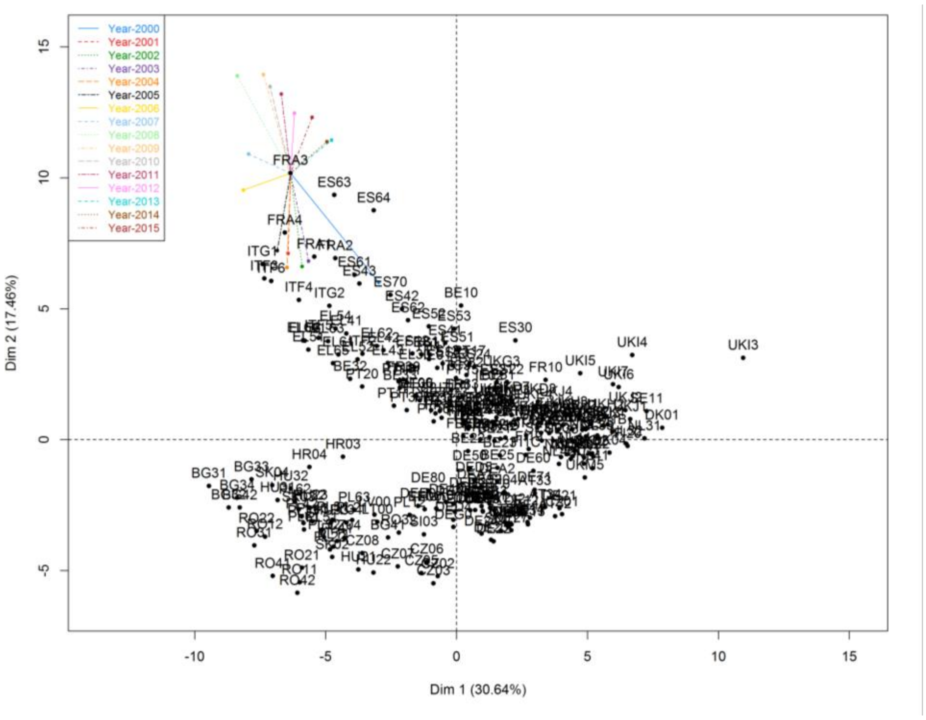
	Correlations				**Squared cosines**			
	F1	F2	F3	F4	F1	F2	F3	F4
pop_dens	0.25	0.23	0.27	**0.62**	0.06	0.05	0.07	**0.39**
pop_14	0.13	0.32	**0.74**	-0.09	0.02	0.12	**0.55**	0.01
pop_1564	-0.23	-0.36	0.09	**0.57**	0.06	0.13	0.03	**0.32**
pop_65	0.07	0.02	**-0.68**	-0.36	0.01	0.01	**0.46**	0.13
gdp_head	**0.68**	0.12	-0.15	**0.5**	0.47	0.02	0.02	0.25
emp_tot	**0.83**	-0.33	-0.02	-0.17	**0.7**	0.13	0.01	0.03
emp_fem	**0.80**	-0.4	0.11	-0.15	**0.64**	0.17	0.02	0.03
emp_mal	**0.76**	-0.19	-0.18	-0.16	**0.58**	0.10	0.04	0.03
unemp_tot	**-0.59**	0.39	0.1	0.19	0.35	0.22	0.02	0.05
unemp_lt	**-0.67**	0.02	-0.15	0.24	**0.45**	0.06	0.03	0.06
unemp_fem	**-0.61**	0.49	-0.02	0.15	0.38	0.27	0.01	0.03
unemp_you	**-0.62**	0.49	0.07	0.07	0.4	0.28	0.02	0.01
low_edu	-0.27	**0.76**	-0.2	-0.26	0.07	**0.57**	0.04	0.07
med_edu	-0.15	**-0.84**	0.11	0.11	0.02	**0.7**	0.01	0.01
high_edu	**0.69**	0.15	0.21	0.27	**0.47**	0.02	0.04	0.07
trng_fem	**0.75**	0.13	0.34	-0.15	**0.56**	0.02	0.12	0.03
trng_mal	**0.81**	0.10	0.29	-0.06	**0.65**	0.01	0.09	0.01
trng_tot	**0.78**	0.12	0.32	-0.11	**0.61**	0.01	0.11	0.01
fert_age	**0.51**	0.42	**-0.54**	0.22	0.26	0.18	0.3	0.05
fert_rate	0.32	0.36	**0.62**	-0.23	0.10	0.14	**0.39**	0.05
lifexp_0	**0.59**	**0.53**	-0.45	-0.04	0.34	0.28	0.2	0.00
lifexp_50	**0.54**	**0.58**	-0.42	-0.07	0.29	0.33	0.17	0.01
neet_fem	**-0.67**	0.40	0.23	0.01	**0.46**	0.17	0.06	0.00
neet_mal	**-0.65**	0.37	0.19	0.08	**0.43**	0.17	0.04	0.01
neet_tot	**-0.71**	0.42	0.20	0.02	**0.52**	0.20	0.04	0.01
bed_hos	-0.21	**-0.54**	-0.10	0.37	0.05	0.29	0.01	0.14
pers_health	-0.05	0.25	**-0.54**	0.29	0.01	0.06	0.29	0.09
emp_agri	**-0.64**	-0.19	0.04	-0.30	**0.41**	0.04	0.00	0.09
emp_indu	-0.30	**-0.69**	-0.18	-0.12	0.09	**0.48**	0.03	0.01
emp_cons	-0.10	0.10	-0.25	-0.27	0.02	0.06	0.08	0.08
emp_trad	0.35	0.31	-0.13	0.22	0.13	0.10	0.02	0.05
emp_fin	**0.66**	0.14	-0.02	**0.55**	**0.43**	0.02	0.00	**0.3**
emp_adm	0.35	**0.56**	0.30	0.00	0.12	**0.31**	0.09	0.00

Source: Own elaboration using EuroStat and Cambridge Econometrics

Factor 1, named *economic development* (ECO-DEV) is associated with a high level of the economic indicators presented in Table 2.2 (GDP and employment), a high level of education and a large number of jobs in financial and business services sectors. It also displays positive correlations with the rates of participation in education and training variables and negative correlations with the unemployment variables, the rate of people neither in employment nor in education and training variables and the agriculture, forestry and fishing sector employment. Factor 2, named *low education* (LOW-EDUC), globally expresses high rate of active population with pre-primary, primary and lower secondary education and low rate of active population with upper secondary and post-secondary levels. It is also associated with a low number of jobs in the industry sector. Factor 3, named *population dynamics* (POP-DYN), is associated with a high percentage of children (population aged less than 15 years) and a low percentage of retired people (population aged more than 65). Factor 3 also shows a positive association with fertility rate. Therefore, a region with a high score on this factor is young and dynamic. Factor 4, named *active population* (ACT-POP) is associated with regions with high population density, high percentage of active adults and also moderately associated with a large number of jobs in financial and business services sectors and with a high GDP per capita. Therefore, regions with a high a value on this factor are those with attractive and competitive employment centers.

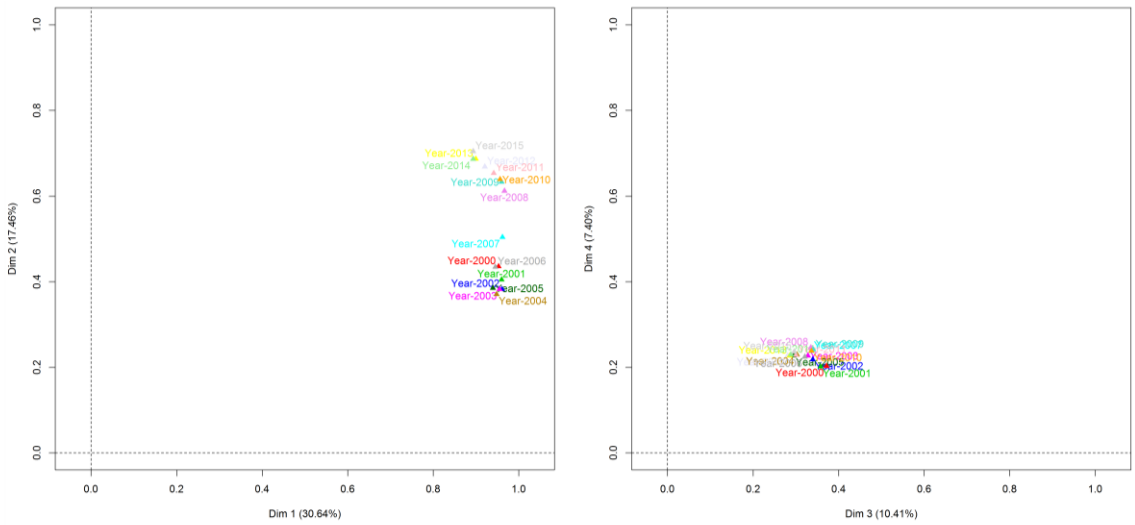
The individuals' graph can be used to identify the regions that contribute the most to the components in the global space defined by MFA. Geometrically, the individual factor score can be interpreted as their projection in the space defined by the principal components. To see the year-by-year regions fit into the global MFA space, we project the dataset of each group as a supplementary element. As an illustration, Figure 3.1 shows the group-year location of the region FRA3 in the global space (focusing on the first-two dimensions). The global position of this region is nothing else that the barycenter (centroid) of its positions from 2000 to 2015. The projection of year-by-year regions into the common global space returns the year-specific factor scores that are used in the subsequent spatio-temporal exploratory analysis.

Figure 3.1: Individuals graphs, with a focus on FRA3



We also take advantage of the group structure to see the contribution of the set of year variables to the dimensions from the MFA (see Figure 3.2). From this Figure, we observe that the year variables' contribution to the first, third and fourth components are fairly stable over time whereas their contribution to the second component is increasing. Specifically, for the second component, there is a gap between the structure of years 2000 to 2007 and years 2008 to 2015 (see left panel of Figure 3.2). The latter group of variables explains more the second component than the former, which could be an effect of the crisis involving some structural change.

Figure 3.2: Contributions of the groups of variables to factors 1 to 4



Factor 1, from its correlations with our original variables and its squared cosines, stands as a variable that provides indications of the economic situation of regions beyond GDP p.c. It can therefore be seen as an answer to the several limits of GDP pointed out in the literature (e.g. Robert *et al*. 2014; Fleurbaey, 2009). The next section is dedicated to the analysis of this factor: we analyze the regional disparities at work within EU-28 and their dynamics using Factor 1, while comparing the results with those obtained with GDP p.c. alone. Then, we complement the picture obtained with the analysis of Factors 2 to 4.

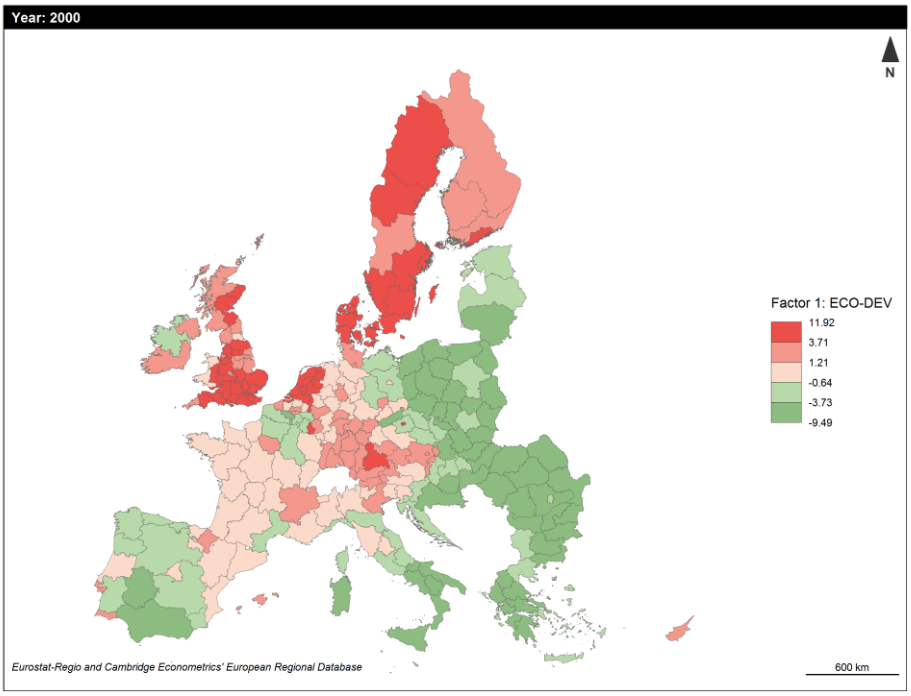
4. Beyond GDP per capita

We analyze regional inequalities and their dynamics from 2000 to 2015 within EU-28 using the first component derived from MFA: ECO-DEV. Throughout the section, we contrast the results obtained with ECO-DEV with those obtained with GDP p.c. We start investigating at a global scale to what extent the distribution of ECO-DEV and its evolution over time is associated with spatial dependence. Then we move at a more local scale to highlight regional dynamics at work and complement the analysis by testing how local patterns detected are different from what would be expected if ECO-DEV values were randomly distributed in space.

4.1 Distribution dynamics and spatial pattern: a global assessment

To start, we display in choropleth maps using a quintile classification (see Figure 4.1), the spatial distribution of ECO-DEV in 2000 and in 2015. The darker the red (green) color, the most (less) developed the corresponding region. The visual inspection of these choropleth maps suggests a spatial clustering of similar values. In 2000, we identify a group of poor regions belonging to Portugal, Spain, Italy, Eastern borders countries (Greece and countries of the former Eastern bloc (Poland, Romania, Bulgaria, etc.) and on either side of France-Belgium border. This is contrasted by rich regions located mainly in United Kingdom, Sweden, Denmark, the Netherlands, Austria and the south-west of Germany. Fifteen years and one financial crisis later, the spatial pattern of ECO-DEV has not significantly changed. Compared to the picture provided by GDP p.c. (see Figure 4.2), we note two main things. First, regions in UK along with Scandinavian regions (excluding Norwegian regions, not in EU-28) appear relatively richer with ECO-DEV than with GDP p.c. Second, northern Italian regions, along with regions from Austria and Germany appears relatively poorer with ECO-DEV compared to GDP p.c. The well documented dualism of the Italian economy is therefore flagrant with GDP p.c. but less so when taking into account other variables.

Figure 4.1: Choropleth maps for ECO-DEV (2000-2015), quintile classification



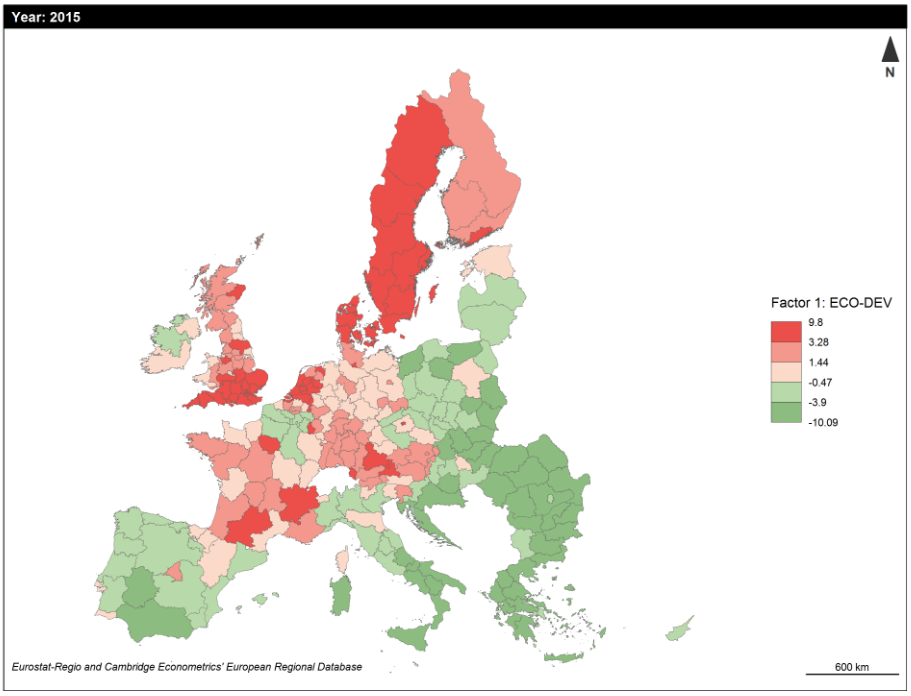
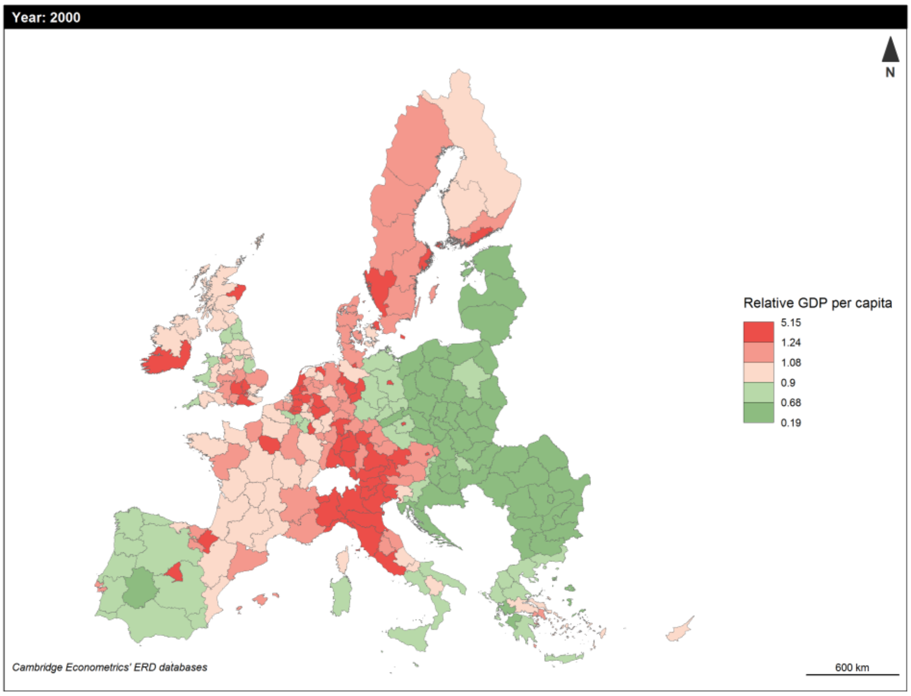
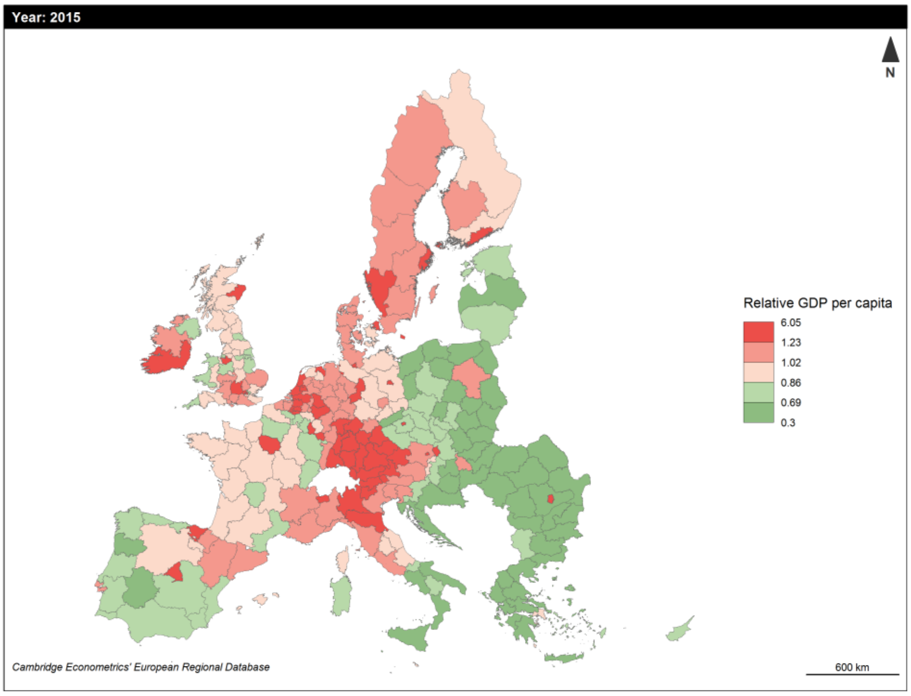


Figure 4.2: Choropleth maps for GDP per capita (2000-2015), quintile classification





This pattern of spatial clustering and its dynamics are explored in more detail using two global indexes: Moran's *I* global spatial autocorrelation statistic I and a global indicator of mobility association (GIMA): τ_W. The latter is an extension of Kendall's rank correlation statistic developed by Rey (2004) and provides indication on space-time concordance, i.e. spatial mobility in the distribution of ECO-DEV over time. As pointed out by Rey (2016), these two global measures should be thought as complements. Indeed, the space-time concordance statistic informs on how the pairwise *ordinal* associations between neighboring values evolve over time, while the evolution of the global autocorrelation statistic captures how pairwise *interval* associations change between time periods. Formally, global Moran's *I* is expressed as follows for the sample of EU-28 regions ECO-DEVs observed in period t:

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

where $z_{i,t}$ is the deviation from the mean of ECO-DEV observed in region i and period t. n is the number of regions and w_{ij} is the (i,j) element of the spatial weight matrix and expresses how region i is spatially connected to region j. By convention $w_{ii} = 0$. s_0 is a scaling factor equal to the sum of all the elements of the weight matrix ($s_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$). When the weight matrix is row-standardized, the expression (1) simplifies as $s_0 = n$. A value over (resp. below) $E(I) = -1/(n-1)$ indicates global positive (resp. negative) global spatial autocorrelation. Inference is based on a permutation approach.

To present the GIMA, we start with the general rank correlation coefficient (Kendall, 1962):

$$\tau(f,g) = \frac{c-d}{n(n-1)/2} \tag{2}$$

where c is the number of concordant pairs and d the number of discordant pairs. Then, the numerator reflects the net concordance between all pairs of observations. In our application $f = z_{t-1}$ and $g = z_t$. τ ranges from -1 (perfect discordance) to 1 (perfect concordance).

A mobility index M is derived from τ as follows: $M = [\tau(f,g) - 1]/(-2)$, with $0 \leq M \leq 1$. Larger M is an indication of greater distributional mixing. Specifically, $M = 0$ implies an absence of rank mobility, while $M = 1$ is an indication of full ranking mobility. Rey (2004) extends this traditional rank correlation measure to incorporate a spatial dimension. Specifically, τ is decomposed as follows:

$$\tau(z_{t-1}, z_t) = \phi\tau_W(z_{t-1}, z_t) + (1-\phi)\tau_{\overline{W}}(z_{t-1}, z_t) \tag{3}$$

where $\phi = \iota W \iota^T / \iota(W + \overline{W})\iota^T$ represents the share of all pairs that involve geographic neighbors. W and $\overline{W} = \iota\iota^T - W - I_{n \times n}$ are matrices capturing respectively neighboring and non-neighboring relationships. ι is the $(n \times 1)$ unit vector. The decomposition in Equation (3) allows to determine to what extent the classic general rank correlation coefficient measure is silent about the correlation patterns between neighboring and non-neighboring regions. This can be inferred based on random spatial permutations of the attributes to develop a distribution for τ_W under the null hypothesis of spatial homogeneity (Rey, 2016). The mobility index (M) can also be decomposed as follows: $M = \phi M_W + (1-\phi)M_{\overline{W}}$, with $M_W = [\tau_W - 1]/(-2)$.

We report in Table 4.1 the values for the global measures of spatial autocorrelation and rank concordance over 2000-2015. The spatial weight matrix used to construct these statistics is based on k-nearest neighbors calculated from the great circle distance between regions' centroids. Since this weighting scheme avoids the problem of isolated regions having non neighbors, it is very useful for our case with on a dataset composed of some islands. The k-nearest neighbors weight matrix is defined as follows:

$$\begin{cases} w_{ij}(k) = 0 \text{ if } i = j, \forall k \\ w_{ij}(k) = 1 \text{ if } d_{ij} \leq d_i(k) \\ w_{ij}(k) = 0 \text{ if } d_{ij} > d_i(k) \end{cases} \tag{4}$$

where $d_i(k)$ is the k^{th} order smallest distance between regions i and j such that region i has exactly k neighbors. We set $k = 10$ to guarantee spatial connection between regions belonging to different countries[7] and avoid a block-diagonal structure of the weights matrix (Le Gallo and Ertur, 2003). With $k = 10$, 34.25% of the 10-nearest neighbors belong to a different country.

The evolution of Moran's I over the period reveals a positive significant and quite stable spatial autocorrelation for all years: the distribution of ECO-DEV is spatially clustered within a given time period (see Table 4.1). This confirms the visual inspection results: rich (resp. poor) regions are localized close to regions with relatively high (resp. low) value of ECO-DEV more often than if their localization were purely random. Interestingly, the estimated standardized Moran's I statistics are much more important than the ones obtained for GDP p.c. (see Table 4.2) and remain stable over time while the level of global spatial autocorrelation for GDP p.c. decreases over time. This reveals the existence of strong disparities within European regions when considering a synthetic index and unlike GDP p.c., the 2008 financial crisis did not have any discernible global effect on the spatial agglomeration of countries.

[7] For example, to connect Greece to Italy.

Table 4.1: Spatial autocorrelation and spatial concordance for ECO-DEV

Year	Moran's I	$E(I)$	$SD(I)$	Standardized value	p-value	Period	τ_W	$E(\tau_W)$	p-value
2000	0.7642	-0,0036	0.0248	30.9050	<0.0001				
2001	0.7659	-0,0036	0.0248	30.9806	<0.0001	2000-2001	0.8652	0.9227	<0.001
2002	0.7712	-0,0036	0.0248	31.1852	<0.0001	2001-2002	0.8681	0.9375	<0.001
2003	0.7861	-0,0036	0.0248	31.7864	<0.0001	2002-2003	0.8030	0.9158	<0.001
2004	0.7928	-0,0036	0.0248	32.0526	<0.0001	2003-2004	0.7956	0.9165	<0.001
2005	0.7769	-0,0036	0.0248	31.4103	<0.0001	2004-2005	0.7970	0.9063	<0.001
2006	0.7444	-0,0036	0.0248	30.1039	<0.0001	2005-2006	0.8681	0.9362	<0.001
2007	0.7199	-0,0036	0.0248	29.1221	<0.0001	2006-2007	0.8563	0.9128	<0.001
2008	0.7071	-0,0036	0.0248	28.6085	<0.0001	2007-2008	0.8548	0.9201	<0.001
2009	0.7015	-0,0036	0.0248	28.3790	<0.0001	2008-2009	0.8370	0.9147	<0.001
2010	0.7168	-0,0036	0.0249	28.9868	<0.0001	2009-2010	0.8504	0.9267	<0.001
2011	0.7348	-0,0036	0.0249	29.7150	<0.0001	2010-2011	0.8756	0.9198	<0.001
2012	0.7549	-0,0036	0.0249	30.5183	<0.0001	2011-2012	0.8415	0.9339	<0.001
2013	0.7724	-0,0036	0.0249	31.2182	<0.0001	2012-2013	0.8385	0.8986	<0.001
2014	0.7691	-0,0036	0.0249	31.0931	<0.0001	2013-2014	0.8844	0.9447	<0.001
2015	0.7633	-0,0036	0.0248	30.8618	<0.0001	2014-2015	0.8711	0.9395	<0.001

Table 4.2: Spatial autocorrelation and spatial concordance for GDP per capita

Year	Moran's I	$E(I)$	$SD(I)$	Standardized value	p-value	Period	τ_W	$E(\tau_W)$	p-value
2000	0.4394	-0,0036	0.0238	18.6218	<0.0001				
2001	0.4308	-0,0036	0.0238	18.2446	<0.0001	2000-2001	0.9581	0.9636	<0.001
2002	0.4218	-0,0036	0.0238	17.8856	<0.0001	2001-2002	0.9537	0.9658	<0.001
2003	0.4074	-0,0036	0.0236	17.3835	<0.0001	2002-2003	0.9473	0.9562	<0.001
2004	0.3867	-0,0036	0.0235	16.6105	<0.0001	2003-2004	0.9480	0.9623	<0.001
2005	0.3621	-0,0036	0.0234	15.6336	<0.0001	2004-2005	0.9502	0.9592	<0.001
2006	0.3508	-0,0036	0.0234	15.1233	<0.0001	2005-2006	0.9364	0.9634	<0.001
2007	0.3232	-0,0036	0.0233	14.0501	<0.0001	2006-2007	0.9440	0.9549	<0.001
2008	0.2958	-0,0036	0.0232	12.9127	<0.0001	2007-2008	0.9207	0.9541	<0.001
2009	0.2709	-0,0036	0.0229	11.9633	<0.0001	2008-2009	0.9381	0.9471	<0.001
2010	0.2749	-0,0036	0.0229	12.1645	<0.0001	2009-2010	0.9205	0.9354	<0.001
2011	0.2857	-0,0036	0.0229	12.6305	<0.0001	2010-2011	0.9420	0.9418	<0.001
2012	0.2924	-0,0036	0.0228	12.9620	<0.0001	2011-2012	0.9410	0.9607	<0.001
2013	0.2876	-0,0036	0.0228	12.7803	<0.0001	2012-2013	0.9513	0.9684	<0.001
2014	0.2745	-0,0036	0.0224	12.4111	<0.0001	2013-2014	0.9648	0.9707	<0.001
2015	0.2703	-0,0036	0.0225	12.1536	<0.0001	2014-2015	0.9523	0.9759	<0.001

We then move to the global indicator of mobility association. It moves the comparative static view (temporal sequence of Moran's I) toward an explicit consideration of spatial dynamics as it formally links a measure in one region in time to another measure for the same region at a different time period. The spatial concordance rank comes as a complement to Moran's I: even if the spatial distribution of ECO-DEV exhibit the same shape over long periods of time, it may actually mask a great internal mixing as regions may move up and down in the distribution of ECO-DEV within neighboring groups. For all the periods (second part of Table 4.1), the degree of rank concordance between neighboring pairs is significantly lower than what would be expected under spatial randomness of rank changes. This would mean (using the transformation from τ_W to M_W) that the mobility between neighboring pairs is significantly more important than the one expected under spatial randomness of rank changes within the observed periods, and since we observe a persistence of spatial clustering observed with Moran's I, this would suggest that this mobility between neighboring pairs is in the same direction. Looking at the results obtained for GDP p.c. in Table 4.2, we note that i) for almost half of the period, there is no significant difference between the mobility rate between neighboring pairs and the mobility rate expected under spatial randomness of rank changes, ii) for the remaining periods, differences exist: the mobility rate between neighboring pairs is significant and more important than the one obtained under spatial randomness. These mobility gaps are however smaller compared to the ones obtained for ECO-DEV. Overall, the mobility between neighboring pairs, higher for ECO-DEV compared to GDP p.c., provides an explanation to the high and persistent Moran's I found with ECO-DEV.

4.2 Going local: a closer look at spatial dependence and its dynamics

We continue the investigation of the spatial dynamics at work in the distribution of ECO-DEV over time at a local level. Indeed, since global Moran's I statistic and the global indicator of mobility association τ_W yield a single result for the entire dataset for a given year, they may mask more complex local dynamics. In our case of positive global autocorrelation, Moran's I fails to discriminate between a spatial clustering of low values and a spatial clustering of high values. Therefore, we use local indicators of spatial association (LISA) in conjunction with Moran scatterplots for a closer view of the spatial dependence and its dynamics, firstly between initial and final periods (Directional LISA) and secondly between a sequence of many periods (Markov LISA).

We start by mapping the significant LISA statistics. The LISA statistic in region i at time t ($L_{i,t}$), which formalizes the relationship between each observation of ECO-DEV and the weighted average of its neighbors (see Anselin, 1995), is defined as:

$$L_{i,t} = z_{i,t} \sum_{j=1}^{n} w_{ij} z_{j,t} \tag{5}$$

with the same notations as before. Then, the LISA statistic is decomposed, i.e. each region in a given time period t is positioned in a Moran scatterplot using the coordinates $z_{i,t}$ and $\sum_{j=1}^{n} w_{ij} z_{j,t}$.[8] Inference is based on 9,999 random permutations and we use the Bonferroni p-value correction to

[8] The four quadrants of the Moran scatterplot report different types of spatial association between a region's ECO-DEV and that of its neighbors. In the first quadrant are located developed regions (regions with a relatively high ECO-DEV), neighbored by similar regions (High-High or HH). Quadrant two contains regions with relatively low ECO-DEV with developed neighbors (Low-High or LH), while quadrant three contains regions with a relatively low ECO-DEV with similar neighbors (Low-Low or LL). Finally, in quadrant four are located developed regions neighbored by regions with a relatively low ECO-DEV (High-Low or HL).

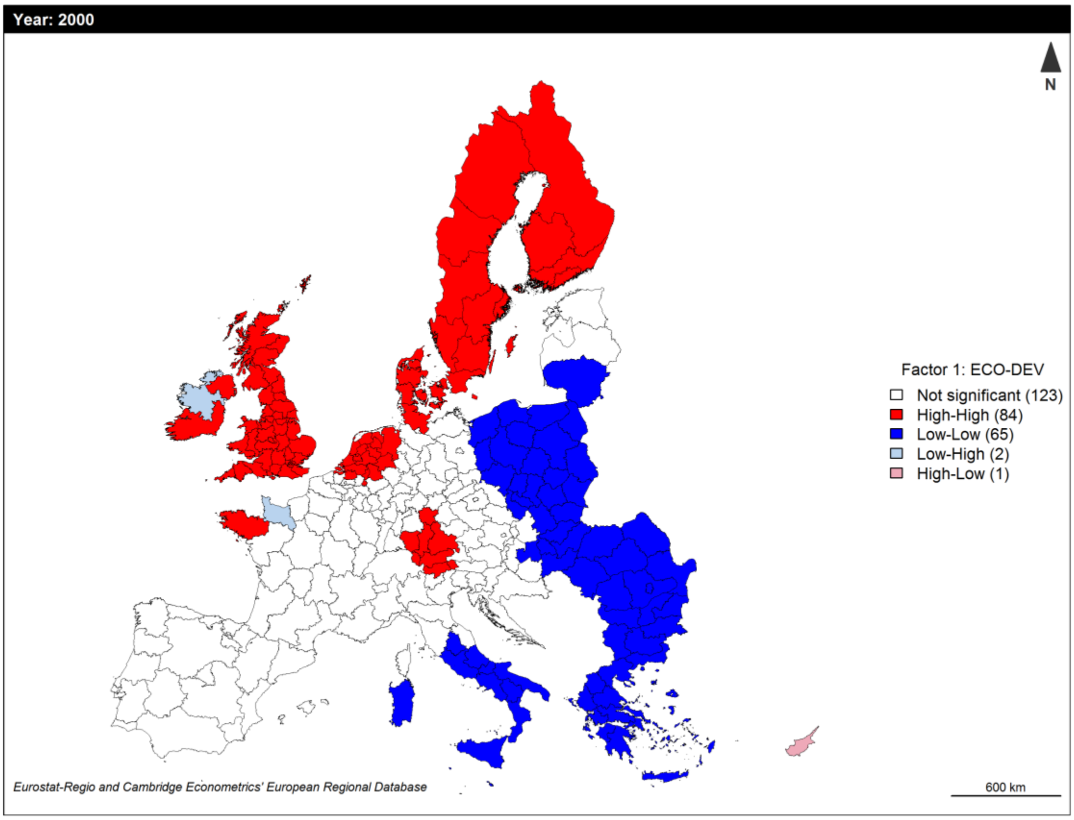
deal with the multiple comparison problem. Specifically, we set to 0.05 the overall significance associated with multiple comparisons. Then since each observation (region) has ten neighbors, the individual significance is set to 0.05/10 = 0.005 as at most 10 comparisons can be made for one observation. The significant clusters identified with the local Moran in 2000 and in 2015 are displayed in Figure 4.3.

Several points can be highlighted. First, the local pattern of spatial association reflects the global pattern of positive spatial autocorrelation since almost all significant clusters are either high-high (hot-spot) or low-low (cold-spot) ones in 2000 and in 2015. Second, at the beginning of the period, Figure 4.3 shows four big clusters of rich regions. The first one is located in United Kingdom. The second one includes regions from Scandinavian countries (except Norway). The last two clusters gather together regions from the Netherlands and from the west of Austria and the south-west of Germany. These clusters are highly persistent over time, the cluster located the west of Austria and the south-west of Germany becomes even bigger in 2015.[9] Third, in 2000, spatial agglomerations of poor regions are located mainly in Greece, in the south of Italy and in countries of the former Eastern bloc. This cluster is also highly persistent over time even if one can note that the position of Croatia and southern Spain and Portugal regions worsens in 2015.[10] While this picture is overall close to the one provided by GDP p.c. (see Figure 4.4), and that found by Ertur and Koch (2006) for the period 1995-2000, three significant changes may be highlighted. First, with GDP p.c., the cluster of rich regions is concentrated only in the south of UK, around London whereas with ECO-DEV, almost all regions from UK are in that cluster. Second, with GDP p.c., a cluster of high values is identified in central Europe and goes across several countries from the north of Italy to the west of Austria and the south-west of Germany. With ECO-DEV, this cluster shrinks significantly. Finally, in 2015, we are able to detect clusters of low value with regions located at the south of Spain and Portugal and in Croatia with ECO-DEV, not detected otherwise with GDP p.c. At this stage, the results show that the number of clusters of similar values identified is more important with ECO-DEV compared to GDP p.c. This would mean that concentration of EU-28 regions within blocks of rich and poor is more pronounced while considering ECO-DEV instead of GDP p.c., in line again with global Moran's I statistics, much more important with ECO-DEV than with GDP p.c. (a standardized value of 30.90 versus 18.24 in 2000 and 30.86 versus 12.15 in 2015).

[9] In total, from 2000 to 2015, more than 96.42% of regions in high-high clusters remain in the same cluster. In addition, 7.32% of regions belonging to the non-significant cluster in 2000 move in the significant high-high cluster in 2015, amplifying the spatial association of high-high regions.

[10] From 2000 to 2015, more than 87% of regions in low-low clusters remain in the same cluster. In addition, 11.38% of regions belonging to the non-significant cluster in 2000 move in the low-low cluster in 2015, amplifying the spatial association of low-low regions.

Figure 4.3: Local Moran clusters for 2000 and 2015 (ECO-DEV) p – value = 0.05 with Bonferroni adjustment



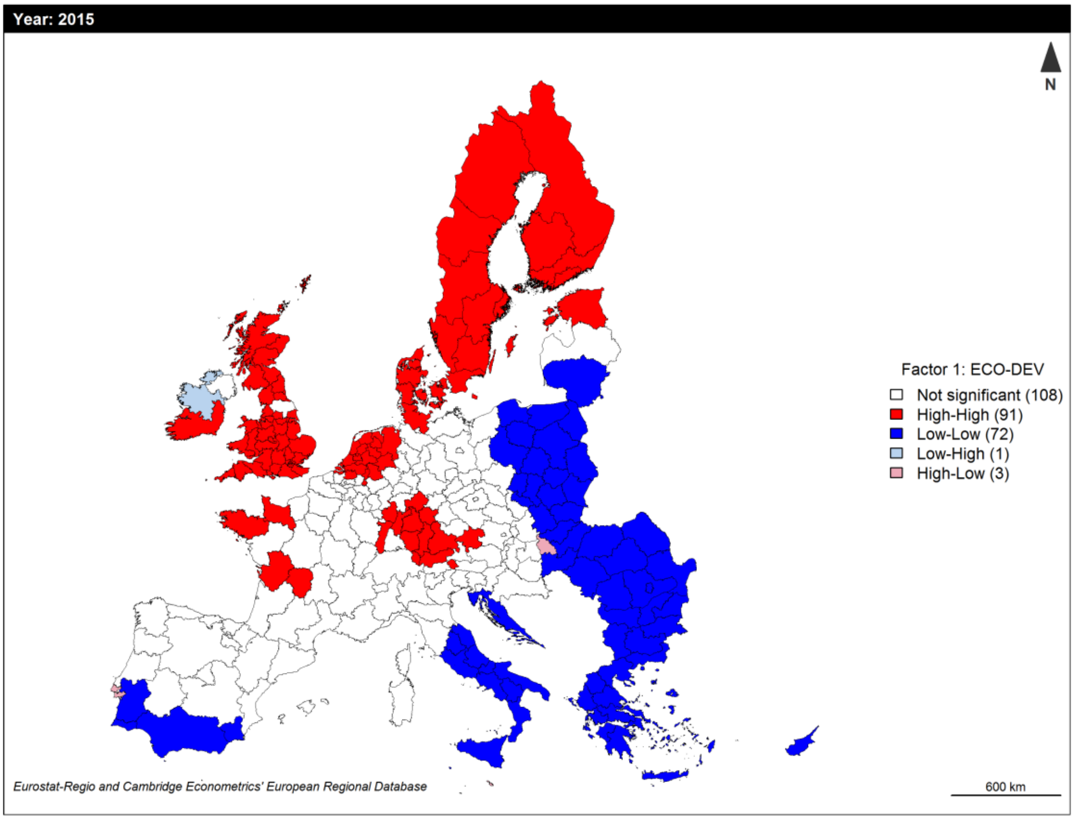
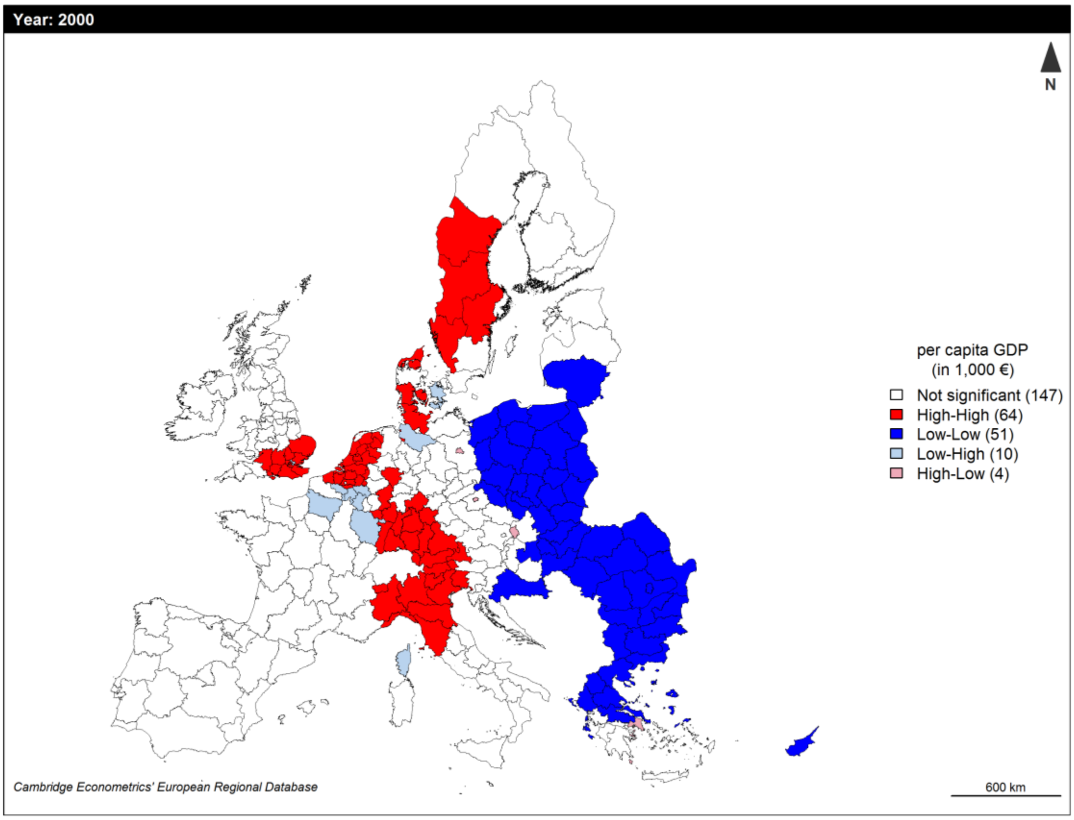
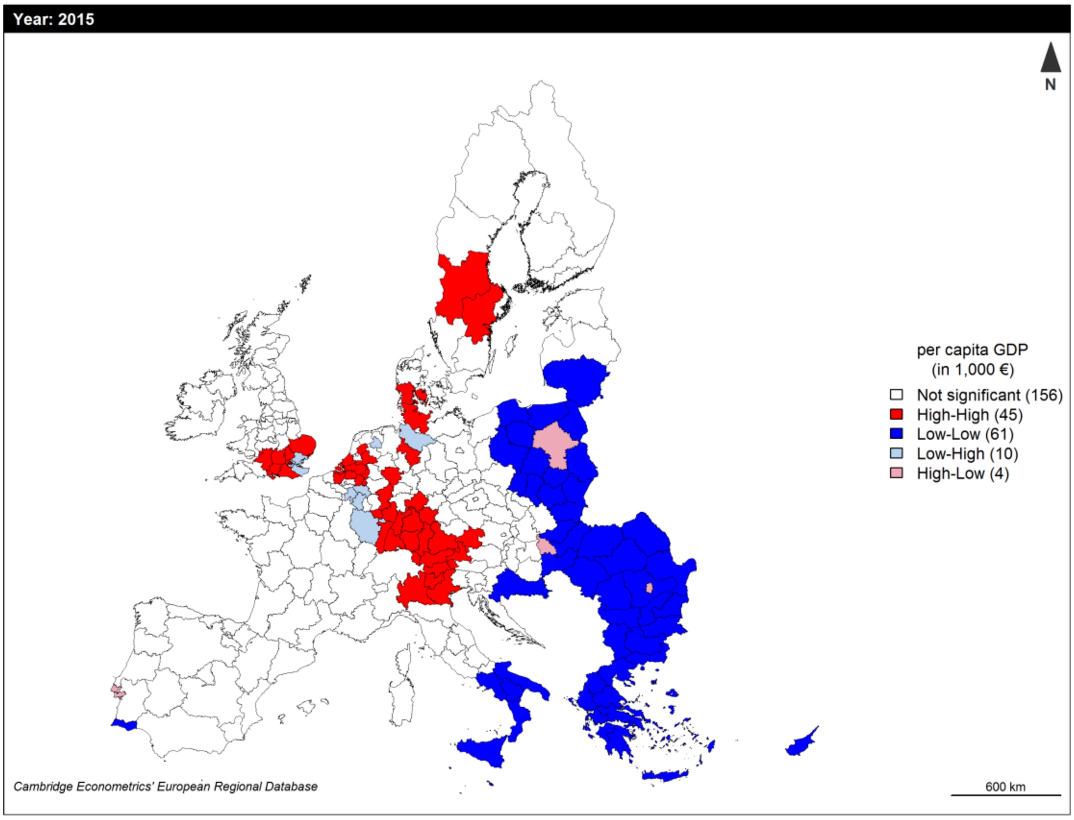


Figure 4.4: Local Moran clusters for 2000 and 2015 (GDP per capita) p – value = 0.05 with Bonferroni adjustment





We now deepen this comparative statics analysis with the directional approach proposed by Rey et al. (2011) which links Moran scatterplots across two time periods and tracks the changes over time. Directional LISA then enables to plot the directions of the movement vectors standardized either by their beginning or ending points. These movement vectors, which represent the transitions that each region has experienced between the first and the last time period considered, centered in the origin of axes, allows a visualization of the direction and magnitude of the spatial dynamics between the two dates.

Moran scatterplots for the initial and final years (2000 and 2015) are displayed in Figure 4.5. We observe that most regions are characterized by positive spatial association, in line with the results of the global autocorrelation statistic. More specifically, 82.17% of regions exhibit association of similar values (44.36% localized in quadrant HH and 37.81% in quadrant LL in 2000. With GDP p.c., the figures obtained are relatively close (see Figure 4.6). The spatial pattern observed in 2000 with ECO-DEV persists and is even more pronounced in 2015: 86.54% of regions are localized in quadrants HH (51.27%) and LL (35.27%). These figures diverge from the ones obtained with GDP p.c. in 2015 for which 79.34% of regions are localized in quadrants HH (34.42%) and LL (44.92%).

Figure 4.5: Moran scatterplots for 2000 and 2015, ECO-DEV

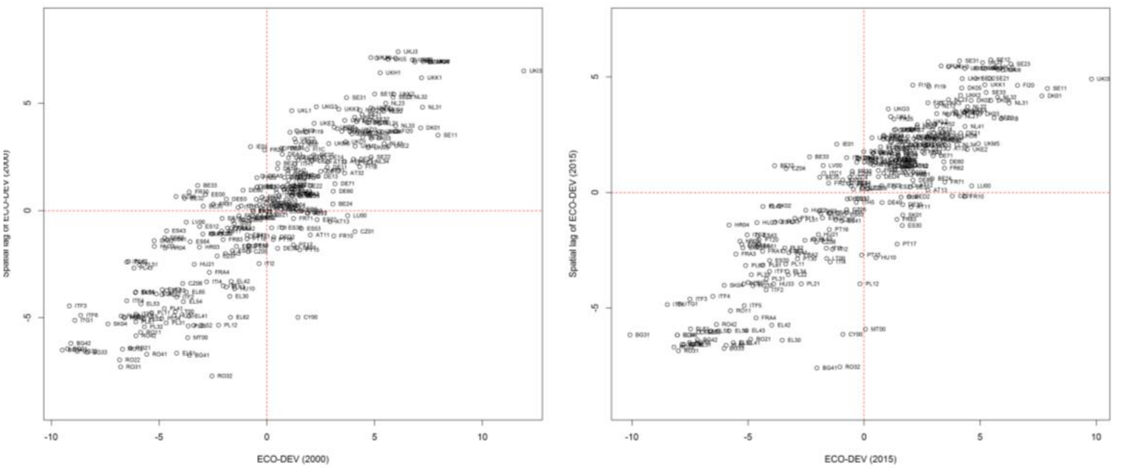
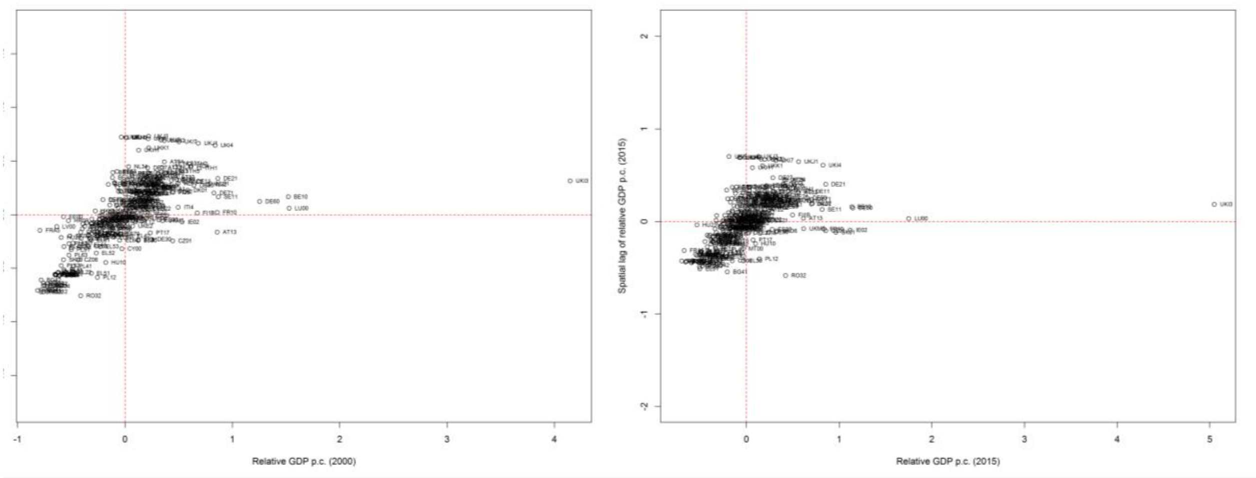


Figure 4.6: Moran scatterplots for 2000 and 2015, GDP per capita



We then contrast the two cross-sectional views of ECO-DEV, i.e. the two Moran scatterplots at the beginning and at the end of the period. This graphically illustrates the transition of each region along with its neighbors between the two times periods in the Moran scatterplot. The transition of each region is represented as a movement vector with the arrowhead pointed at its location in the ending period. Then, we normalize the direction vectors to obtain a standardized directional Moran scatter plot: the vectors are standardized at the origin to reflect their positions in the 2000 Moran scatterplot. Moves to north-east in Figure 4.7 (left panel) reflect simultaneous positive co-movements (gain) of a region and its neighbors in ECO-DEV distribution. Conversely, movements to south-west reflect a simultaneous worsening of the position of the region and that of its neighbors in the distribution of ECO-DEV (see Rey, 2001). We also report in Figure 4.7 (right panel), movements with points instead of arrows to ease the visualization. Indeed, with arrows, long arrows may hide the presence of short ones. To help the interpretation of movements displayed in Figure 4.7, we also provide in Figure 4.8 their dynamics, depending on the location of regions at the beginning of the period. For example, in the first panel of Figure 4.8 (top-left), points in red are the regions in 2015 that were located in the HH quadrant in 2000. Several observations can be made from the observation of these movements. First, most regions that were located in the HH quadrant in 2000 either improve their situation along with their neighbors or worsen it along with their neighbors. One can note, looking at the length of the line from the origin of the scatter plot to the region position in 2015 that the magnitude of the deterioration of the economic situation is more important than the improvement one. Second, the economic situation of regions in LL and in HL is somewhat balanced in the four possible directions. Third, the situation of almost all regions that were located in the LH quadrant in 2000 improves in 2015. This may reflect a limited convergence process at work in the EU-28 regions from 2000 to 2015.

Figure 4.7: Standardized Moran scatterplots for 2000-2015, ECO-DEV

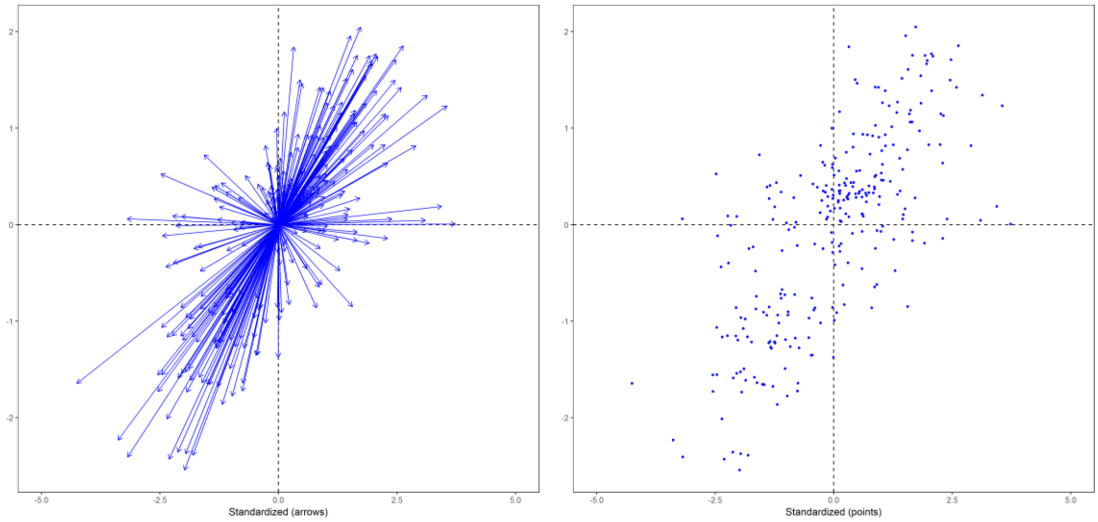
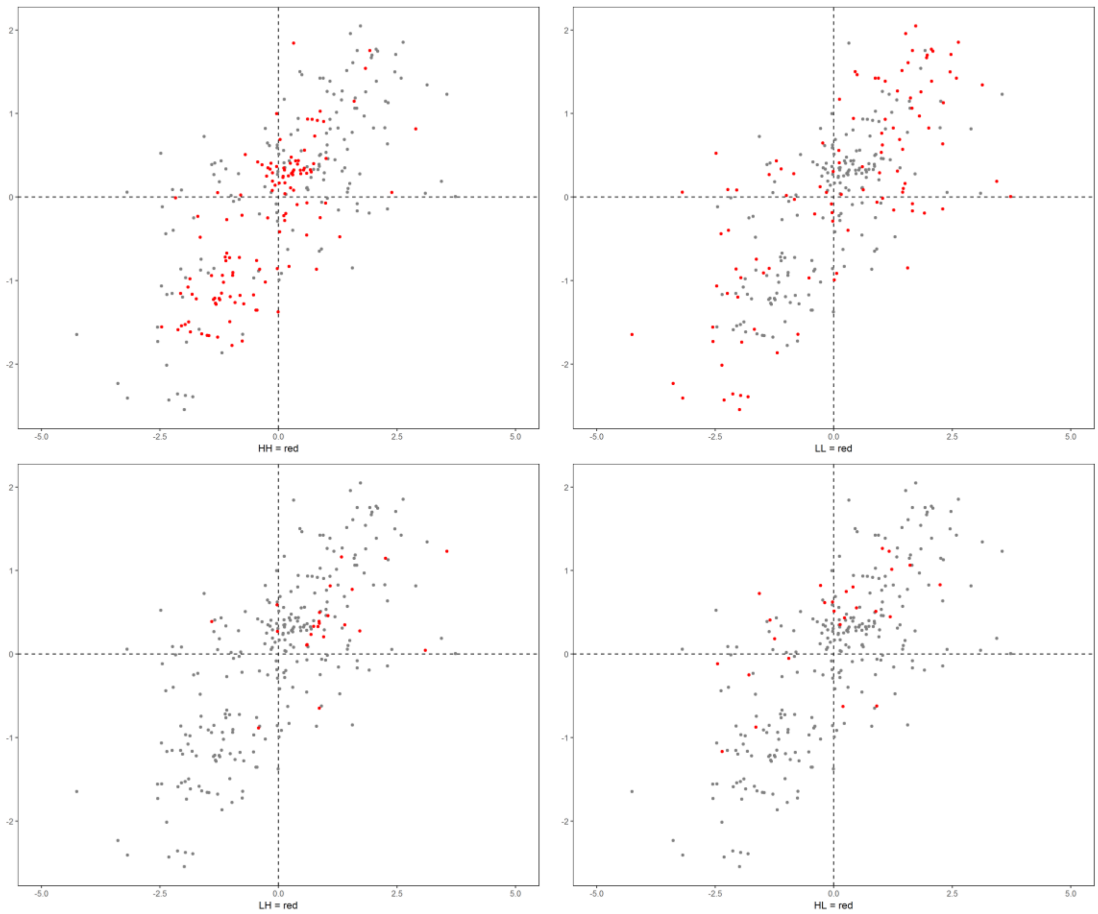


Figure 4.8: Movements between 2000-2015, ECO-DEV



Focusing on quadrants 1 and 3, where the vast majority of the regions are located, there are differences between the pictures provided by GDP p.c. (see Figures 4.9 and 4.10) and ECO-DEV. Regarding these quadrants, regions' movements are more dispersed with ECO-DEV compared to GDP p.c. Moreover, for the first quadrant and with the GDP p.c., the economic position of most regions in HH in 2000 and of their neighbors worsens during the period of analysis while with ECO-DEV their situation is more balanced as seen above. All in all, the (limited) convergence pattern detected above with ECO-DEV is more pronounced when the focus is on the GDP variable, in line with the results obtained with the global analysis. This would mean that ECO-DEV adjusts slower over time compared to GDP p.c.

Figure 4.9: Standardized Moran scatterplots for 2000-2015, GDP per capita

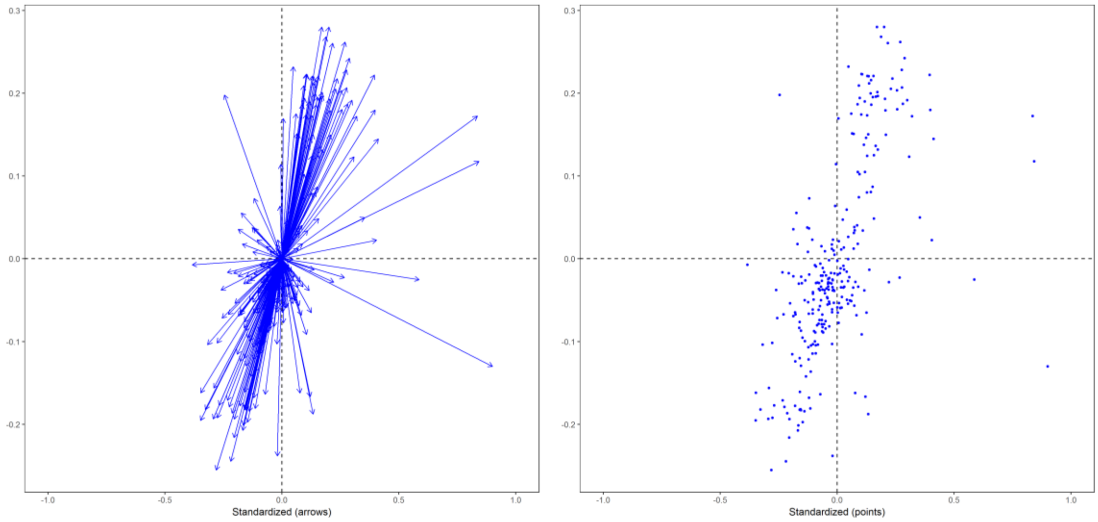
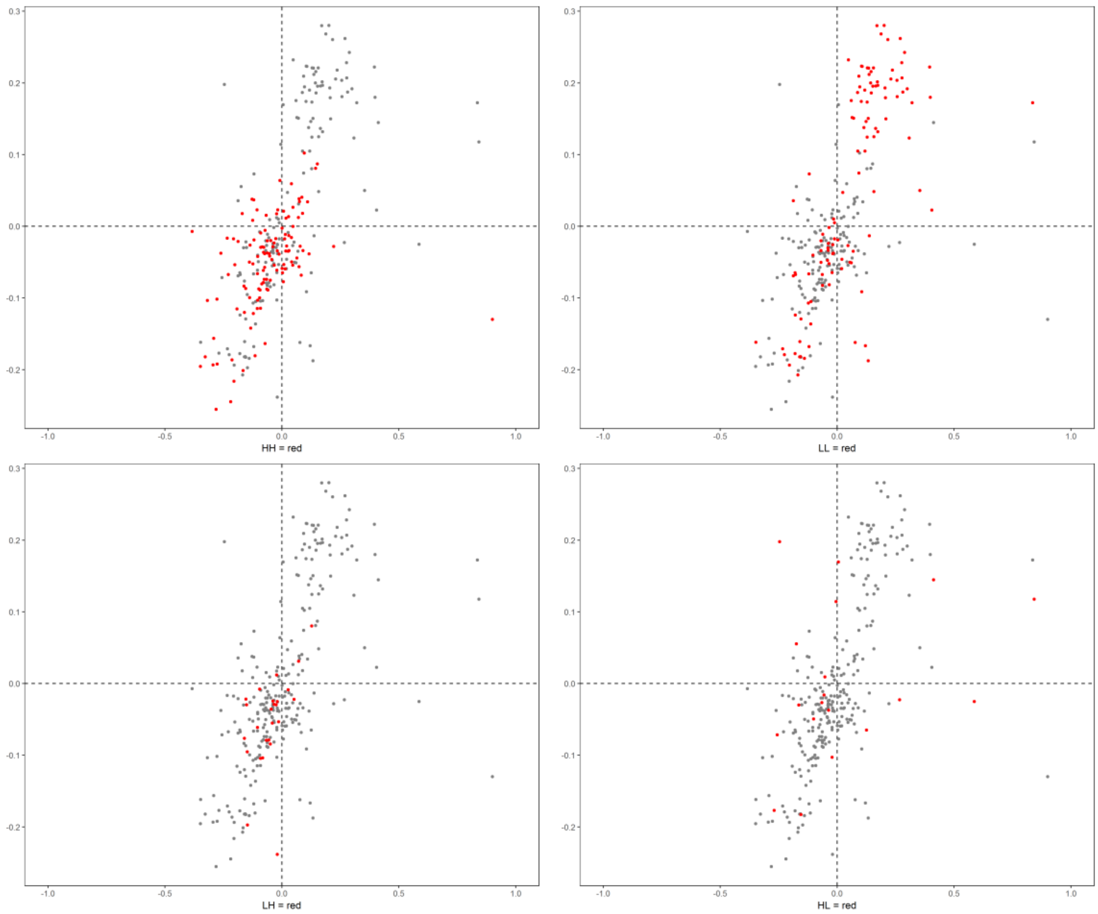


Figure 4.10: Movements between 2000-2015, GDP per capita



Next, from the standardized directional Moran scatter plot, we construct a rose diagram to gain additional insights on the regions' economic dynamics. The rose diagram reports the frequency of moves across different directions. We use the eight-class rose diagram, depicted in the left panel of Figure 4.11. One can observe that the predominant direction involves upward moves of regions and their neighbors in the ECO-DEV distribution (122 regions). In this block, the number of regions, the situation of which improves more their neighbors (79) is more important than the opposite (43). The second most important direction represents downward moves (90 regions). In this block, the number of regions, the situation of which worsens more than the one of neighbors (58) is less important than the opposite (32). Besides these two important categories of moves, one can find atypical movements involving opposite trajectories for a region and its neighbors (63 regions). The visual dominance of the upward moves suggests an asymmetric convergence pattern. We also analyze whether the pattern provided by the rose diagram is different from what would be expected if values of ECO-DEV were randomly distributed in the European space. The results are reported in the right panel of Figure 4.11. It appears that the movements of the regions and their neighbors are different from random movements at 5% for quadrants with important number of movements, with the exception of the movements corresponding to the case where the situation of regions deteriorates more than that of their neighbors. The GDP p.c. gives a different picture (see Figure 4.12): the position of the vast majority of regions worsens during the period.

Figure 4.11: Rose diagram and p-values for k = 8, ECO-DEV

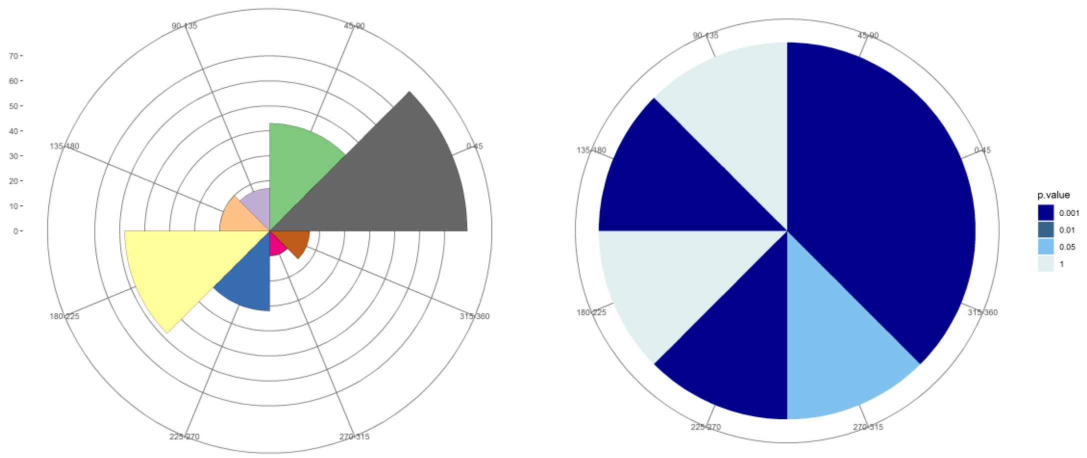
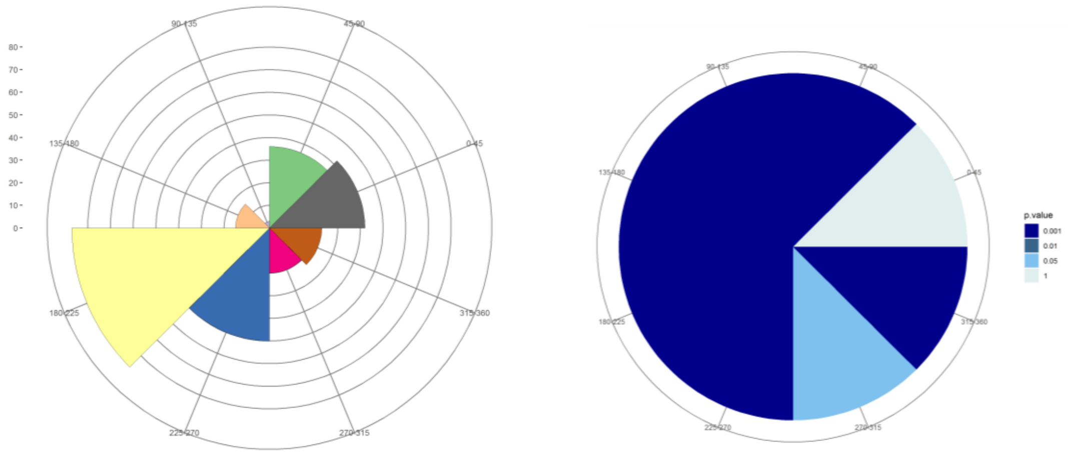


Figure 4.12: Rose diagram and p-values for k = 8, GDP per capita



For identification purposes, we plot the regions displaying the direction of moves in Figure 4.13. For ease of reading, we use the same colors as in the eight class rose diagram. First, we observe that regions from Poland, Romania, Bulgaria, Sweden, Germany, Austria and France improved their positions from 2000 to 2015. Second, some Italian, UK and Greek, Croatian and Dutch regions worsened their position from 2000 to 2015. The picture provided by GDP p.c. is completely different (see Figure 4.14). Indeed, with this variable, only few regions from Poland, Romania, Bulgaria, improved their positions. Moreover, most French regions along with regions from Germany, Austria, eastern European countries, UK, Italy, Greece and Sweden, Denmark and Finland worsened their positions during the period 2000-2015. This result implies that the assessment of regional economic performance provided by GDP p.c. for French regions for instance is stricter than the one provided by ECO-DEV.

Figure 4.13: Identification of regions from the 8-class rose diagram, ECO-DEV

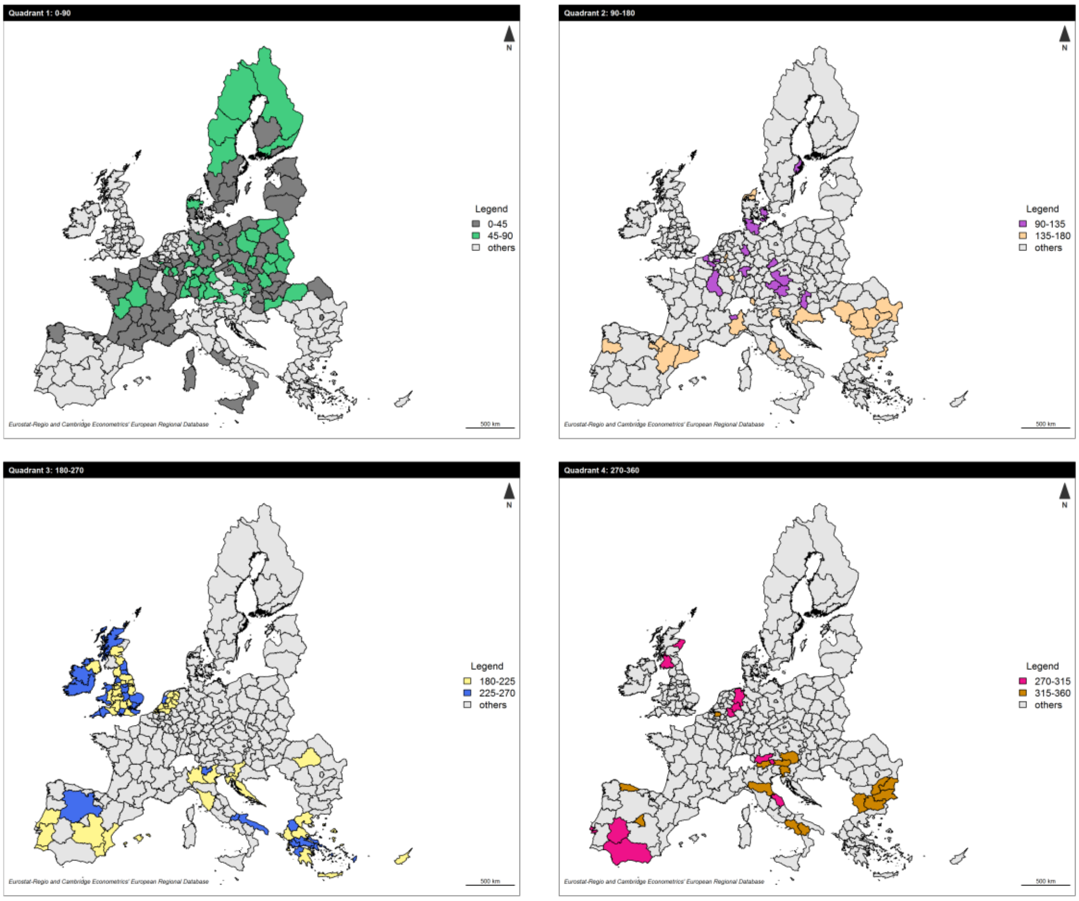
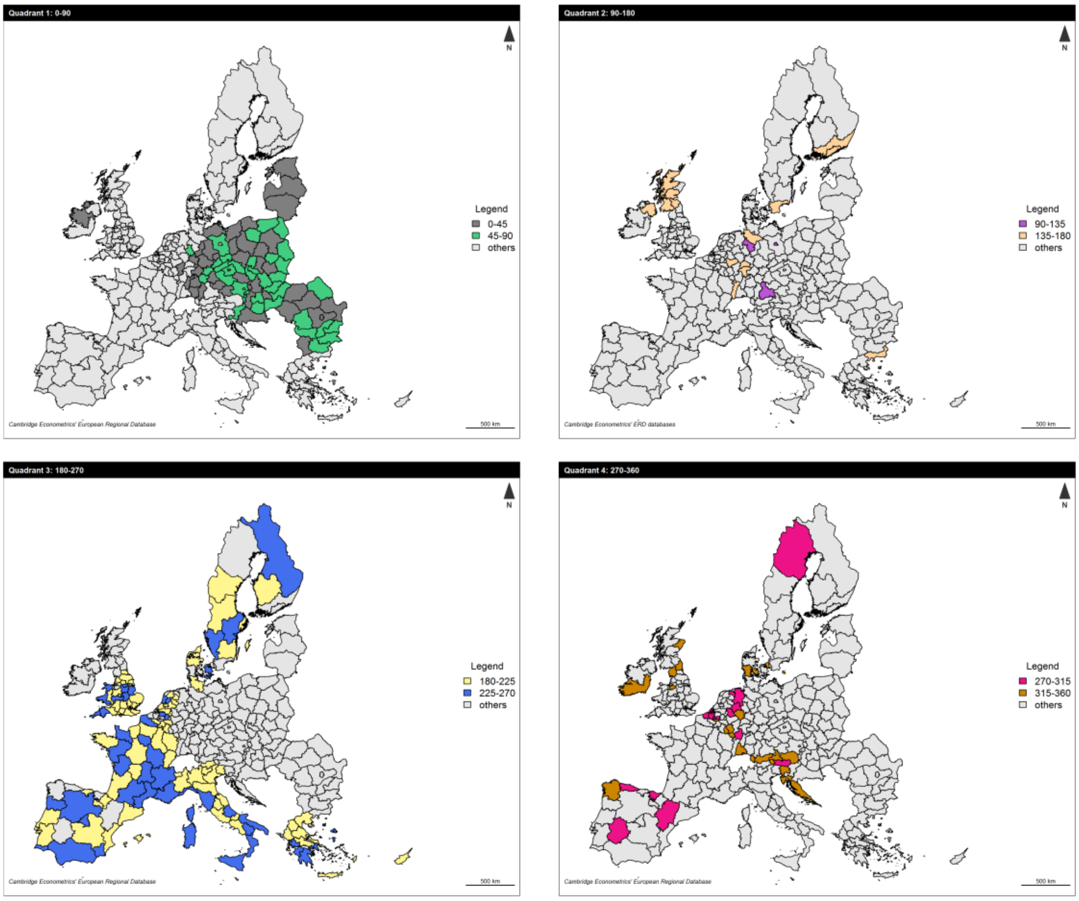


Figure 4.14: Identification of regions from the 8-class rose diagram, GDP per capita



We finish the analysis with an explicit dynamic consideration using Markov LISA by investigating the dynamics of spatial dependence over the study period (from 2000 to 2015). Specifically, we draw as many Moran scatterplots as there are time periods and define a move as a movement across one of the four quadrants of the Moran scatterplot. From this, we define a discrete LISA Markov chain where the states of the chain are the four quadrants of the scatterplot in a given period. Between any two time periods, the region position in the scatterplot may change. Collecting all these transitions enables the estimation of the Markov transition probabilities reported in Table 4.3. The chain has been estimated using maximum likelihood. The examination of these probabilities reveals several interesting characteristics about the spatial dynamics of ECO-DEV. First, the staying probabilities, i.e. the probability of remaining in one state between two time periods, are highest for quadrants 3 (LL) and 1 (HH) of the Moran scatterplot, followed by quadrants 2 (LH) and 4 (HL). Compared to regions in HH and LL, those in HL and LH are more likely to cross the scatterplot quadrants. Second, considering a region in the initial state LH, the movement to HH, which involves a change for the region but not its spatial lag, occurs more frequently than movement to LL, which involves a change in the position of the spatial lag but not the focal region. Similarly, for regions in HL in the initial state, moves to HH are more frequent than moves to LL. This confirms the moderated convergence pattern detected above. The relative mobility in this Markov LISA transition matrix[11] is

[11] This statistic (δ) is calculated as follows (see [Rey(2001)] for more details): $\delta = (k - \sum_i P_{ii})/(k-1)$, where P_{ii} is the diagonal element of the LISA Markov transition matrix P and k the number of total classes. With no inter-class transitions, $\delta = 0$, and the more the inter-class mobility, the larger δ. The maximum value is $k/(k-1)$.

relatively small (0.1652), confirming the persistence of spatial dependence also highlighted by GIMA. The last row of Table 4.3 shows the ergodic probabilities which gives an indication on the long term probabilities in each class. The higher ergodic probabilities are associated with the HH and LL columns, meaning that, only a few LL regions and a lot of HH ones will exist in the long run.

Table 4.3: LISA Markov transition probabilities (4 classes), ECO-DEV

		End			
		HH	LL	LH	HL
Beginning	HH	0.9661	0.002	0.0174	0.0145
	LL	0.0039	0.9662	0.0163	0.0137
	LH	0.1391	0.0596	0.798	0.0033
	HL	0.1398	0.0824	0.0036	0.7742
	π	0.5777	0.2921	0.0744	0.0557

When we compare these results to the GDP p.c. ones and focus on quadrants 1 and 3 (where the majority of regions are concentrated)– see Table 4.4 – we can note that staying probabilities are almost the same for ECO-DEV and GDP p.c. This means that most regions in these quadrants, while improving or worsening their positions stay in their starting quadrant over time.

Table 4.4: LISA Markov transition probabilities (4 classes), GDP per capita

		End			
		HH	LL	LH	HL
Beginning	HH	0.9643	0.0038	0.02	0.0119
	LL	0.0017	0.9739	0.0186	0.0058
	LH	0.0374	0.065	0.8957	0.002
	HL	0.0386	0.0514	0	0.91
	π	0.273	0.5108	0.1433	0.0729

In order to account non significant LISA values, we move from the four-class to a five-class LISA Markov: an additional class is added corresponding to regions associated with non significant LISA statistics while the remaining classes corresponding to the 4 quadrants of the Moran scatterplot only including regions associated to a significant LISA statistics. The corresponding estimated probabilities are reported in Table 4.5 for ECO-DEV. The staying probabilities are again the highest for quadrants 3 (LL) and 1 (HH) of the Moran scatterplot and the non-significant case, followed by quadrants 4 (HL) and 2 (LH). As for the four-class LISA Markov, i)compared to regions in HH and LL, those in HL and LH are more likely to cross the scatter plot quadrants; ii) for a region in the initial state LH, the movement to HH, which involves a change for the region but not its spatial lag, occurs more frequently than movement to LL, which involves a change in the position of the spatial lag but not the focal region. For regions in HL in the initial state, moves to HH are less frequent than moves to LL (for the four-class case, we had the opposite observation). The convergence pattern is explained in part with the movements of regions in LH which mostly either stay in the same quadrant or move in quadrant 1 (improvement).

We set up a formal test for co-movement dependence, deriving the LISA chain into two marginal discrete chains: one for the focal unit and one for the the spatial lag chain (the neighbors). Each of these marginal chains has two states H or L, depending of their position relatively to the mean value of ECO-DEV in a given time period. The test of the difference of these two transitions matrices resulted in $\chi^2(9)$ = 4854.65 , p-value $<$ 0.0001. It indicates that the movement of regions' ECO-DEV in the distribution is dependent of the movement of the neighboring regions values.

Table 4.5: LISA Markov transition probabilities (5 classes), ECO-DEV

		End				
		HH	LL	LH	HL	non sign.
Beginning	HH	0.9581	0.0000	0.0054	0.0000	0.0365
	LL	0.0000	0.9749	0.0000	0.0029	0.0222
	LH	0.3600	0.0000	0.4000	0.0000	0.2400
	HL	0.0000	0.1053	0.0000	0.7368	0.1579
	non sign.	0.0301	0.0171	0.0040	0.0023	0.9465
	π	0.3397	0.2613	0.0057	0.0062	0.3872

When we compare these results to those obtained with GDP p.c. (see Table 4.6), we note, focusing on quadrants 1 (HH) and 3 (LL) where the vast majority of regions are concentrated, that the staying probability in LL is relatively stable while the one in HH is higher for ECO-DEV compared to GDP p.c. This again confirms the fact that regions are less integrated when assessed with ECO-DEV compared to GDP p.c.

Table 4.6: LISA Markov transition probabilities (5 classes), GDP per capita

		End				
		HH	LL	LH	HL	non sign.
Beginning	HH	0.8887	0	0.0122	0	0.0991
	LL	0	0.9611	0	0.0054	0.0335
	LH	0.036	0	0.8849	0	0.0791
	HL	0	0.0282	0	0.8873	0.0845
	non sign.	0.0249	0.0147	0.0037	0.0012	0.9555
	π	0.1412	0.2283	0.0336	0.0171	0.5799

To summarize the results obtained in this section, we show that ECO-DEV is significantly spatially clustered in Europe. Indeed, global Moran's I and global GIMA τ_W reflect the existence of a significant, positive and persistent spatial dependence in the distribution of ECO-DEV over years. Thanks to local Moran statistics, we find that most significant clusters are either high-high (hot-spot) or low-low (cold-spot). Globally, high values of ECO-DEV are concentrated in the center of Europe, UK and Scandinavian regions. Backwards regions are concentrated in Eastern European countries and southern regions of Italy and Spain. We then moved to the directional LISA and Markov LISA to further analyze the dynamics of ECO-DEV spatial distribution. Using the former method, we highlight that the predominant direction involves upward moves of regions and their neighbors in ECO-DEV distribution. This is, most regions, along with their neighbors improve their relative position from 2000 to 2015. The second most important direction represents downward moves. The closer examination of these movements, with respect to the quadrant of origin, suggests that there is a convergence process at work in the EU-28 regions from 2000 to 2015. From the Markov LISA analysis, we show that the staying probabilities are relatively high for all quadrant but LH and HL.

Since most regions are either in HH or LL, this would mean that the convergence pattern detected above is somewhat moderated: most regions which improves/worsens their situation, along with their neighbors remains in their quadrant of origin.

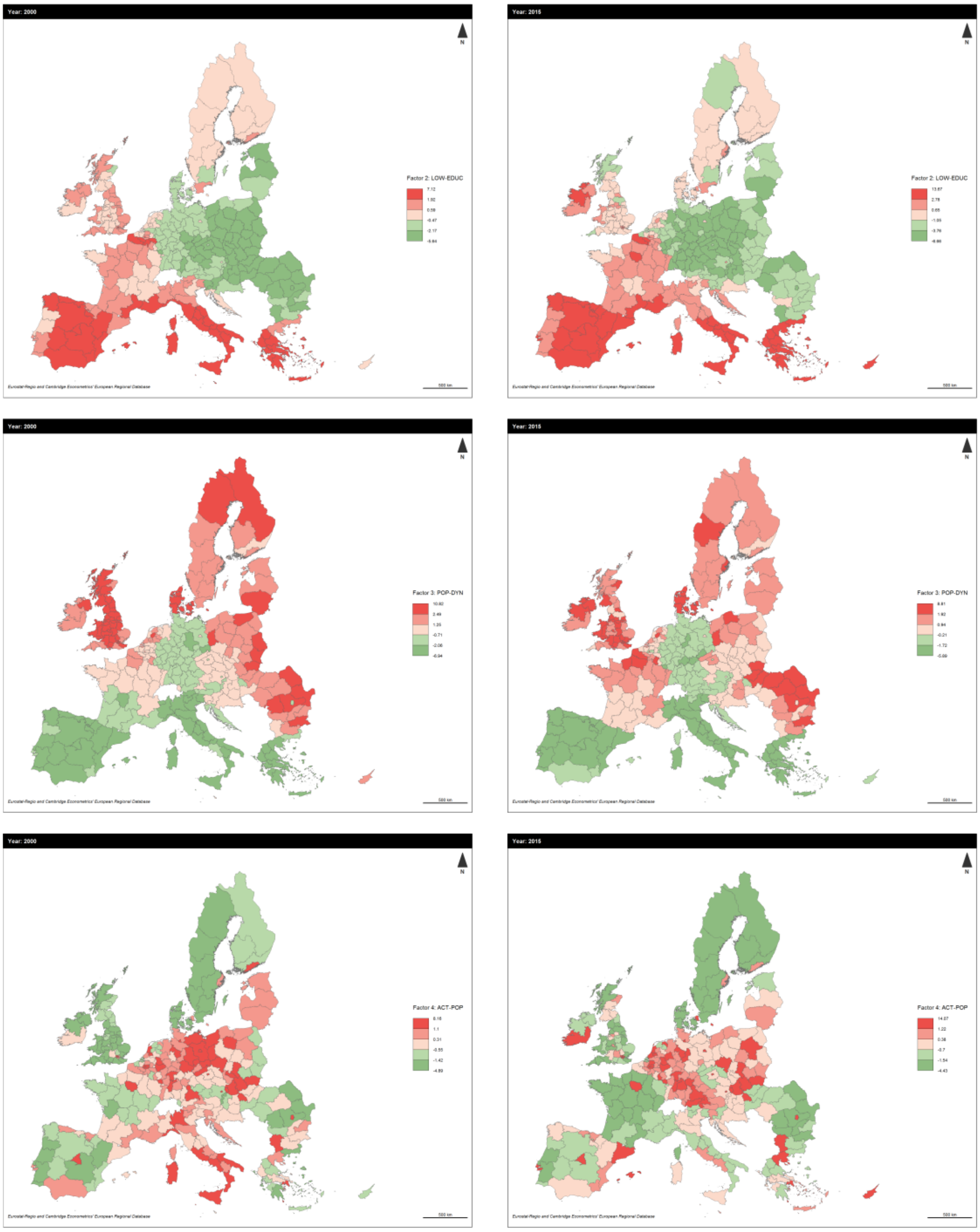
Also, we systematically contrast results obtained with those obtained with GDP p.c. With local Moran statistics, we observe that the number of clusters of similar values identified is more important with ECO-DEV compared to GDP p.c. This observation, in conjunction with the higher values of global Moran's I and the results of GIMA τ_W, shows that the clustering of EU-28 regions within blocks of rich and poor is more pronounced when considering ECO-DEV instead of GDP p.c. Therefore, the magnitude of economic integration shown by GDP p.c. should be considered with caution. Regarding the dynamics, one can note that the convergence pattern detected with ECO-DEV is less pronounced with the GDP variable. This would mean that ECO-DEV adjusts slower over time compared to GDP p.c. In other words, some of the original variables contributing to ECO-DEV (from the MFA) must be somewhat rigid. Indeed, for countries like France, the dynamics of regional GDP p.c. is negative while with ECO-DEV, these regions are doing well. That would mean that variables like female employment rate or young people neither in employment nor in education and training are limiting the effect of GDP p.c. fall and are acting as economic stabilizers. Conversely, some regions from Eastern countries are doing well with the GDP p.c. and not with ECO-DEV. As explained above, this would mean that the GDP variables is less rigid than the others in ECO-DEV. It could also mean that there are some intra-NUTS-2 GDP p.c. disparities within these poor regions. At any rate, the results clearly highlight the fact that in some NUTS-2 regions, the GDP p.c. is a poor indicator of the economic well-being.

5. Complementary analysis: what about the other factors?

We briefly present in this section the results obtained for the remaining three factors: *low education* (LOW-EDUC), *population dynamics* (POP-DYN) and *active population*(ACT-POP) from the MFA.

Visually, the choropleth maps of these variables suggest spatial association of similar values, more pronounced for LOW-EDUC and POP-DYN compared to ACT-POP (see Figure 5.1). Regarding the variable LOW-EDUC, one can identify in 2000 a group of regions with a high percentage of active people with a pre-primary, primary and lower secondary education, and a low percentage of active population with upper secondary and post-secondary education belonging mainly to Portugal, France, Spain, Italy and Greece. Regions with exact opposite characteristics are located in the core center of Europe and in countries of the former Eastern bloc. These observations are globally in line with Rodriguez-Pose and Tselios (2011) who found that Portuguese, followed by the Spanish, French, Italian and British are the least educated in Western Europe in 2015, whereas Denmark and Sweden have the highest percentage of people with secondary education. For the factor POP-DYN, we have in 2015 a group of regions with a high percentage of retired people and a low percentage of children in Spain, Portugal, Italy, Greece, Germany and in the south-west of France. Conversely, regions with a high proportion of children, along with a low proportion of retired are clustered mainly in UK, countries of the former Eastern bloc and in Scandinavian countries. The last factor (ACT-POP), as mentioned above, is significantly less clustered compared to factors 2 and 3. Also, spatial patterns detected with ACT-POP are relatively less persistent than the ones detected with LOW-EDUC and POP-DYN.

Figure 5.1: Spatial patterns of factors 2 to 4 for 2000 and 2015 (quintile classification)



These observations are confirmed by the estimation of global Moran's I statistic, which is respectively equal to 31.05, 29.46 and 10.37 for LOW-EDUC, POP-DYN and ACT-POP in 2000 and 29.96, 27.20 and 8.55 in 2015 (standardized values, see Table 5.1). Focusing on the dynamics of estimated Moran $I's$ among years, one can note a feature common to the three factors: the evolution of Moran $I's$ can be characterized by three sub-periods. Indeed, while during the first sub-period (between 2000 and 2003 for LOW-EDUC and ACT-POP, 2000 and 2002 for POP-DYN), no discernible trend in the evolution of LOW-EDUC, POP-DYN and ACT-POP Moran I is detected, one can observe a continuous decrease in the second sub-period (from 2004 to 2008 for LOW-EDUC, 2003 to 2007 for POP-DYN and 2004 to 2010 for ACT-POP) in the standardized value of the statistic, for the three factors. During the last period, Moran $I's$ globally continuously increases, without reaching the 2000 levels.

Table 5.1: Moran's I statistic for factors 2-4

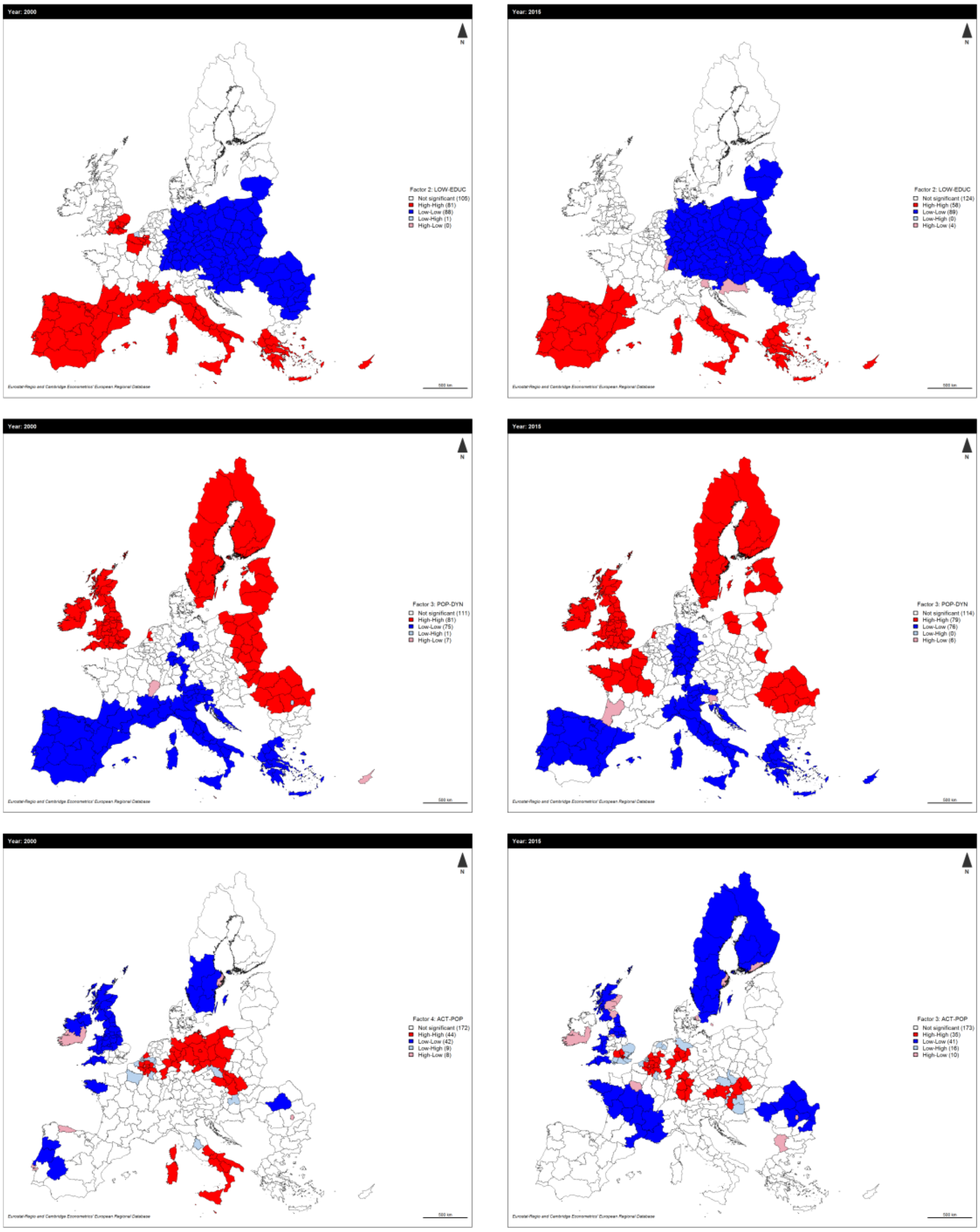
	LOW-EDUC		POP-DYN		ACT-POP	
Year	Moran's I	s. value	Moran's I	s. value	Moran's I	s. value
2000	0.7675	31.056	0.7275	29.4620	0.2529	10.3732
2001	0.7624	30.8538	0.6772	27.4444	0.2547	10.4368
2002	0.7782	31.4778	0.6584	26.6918	0.2467	10.1430
2003	0.7804	31.5700	0.6817	27.6346	0.3002	12.2845
2004	0.7867	31.8230	0.6618	26.8376	0.3014	12.3294
2005	0.7703	31.1729	0.5988	24.3299	0.2958	12.1140
2006	0.7242	29.3542	0.5493	22.3561	0.2641	10.8473
2007	0.6901	27.9883	0.5254	21.3878	0.2067	8.5592
2008	0.6595	26.7771	0.5443	22.1444	0.1879	7.8180
2009	0.6716	27.2575	0.5502	22.3694	0.1717	7.1817
2010	0.6763	27.4528	0.5514	22.4033	0.1537	6.4545
2011	0.7134	28.9189	0.5512	22.4068	0.1667	6.9940
2012	0.7373	29.8726	0.5959	24.1853	0.1887	7.8879
2013	0.7587	30.7118	0.6413	25.9775	0.2086	8.6817
2014	0.7537	30.5109	0.6662	26.9869	0.2148	8.9341
2015	0.7397	29.9669	0.6715	27.2051	0.2056	8.5530

Note: all statistics are significant at p = 0.00001; s. value = standardized value

Figure 5.2 displays the significant LISA statistics for the three factors. With respect to LOW-EDUC, two findings can be emphasized. First, at the beginning of the period, we detect one big cluster of low-low values composed of regions from Germany, Austria and from countries of the former Eastern bloc. That is, these regions exhibit a low rate of active population with pre-primary, primary and lower secondary education, and high rate of active population with upper secondary and post-secondary levels. This cluster is highly persistent over time. Second, we observe four clusters of high values: the biggest consists of regions of Spain, Portugal, along with southern regions of France and Italy. The three remaining ones (of moderate size) are composed of regions from Greece, regions from the north of France and regions from the south of UK. These clusters are also persistent over time, even if when we move to 2015, the north of France and the south of UK clusters vanishes and most regions in the south or France are no longer in the high-high cluster. Regarding POP-DYN, local Moran statistics identify three clusters of high values in 2000. The first one includes regions from UK while the second and third are composed respectively of regions from Sweden and Finland and from regions from the former Eastern bloc. Recall that a region with a high score on this variable is probably young and dynamic. Therefore these clustered regions are the youngest and dynamic paces in EU-28. These clusters are also persistent over time. One can note however, when we move to 2015, the creation of an extra cluster with regions from the north of France on the one hand and on the other hand, that most regions from Poland, previously in the high-high cluster are no longer clustering. Beside these high-high clusters, we identify in 2000 two clusters of low-low regions. The first one consist of regions from Portugal, Spain, southern regions of France, Italy, Croatia and some regions from the core center of Europe (Austria and Germany). The second cluster is composed with regions from Greece. These regions are characterized by a high proportion of people aged more than

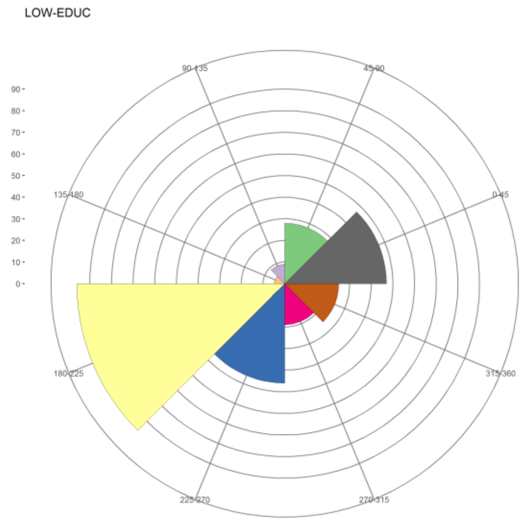
50. Note to finish that when moving to 2015, one can observe that regions from southern regions of France are no longer in this low-low cluster on the one hand and that on the other hand, the first low-low cluster described above expands toward the center of Europe. The last factor, ACT-POP is much less concentrated compared to the first two. This explains the relatively low value of Moran's I statistics for this variable. Moreover, unlike the previous variables analyzed, the clusters detected in 2000 for ACT-POP are less persistent over time. For example, one can observe a cluster of low values composed of regions in the west of France in 2015, not present in 2000.

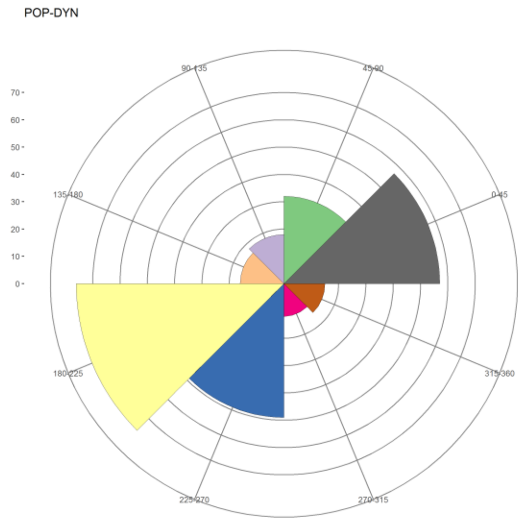
Figure 5.2: Local Moran clusters for 2000 and 2015 (factors 2 to 4) – p-value with Bonferroni adjustment

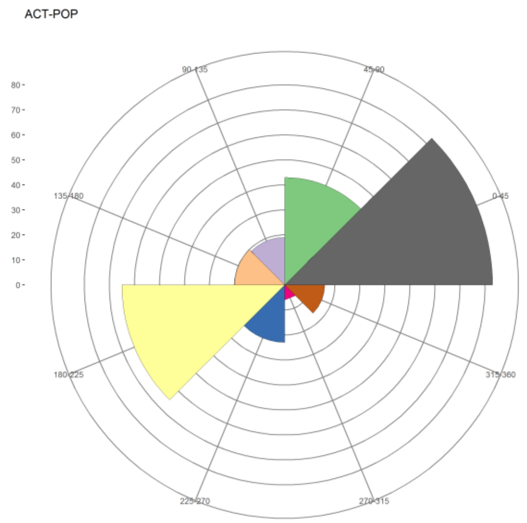


We then implement the directional LISA methodology to further investigate spatial dynamics within the distribution of LOW-EDUC, POP-DYN and ACT-POP. The corresponding Rose diagrams are displayed in Figure 5.3. It appears that for LOW-EDUC and POP-DYN, the predominant direction of moves between 2000 and 2015 involves downward moves of regions and their neighbors, followed by upward moves while for ACT-POP we have the opposite observation. Regarding LOW-EDUC, this would mean that most regions, along with their neighbors' rate of active population with pre-primary, primary and lower secondary education, is decreasing, along with an increase in the rate of active population with upper secondary and post-secondary levels. That is, people of most European regions are getting more educated over time. For POP-DYN, direction of moves suggests that for the majority of regions, the percentage of children is decreasing while the percentage of retired people is increasing over the period. This trend is however followed by a second category of regions exhibiting the opposite picture. Finally, in most regions population density along with the percentage of active population is increasing. This would mean that most regions are becoming more attractive and competitive.

Figure 5.3: Rose diagrams for k = 8 (factors 2 to 4)







As a complement to Rose diagrams, we plot regions displaying the direction of moves identified for each of our three variables. With respect to LOW-EDUC (see Figure 5.4), the majority of regions from UK, core central Europe and from the former Eastern bloc are getting more educated over time. Note also that some regions in France, Spain, Portugal, Italy and Greece follow the exact opposite path. Regarding the variable POP-DYN (see Figure 5.5), one can observe that most regions from France, Portugal, Spain, Italy and Croatia are becoming younger and conversely, regions from Scandinavia, Greece and from the former Eastern bloc are becoming aged. Finally, from the observation of ACT-POP (see Figure 5.6), we observe that most regions from Spain, Portugal, UK and the those located at the center of Europe are increasing their attractiveness over time. Conversely, most regions from

France, Italy, Scandinavian countries, north of Germany and west of Poland are losing grounds on the competitiveness race. The results of the inference (see Figure 5.7) highlight that the movement of regions and their neighbors is different from a random one at 5% for quadrants 1 and 3 (where are concentrated the majority of movements), with one exception however. Indeed, for LOW-EDUC, the movements corresponding to the situation where both the region and its neighbors increase but more so for the region itself is not significantly different from a random movement over time. As the majority of movements are concentrated in quadrant 1 and 3, globally the conclusions made above from the observation of Figure 5.4 remain valid.

Figure 5.4: Identification of regions from the 8-class rose diagram (factor 2)

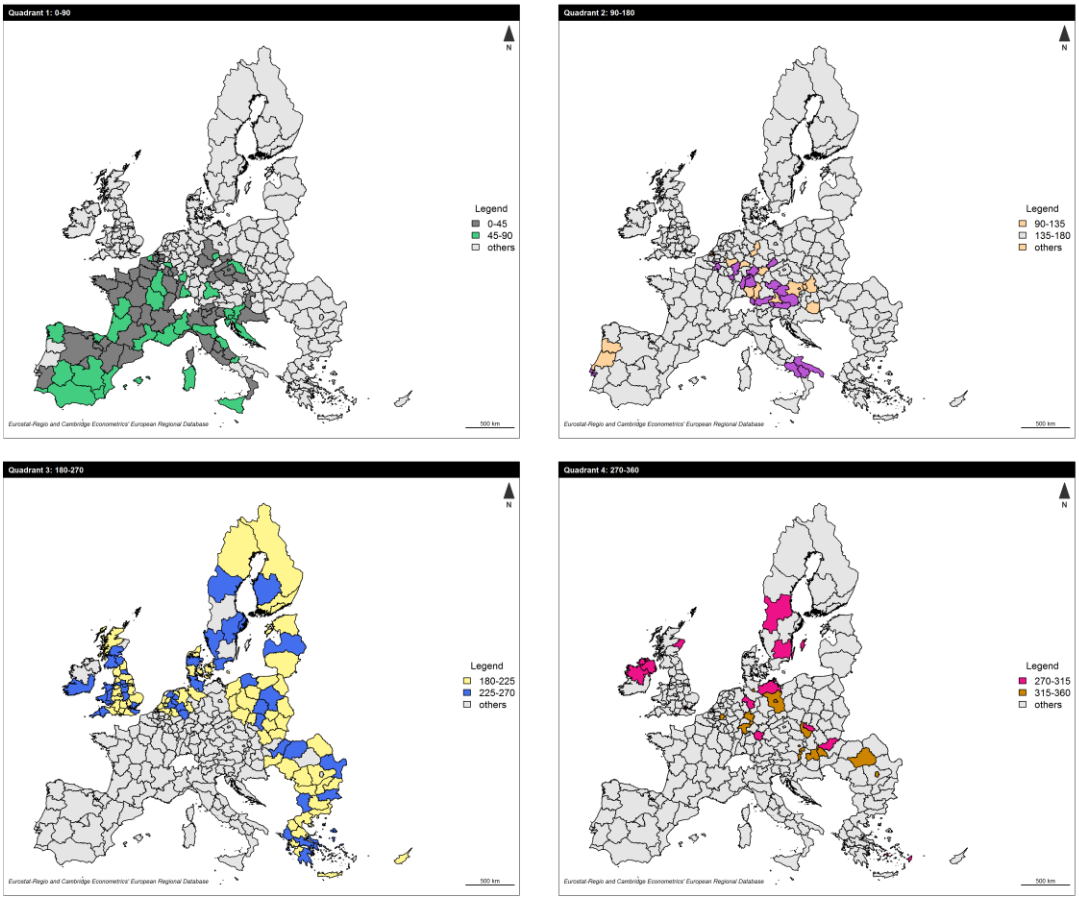


Figure 5.5: Identification of regions from the 8-class rose diagram (factor 3)

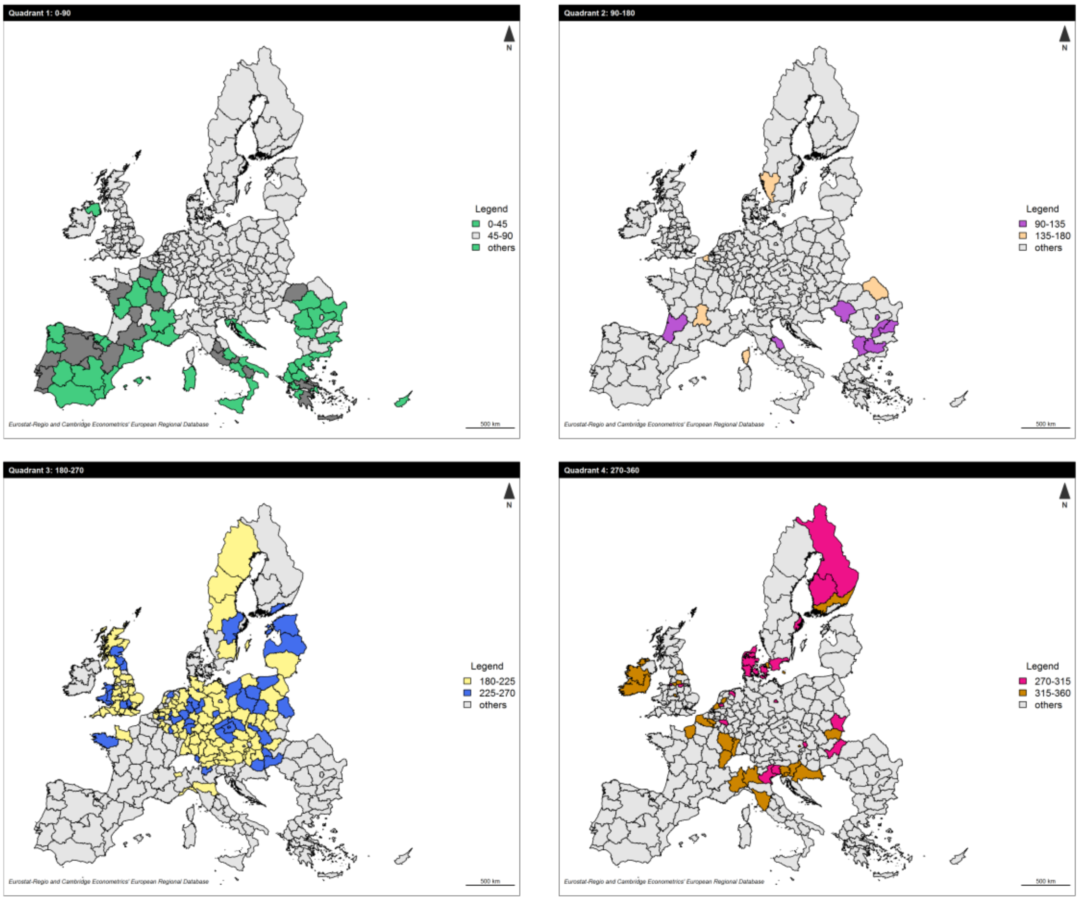


Figure 5.6: Identification of regions from the 8-class rose diagram (factor 4)

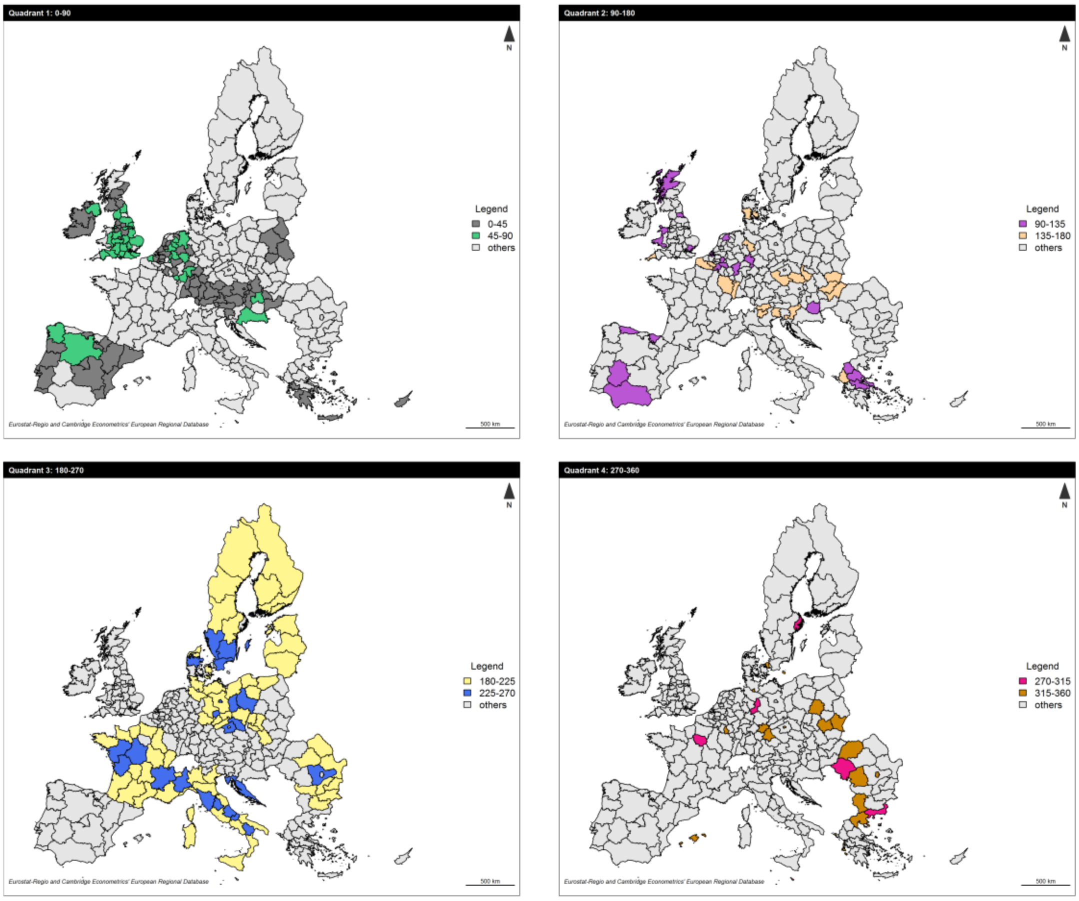
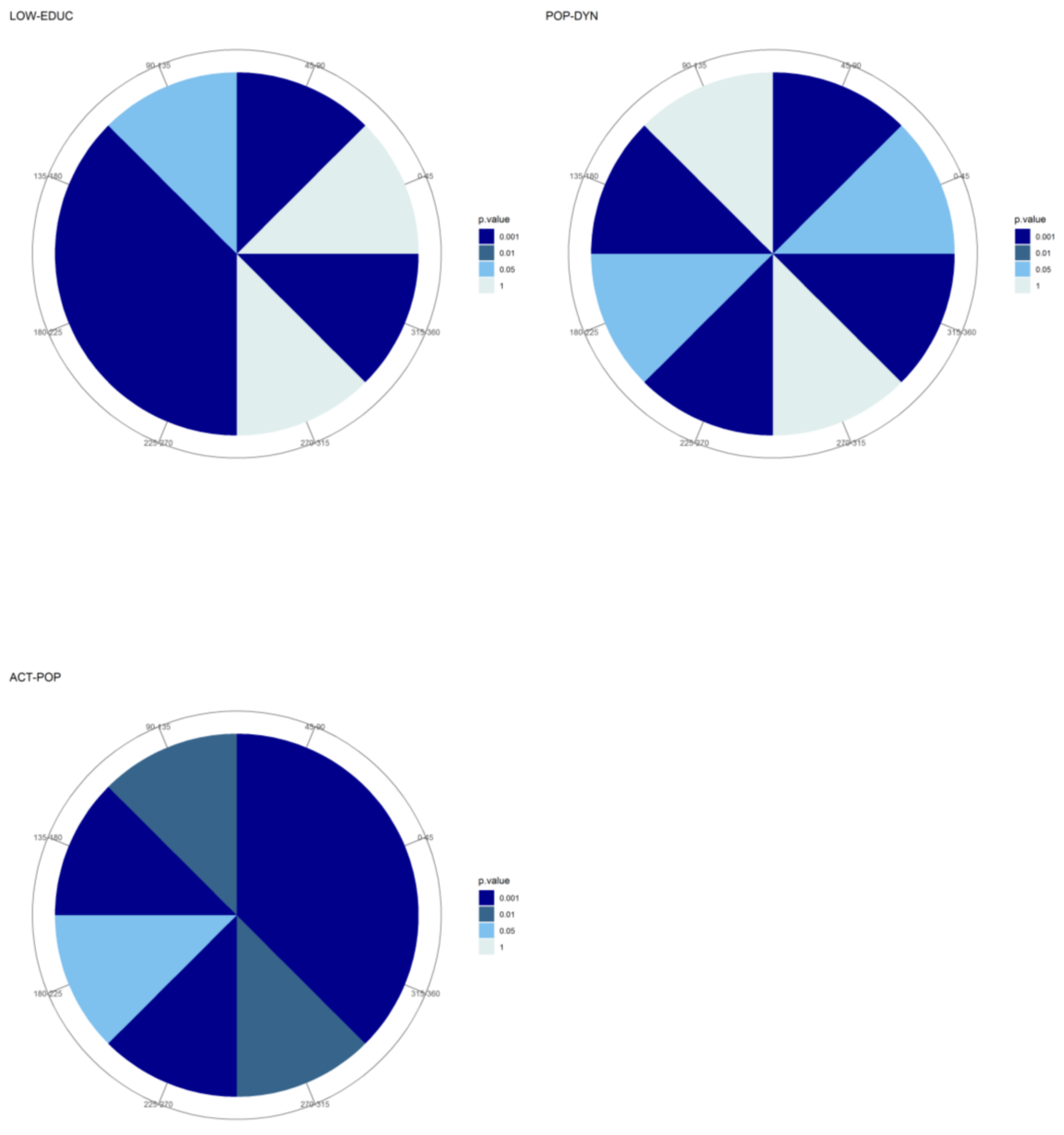


Figure 5.7: p-value diagrams for k = 8 (factors 2 to 4) - p-value = 0.05



We finally estimate LISA Markov chains over the study period (from 2000 to 2015). The obtained results are reported in Table 5.2. As for ECO-DEV, the staying probabilities are relatively important. They are the highest for quadrant 1 (HH) and 3 (LL). Regions in quadrants 2 (LH) and 4 (HL) are less stable than those in quadrants 1 and 3 and are thus more likely to cross the scatterplot quadrants. One can note also that the relative mobility on Markov transition matrix is quite stable amongst factors 2 to 4 (0.1592, 0.1510 and 0.1350 respectively for LOW-EDUC, POP-DYN and ACT-POP). That is, the lower staying probabilities in quadrant 1 (HH) and 3 (LL) for ACT-POP compared to LOW-EDUC and POP-DYN are almost offset by its higher staying probabilities observed in quadrant 2 (LH) and 4 (HL), in comparison to those observed for LOW-EDUC, and POP-DYN.

Table 5.2: LISA Markov transition probabilities (4 classes) for factors 2-4

			End			
			HH	LL	LH	HL
Factor 2	**Beginning**	HH	0.9627	0.0065	0.0244	0.0065
		LL	0.0023	0.9764	0.0063	0.015
		LH	0.133	0.0851	0.7819	0
		HL	0.1075	0.0914	0	0.8011
		π	0.4018	0.4892	0.0593	0.0496

			End			
			HH	LL	LH	HL
Factor 3	**Beginning**	HH	0.9733	0.0011	0.0165	0.0091
		LL	0.0029	0.9608	0.0164	0.0199
		LH	0.1263	0.0877	0.786	0
		HL	0.0669	0.1024	0.0039	0.8268
		π	0.5399	0.3267	0.0679	0.0656

			End			
			HH	LL	LH	HL
Factor 4	**Beginning**	HH	0.922	0.0045	0.0503	0.0233
		LL	0.004	0.944	0.0273	0.0247
		LH	0.0856	0.0548	0.8596	0
		HL	0.0552	0.0737	0.0018	0.8692
		π	0.3136	0.3735	0.1867	0.1262

The five-class LISA Markov are in Table 5.3. Even if the estimated probabilities are lower compared to the ones from the four-class method, they remain relatively important. Also, as seen in the last section, compared to regions in HH and LL, those in HL and LH are more likely to cross the scatter plot quadrants. Two additional observations can be made. First, for regions in the initial state HL, their probability to move to HH decreases and is even null and their probability to move to LL increases. Second, for regions in the initial state LH, their probability to move to HH increases and their probability to move to LL decreases and is even null.

Table 5.3: LISA Markov transition probabilities (5 classes) for factors 2-4

			End				
			HH	LL	LH	HL	non sign.
Factor 2	**Beginning**	HH	0.9379	0.0000	0.0010	0.0000	0.0612
		LL	0.0000	0.9667	0.0000	0.0068	0.0265
		LH	0.2222	0.0000	0.5556	0.0000	0.2222
		HL	0.0000	0.2059	0.0000	0.7353	0.0588
		non sign.	0.0245	0.0222	0.0012	0.0023	0.9498
		π	0.1833	0.3614	0.0016	0.0131	0.4405

			End				
			HH	LL	LH	HL	non sign.
Factor 3	**Beginning**	HH	0.9403	0.0000	0.0026	0.0000	0.0571
		LL	0.0000	0.9444	0.0000	0.0121	0.0435
		LH	0.2174	0.0000	0.7391	0.0000	0.0435
		HL	0.0000	0.1000	0.0000	0.8000	0.1000
		non sign.	0.037	0.0395	0.0019	0.0038	0.9179
		π	0.2613	0.3199	0.0054	0.0266	0.3869

			End				
			HH	LL	LH	HL	non sign.
Factor 4	**Beginning**	HH	0.8505	0.0000	0.0309	0.0000	0.1186
		LL	0.0000	0.9039	0.0000	0.0142	0.0819
		LH	0.0829	0.0000	0.8244	0.0000	0.0927
		HL	0.0000	0.014	0.0000	0.8392	0.1469
		non sign.	0.0238	0.0227	0.0102	0.0063	0.9371
		π	0.1316	0.1517	0.059	0.0375	0.6202

6. ESDA on microdata

This section performs an ESDA on the data provided by the report D2.3 of Work Package 2 (WP2) of the IMAJINE Project. These data are described in report D2.2 'Literature Review on Disaggregation Methodologies'. The study area is depicted in Figure 6.1. There are 21,781 observations distributed in 10 countries.

Figure 6.2. displays the distribution of average household income in 2011 using the Jenks algorithm to define the classes. Note that we did some adjustments for outliers, the values of which were exceedingly high without justification given their economic situation (some units in the center and East of France, in Spain, in Belgium, in Germany and the Netherlands). In these cases (30 in total), the values of revenue per capita were replaced by the mean of their neighbors (defined as first-order contiguity). Strong spatial heterogeneity is evident from this figure between and within countries. Clusters of rich units are apparent in South of UK, Germany, Netherlands, East of France, and capital regions. The Italian dualism is quite apparent as well together with the lagging situation of Portuguese regions, most Spanish regions and especially Romanian regions.

Figure 6.1: Study area

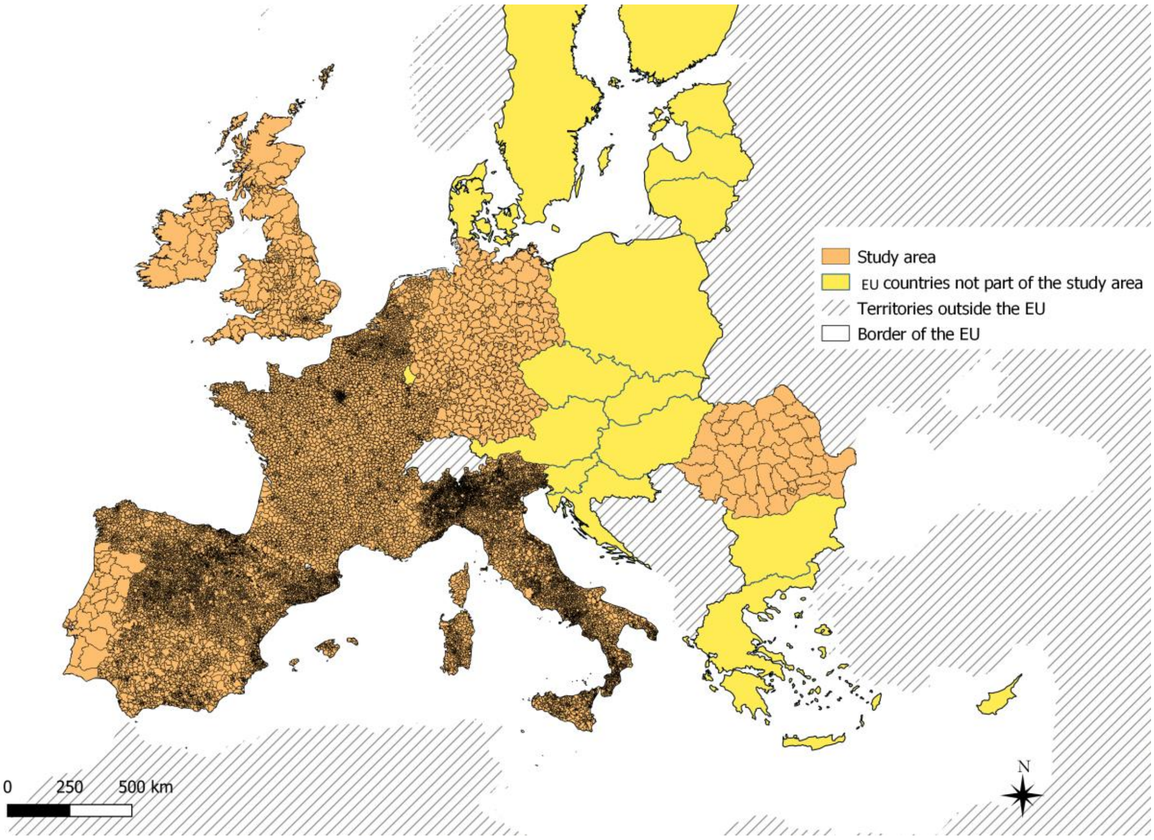


Figure 6.2: Distribution of average household income, 2011

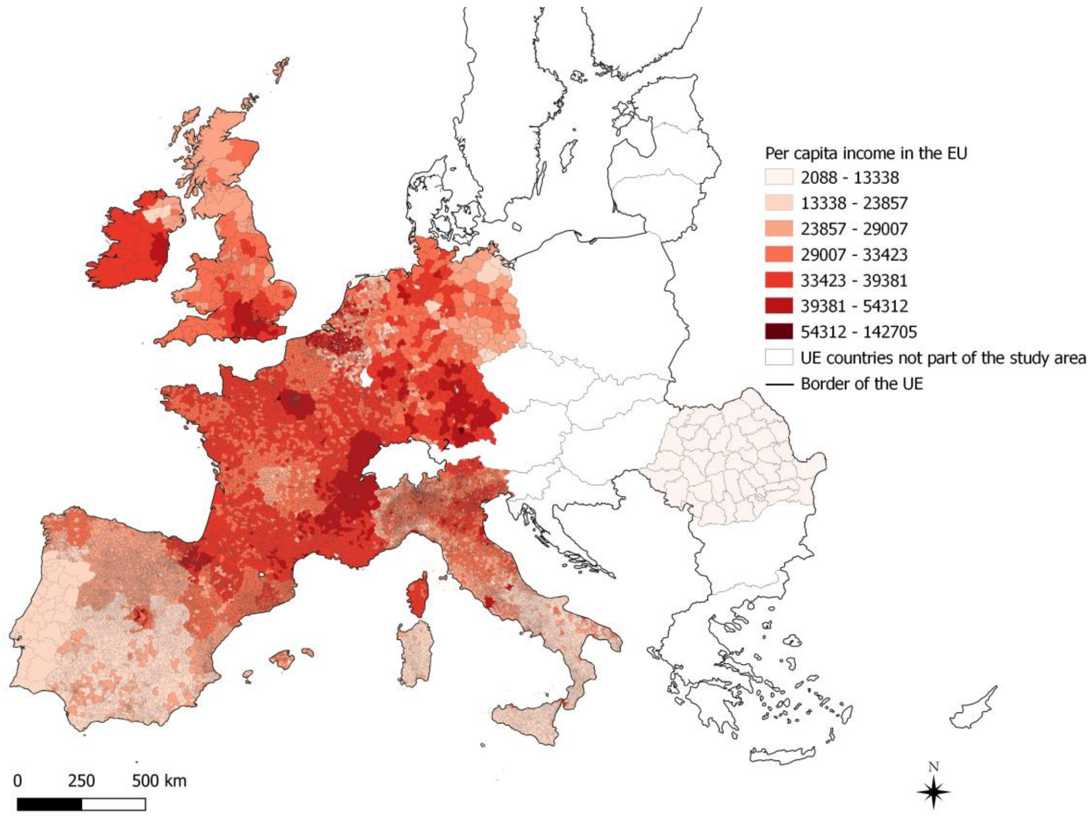


Table 6.1. further illustrates these disparities by displaying the mean income per capita by country in the EU in 2011.

Table 6.1: Mean income per capita by country in the EU in 2011

Belgium	France	Ireland	Italy	Germany	Netherlands	Portugal	Romania	Spain	United Kingdom	European average
35,001	35,799	38,807	27,628	33,312	29,248	16,920	3,117	26,956	33,305	29,127

We now investigate in more depth the pattern of spatial clustering in the distribution of mean income per capita in 2011. For that purpose, we first define our weights matrices. As in the previous sections, we use the row-standardised *k*-nearest neighbors matrix as defined in Eq. 4 with 7, 10 and 15 neighbors. We then compute the global Moran's *I* statistic (see Eq. 1) and perform an inference based on 9,999 permutations. The results are displayed in Table 6.2.

Table 6.2: Moran's I statistics for per capita income in Europe, 2011

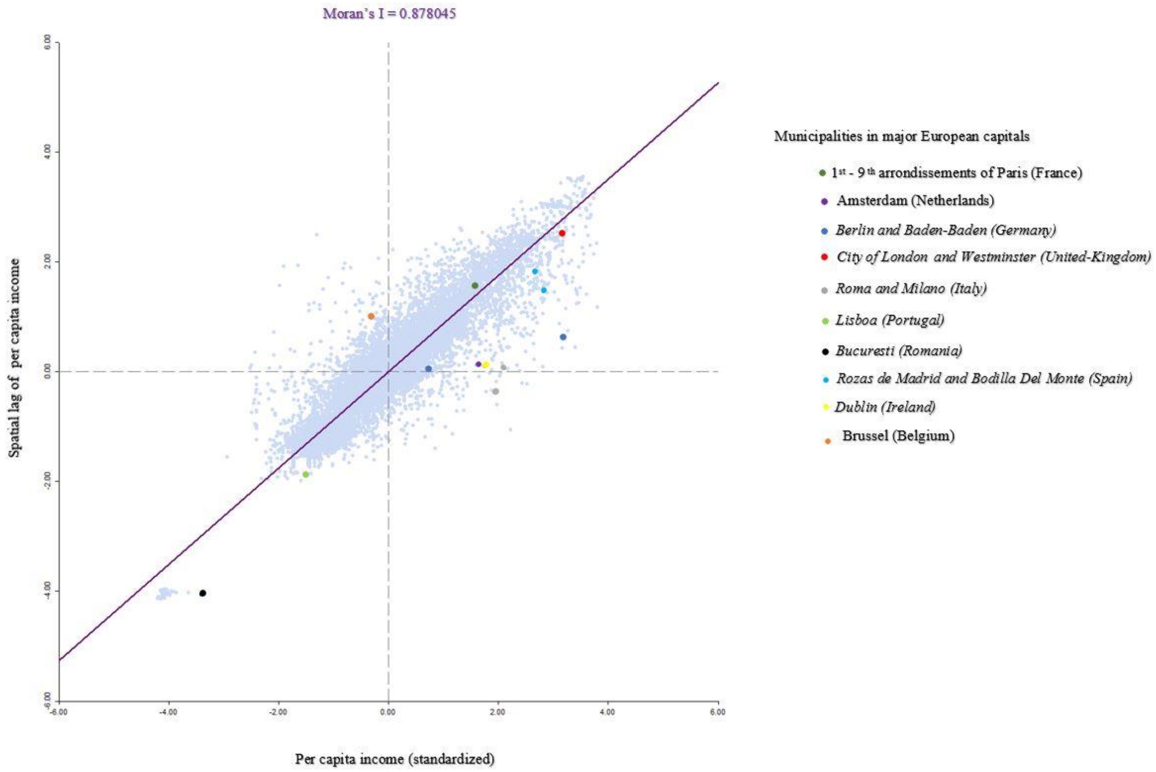
k-nearest neighbors weight matrices	Moran's *I*	expectation	Variance	*p*-value
7	0.8707		5.70e-06	2.2e-16
10	0.8781	4.61-05	8.52e-06	2.2e-16
15	0.8707		5.70e-06	2.2e-16

Note: inference is based on the permutation approach.

The results confirm the presence of positive and strongly significant spatial autocorrelation in the distribution of local per capita income in Europe in 2011. As the value k = 10 maximises the standardised value of Moran's *I*, we continue with this value for subsequent analysis.

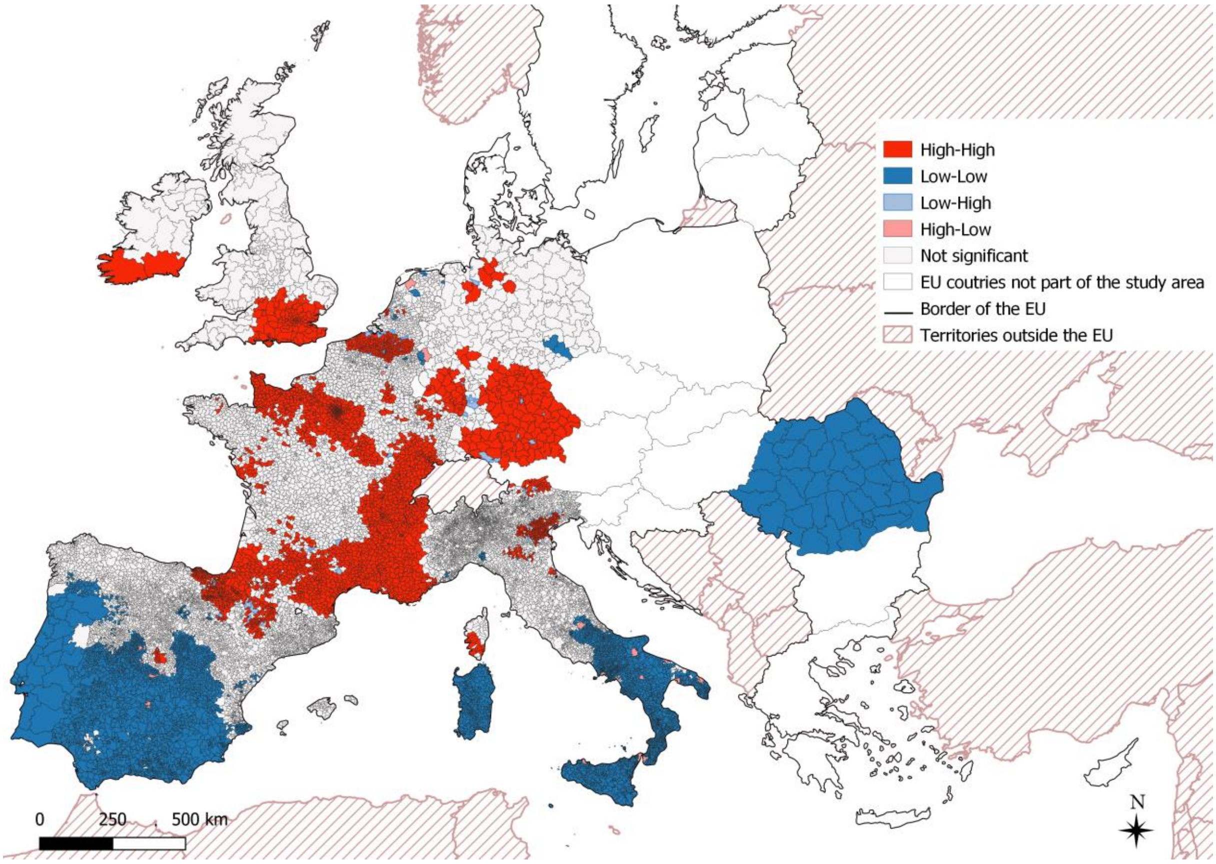
This global pattern of spatial autocorrelation may mask spatial atypical locations. Therefore, we proceed with the computation of a Moran scatterplot that is displayed in Figure 6.3. The figure also highlights the points corresponding to the capital regions of the 10 countries of the sample. It appears that the vast majority of observations (92%) are located in the HH and LL quadrants, each of them grouping 46% of the observations. The quadrants LH and HL where the spatial atypical observations are located each group 4% of the observations. The observations with high per capita income and rich neighbors compared to the sample mean are mainly to be found in France, Germany, Ireland, UK, Belgium and the Netherlands but also in the North of Italy and some Spanish municipalities around Madrid and close to the French border. Conversely, the observations with low per capita income surrounding by poor neighboors compared to the sample mean are mostly located in southern and the CEECs. The observations in the LH quadrant are to be found in France, Germany, Belgium, Netherlands and northern Italy. These observations are municipalities experiencing economic problems that do not allow them to benefit from the dynamism of their neighbors. Conversely, observations located in the HL quadrant are municipalities located in Spain and Italy.

Figure 6.3: Moran scatterplot for average household income, 2011



Finally, we compute the local Moran statistics with an inference based on the permutation approach and applying the Bonferroni adjustment to deal with the multiple comparison problem. The results are displayed in Figure 6.4. Interesting patterns are apparent. The significant HH observations (in red) are located around the capital regions in France and UK, a large number of French regions around the Swiss, Italian and Spanish borders, some areas in south of Ireland, northern Italy, in Germany and north of Europe. The significant LL observations concern the whole of Romania, south of Italy, Portugal and a large part of Spain. Some significant HL areas appear in South of Spain and South of Italy while some significant LH areas are to be found in Germany. Note that when the Bonferroni adjustment is applied, almost all HL or LH observations are not significant, meaning that the local spatial pattern of spatial autocorrelation is vastly characterized by a spatial heterogeneity between HH and LL areas.

Figure 6.4: Local Moran clusters for 2011, income per capita – p-value = 0.05 with Bonferroni adjustment



7. Conclusion

The report makes two contributions.

On the one hand, we analyze socioeconomic disparities at work in a sample of 275 regions in EU-28 from a dynamic perspective (2000-2015). Starting from a wide set of socioeconomic indicators from Cambridge Econometrics' European Regional and Eurostat REGIO databases, we show that the use of Multiple Factor Analysis (MFA) is an appropriate dimension reduction tool for dynamic analyses. Indeed, unlike principal component analysis (PCA) with which we are likely to end with factors that are not comparable over years, MFA builds up a common global space defined by several global components while taking into account the dataset structure (year-by-year observations on EU-28 regions). For each year, the projection of observations on regions leads to the yearly scores used in the dynamic analysis. Several interesting observations emerge from this analysis.

First, using the first factor of MFA, which provides indications on economic condition (ECO-DEV), we show that on the one hand European regions are spatially clustered and that on the other hand, most regions, along with their neighbors improves their relative position from 2000 to 2015. Globally, we reveal that there is a convergence process at work in the EU-28 regions from 2000 to

2015 but that this convergence pattern is moderated as most regions which improves/worsens their situation, along with their neighbors remains in their quadrant of origin in the Moran scatter plot.

Second, when we compare these results with those obtained for the usual indicator of economic activity, i.e. GDP per capita, we show that the convergence pattern detected with ECO-DEV is less pronounced than with GDP p.c.. This would means that ECO-DEV adjusts slower over time compared to GDP p.c. In other words, some of the original variables contributing to ECO-DEV (from the MFA) must be relatively rigid.

Third, pictures provided by the remaining interesting factors, i.e. factors 2 to 4 are completely different from the one provided by ECO-DEV. One can note that people of most European regions are getting more educated over time. Also, most regions from France, Portugal, Spain and Italy are becoming younger. This is however contrasted by the opposite trend of almost equal strength for regions from UK and Eastern countries. Finally, most regions from France, Italy, Scandinavian countries, north of Germany and west of Poland are are losing grounds on the competitiveness race.

All these results point to the limits of GDP p.c. as as single indicator of development. Several research directions could be further investigated. In particular, conditionally to the availability of data, a multiscalar analysis could be undertaken. Indeed, Dias Dapena *et al*. (2018) and Rubiera-Morollon Paredes] show that for per capita GDP, a general process of convergence in the EU co-exists with intranational processes of divergence. It could be interesting to analyze whether such differences due to spatial scale also exist for the MFA factors, notably economic development.

On the other hand, we perform ESDA on the local microdata and show that the spatial distribution of local income per capita is strongly and positively spatially autocorrelated and that the local spatial pattern of spatial autocorrelation is vastly characterized by a spatial heterogeneity between HH and LL areas. Further analysis could include a systematic multiscalar approach.

References

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *WiREs Computational Statistics*, *5*, 149–179.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*, 93-115.

Botzen, K. (2016). Social capital and economic well-being in Germany's regions: an exploratory spatial data analysis. *Region, 3*, 1-24.

Breau, S., Shin, M.J., & Burkhart, N. (2018). Pulling apart: new perspectives on the spatial dimensions of neighborhood income disparities in Canadian cities. *Journal of Geographical Systems, 20*, 1-25.

Chapman, S.A., & Meliciani, V. (2012) Income disparities in the enlarged EU: socio-economic, specialisation and geographical clusters. *Tijdschrift voor Economische en Sociale Geografie, 103*, 293-311.

Chocolatá, M., & Furková, A. (2017). Regional disparities in education attainment level in the European Union: a spatial approach. *Baltic Journal of European Studies, 7,* 107-131.

Dall'erba, S. (2005). Distribution of regional income and regional funds in europe 1989–1999: An exploratory spatial data analysis. *Annals of Regional Science*, *39*, 121-148.

Del Campo, C., Monteiro, C. M., & Soares, J.O. (2008). The european regional policy and the socio-economic diversity of European regions: A multivariate analysis. *European Journal of Operational Research*, *187*, 600–612.

Di Liberto, A., & Usai, S. (2013). TFP convergence across European regions: a comparative spatial dynamics analysis. In Crescenzi, R., & Percoco, M. (Eds.), *Geography, Institutions and Regional Economic Performance*. *Advances in Spatial Science*, Springer-Verlag, Berlin, 39-58.

Diaz Dapena, A., Rubiera-Morollon, F., & Paredes, D. (2018). New approach to economic convergence in the EU: a multilevel analysis from the spatial effects perspective. *International Regional Science Review*, *forthcoming*.

Durlauf, S., & Johnson, J.P.A., & Temple, J.R.W. (2005). Growth econometrics. In Aghion P., & Durlauf, S.N. (Eds.), *Handbook of Economic Growth Volume 1A*, North-Holland Publishing Company, 555-677.

Erdem, U. (2016). Regional human capital distribution and disparities in Turkey. *Review of Urban & Regional Development Studies, 28*, 16-31.

Ertur, C., & Koch, W. (2006). Regional disparities in the european union and the enlargement process: an exploratory spatial data analysis, 1995–2000. *Annals of Regional Science*, *40*, 723–765.

Ertur, C., & Le Gallo, J. (2003) An exploratory spatial data analysis of European regional disparities, 1980-1995. In Fingleton, B. (Ed.), *European Regional Growth*, *Advances in Spatial Science*, Springer-Verlag, Berlin, 55-98.

Escofier, B., & Pages, J. (1983). Méthode pour l'analyse de plusieurs groupes de variables. Application à la caractérisation de vins rouges du val de loire. *Revue de Statistique Appliquée*, *31*, 43–59.

Escofier, B., & Pages, J. (1994). Multiple factor analysis (afmult package). *Computational statistics & data analysis*, *18*, 121–140.

European Commission (2004). *A New Partnership for Cohesion. Convergence, Competitiveness, Cooperation. Third Report on Economic and Social Cohesion*.

Ezcurra, R., Pascual, P., & Rapún, M. (2007). Ths spatial distribution of income inequality in the European Union. *Environment and Planning A, 39,* 869-890.

Fazio, G. & Lavecchia, L. (2013). Social capital formation across space: proximity and trust in European regions. *International Regional Science Review, 36,* 296-321.

Fleurbaey, M. (2009). Beyond gdp: The quest for a measure of social welfare. *Journal of Economic literature*, *47* , 1029-75.

Gregory, T., & Patuelli, R. (2015). Demographic ageing and the polarization of regions: an exploratory space-time analysis. *Environment and Planning A, 47,* 1192-2010.

Gutiérreza, M.L.S., & Rey, S.J. (2013). Space-time income distribution dynamics in Mexico. *Annals of GIS, 19,* 195-207.

Kalogirou, S. (2010). Spatial inequalities in income and post-graduate educational attainment in Greece. *Journal of Maps, 6,* 189-206.

Kendall, M. G. (1962). *Rank Correlation Methods*. 3rd ed. London: Griffin.

Le Gallo, J., & Ertur, C. (2003). Exploratory spatial data analysis of the distribution of regional per capita gdp in Europe, 1980- 1995. *Papers in regional science*, *82*, 175–201.

López-Bazo, E., Vaya, E., Mora, A. J., & Surinach, J. (1999). Regional economic dynamics and convergence in the European Union. *Annals of Regional Science*, *33*, 343-370.

Hierro, M., & Maza, A. (2009). Structural shifts in the dynamics of the European income distribution. *Economic Modelling*, *26*, 733-739.

Pagès, J. (2014). *Multiple Factor Analysis by Example using R* . Chapman and Hall/CRC.

Rey, S.J. (2001). Spatial empirics for economic growth and convergence. *Geographical analysis*, *33*, 195–214.

Rey, S.J. (2004). Spatial analysis of regional income inequality. *Spatially integrated social science*, *1*, 280–299.

Rey, S.J. (2016). Space–time patterns of rank concordance: Local indicators of mobility association with application to spatial income inequality dynamics. *Annals of the American Association of Geographers*, *106*, 788–803.

Rey, S.J., & Le Gallo, J. (2009). Spatial analysis of economic convergence. In Mills, T.C., & Patterson, K. (Eds.), *The Palgrave Handbook of Econometrics volume ii: Applied Econometrics*. Palgrave-MacMillan, 1251-1292.

Rey, S.J., & Ye, X. (2010). Comparative spatial dynamics of regional systems. In Paez, A., Le Gallo, J., Dall'erba, S., & Buliung, R. (Eds.), *Progress in Spatial Analysis*, *Advances in Spatial Science*, Springer-Verlag, Berlin, 441-463.

Rey, S.J., Murray, A.T., & Anselin, L. (2011). Visualizing regional income distribution dynamics. *Letters in Spatial and Resource Sciences*, *4*, 81–90.

Robert, C., Kubiszewski, I., Giovannini, E., Lovins, H., McGlade, J., Pickett, K., Vala Ragnarsdottir, K., Roberts, D., De Vogli, R., & Wilkinson, R. (2014). Time to leave gdp behind. *Nature*, 505.

Rodríguez-Pose, A., & Tselios, V. (2011). Mapping the european regional educational distribution. *European Urban and Regional Studies*, *18*, 358–374.

Tselios, V., & Stathakis, D. (2019) Exploring regional and urban clusters and patterns in Europe using satellite observed lighting, *Environment and Planning B*,, forthcoming.

Appendix

A1. Interpolation techniques used

Since some variables in some regions (NUTS2 level), display some missing value, we use a combination of the following four interpolation and extrapolation techniques to fill in these missing values: (a) the national variation rate is applied to regional values (when we have at least data at one point of time for a given variable and NUTS2 region and for the considered variable, its country-level the growth rate); (b) the country-level values are used at the NUTS2 level (when we have no data for a given variable expressed in percentage and NUTS2 region, but the country level values); (c) the country-level values weighted by the population are used at the NUTS2 level (when we have no data for a given variable expressed in numbers, euros, etc. and NUTS2 region, but the country level values); (d) we consider the mean of available years (when there is no country-level and NUTS2 values).

We report in table A1 for all variables considered in the analysis, the interpolation rate applied to fill in missing values from 2000 to 2015 by methods (a), (b), (c), and (d). Table 8 also displays in the last column the total interpolation rate. It gives an idea on the percentage of missing values for each variable. Note that the option (d) rate, which is the most questionable interpolation technique, for almost all our variables does not exceed 0.6% of the total sample (the highest rates are observed for the variable bed_hos with 3.5% missing values filled in with this method).

Table A1: Interpolation rate by techniques (in %)

Code	a	b	c	d	Total
Demography					
pop_dens	4.16	0	0	0.19	4.35
pop_14	2.54	0	0	0	2.54
pop_1564	2.54	0	0	0	2.54
pop_65	2.54	0	0	0	2.54
lifexp_0	6.84	0	0	0.02	6.86
lifexp_50	6.84	0	0	0.02	6.86
fert_age	5.46	0	0	0	5.46
fert_rate	5.46	0	0	0	5.46
Economy					
gdp_head	0.7	0	0	0	0.7
emp_agri	0	0	0	0	0
emp_indu	0	0	0	0	0
emp_cons	0	0	0	0	0
emp_trad	0	0	0	0	0
emp_fin	0	0	0	0	0
emp_adm	0	0	0	0	0

Employment					
emp_tot	4.24	0	0	0.04	4.28
emp_fem	4.28	0	0	0.04	4.32
emp_mal	4.24	0	0	0.04	4.28
unemp_tot	5.48	0.36	0	0.04	5.88
unemp_lt	11.49	2.17	0	0.58	14.24
unemp_fem	9.27	0.36	0	0.04	9.67
unemp_you	12.98	1.45	0	0.04	14.47
neet_fem	12.23	3.99	0	0.36	16.58
neet_mal	15.15	4.71	0	0.36	20.22
neet_tot	6.5	1.45	0	0.36	8.31
Education					
low_edu	5.07	0.36	0	0.04	5.47
med_edu	5.07	0.36	0	0.04	5.47
high_edu	5.07	0.36	0	0.04	5.47
trng_fem	8.97	0.7	0	0.53	10.2
trng_mal	8.8	1.77	0	0.53	11.1
trng_tot	6.69	0.36	0	0.53	7.58
health					
bed_hos	4.26	32.42	0	3.5	40.18
per_health	5.67	32.97	0	1.9	40.54

Table A2: NUTS2 regions distribution per EU-28 countries

Country	Sigle	Nb. of regions	Country	Sigle	Nb. of regions
Austria	AT	9	Ireland	IE	2
Belgium	BE	11	Italy	IT	21
Bulgaria	BG	6	Lithuania	LT	1
Cyprus	CY	1	Luxembourg	LU	1
Czech Republic	CZ	8	Latvia	LV	1
Germany	DE	38	Malta	MT	1
Denmark	DK	5	Netherlands	NL	12
Estonia	EE	1	Poland	PL	16
Greece	EL	13	Portugal	PT	7
Spain	ES	19	Romania	RO	8
Finland	FI	5	Sweden	SE	8
France	FR	27	Slovenia	SI	2
Croatia	HR	2	Slovakia	SK	4
Hungary	HU	7	United Kingdom	UK	40