Beyond GDP: an analysis of the socio-economic diversity of European regions

Ayouba, Kassoum; Le Gallo, Julie; Vallone, Andrès

Postprint / Postprint
Zeitschriftenartikel / journal article

Beyond GDP: an analysis of the socio-economic diversity of European regions☆

Kassoum Ayouba[a], Julie Le Gallo[a,*], Andrés Vallone[b]

[a] *CESAER UMR 1041, AgroSup Dijon, INRA, Université de Bourgogne Franche-Comté, 26 boulevard Petitjean, 21079 Dijon Cedex, France*
[b] *Universidad Católica del Norte, Escuela de Ciencias Empresalariales, Larrondo 1281, Coquimbo, Chile*

Abstract

This paper aims to analyze the socioeconomic diversity of European Union (EU-28) regions from a dynamic perspective. For that purpose, we combine a series of exploratory space-time analysis approaches to multiple Factor Analysis (MFA) applied to a large range of indicators collected at the NUTS-2 level for the period 2000-2015 for the EU-28. First, we find that the first factor of MFA, interpreted as economic development (ECO-DEV), is spatially clustered and that a moderate convergence process is at work between European regions from 2000 to 2015. Second, when comparing these results with those obtained for GDP per capita, we show that the convergence pattern detected with GDP per capita is more pronounced: ECO-DEV adjusts slower over time compared to GDP per capita. Third, pictures provided by the remaining interesting factors, capturing educational attainment, population dynamics and employment, are very different.

Keywords: Multiple Factor Analysis, Exploratory Space-Time Analysis, European regions, Spatial autocorrelation
JEL codes: C14, O18, R12

☆This research was funded by H2020 IMAJINE (Integrative Mechanisms for Addressing Spatial Justice and Territorial Inequalities in Europe), Grant agreement 726950. We thank V. Larmet for excellent research assistance. We also thank E. Elguezabal, G. Lafferté, J. Mischi, R. Sinthon, L. Védrine and the participants of the research seminar held in Saint Etienne on February 14, 2019 for useful comments.

*Corresponding author
Email addresses: ayouba.kassoum@inra.fr (Kassoum Ayouba), julie.le-gallo@agrosupdijon.fr (Julie Le Gallo), avallone@ucn.cl (Andrés Vallone)

1. 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 seminal study by Barro et al. (1991). 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 et al., 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 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 starting inter alia with Rey & Montouri (1999) and Lopez-Bazo et al. (1999), 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 & 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; Rey, 2014). Our paper, by implementing a large range of 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. Applications making use of exploratory spatial data analysis applied to the distribution of GDP or income per capita in European regions include, among others, Lopez-Bazo et al.

(1999), Le Gallo & Ertur (2003), Dall'erba (2005) or Ertur & Koch (2006). Yet, other dimensions of disparities might be interesting. For instance, exploratory spatial data analysis methods were applied on educational attainment and inequality in European regions (Rodríguez-Pose & Tselios, 2011; Chocholatá & Furková, 2017; Kalogirou, 2010), on human capital in Turkish districts (Erdem, 2016) or on social capital (Fazio & Lavecchia, 2013; Botzen, 2016) and demographic ageing (Gregory & 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 aim of this article is 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 & Ye, 2010), Mexican states (Gutiérreza & Rey, 2013), Canadian cities (Breau et al., 2018) or other measures, such as total factor productivity (Di Liberto & 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 therefore 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 remainder of the paper is organized as follows. 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. Section 6 provides some concluding comments and suggests some avenues for future research.

2. Data description

Table 1 presents the socio-economic variables collected for the empirical analysis. They are grouped into five broad categories: demography, economy, employment, education, and health. Economy 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 1 includes 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 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 or means 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), employment variables (young people neither in employment nor in education and training[3]) and health variables (hospital beds and health personnel per 100,000 inhabitants).

Our sample includes 275 regions at the NUTS-2[4] level in 28 European countries over the period 2000-2015 (see Table A2 for regions' distribution per EU-28 countries).[5] We report in Table B3 in the appendix 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 1:1.2 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 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.

3. Regional socio-economic indicators

Since we collected data for numerous variables (see Table 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

[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.

[3]neet_fem (resp. neet_mal) represents the share of young female (resp. male) people (population ages 15-24) who are not in employment, education or training, as a percentage of the total number of young female (resp. male) people. neet_tot is the indicator computed without sex consideration.

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

[5]We remove the remote French island Mayotte.

Table 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) 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)
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 & training (in %)
neet_mal	Young male people neither in employment nor in education & training (in %)
neet_tot	Young people neither in employment nor in education & training (in %)
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

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).[6] 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[7] 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 2 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%. The first four factors explain more than 65% of the total variance.

Table 2: Principal component 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

Table 3 contains the correlations between the first four factors and the original variables. Since each variable appears sixteen times, as much correlation coefficients can be computed with

[6] The method has been introduced in Escofier & Pages (1983, 1994). For an extensive and comprehensive review, see Pagès (2014) and Abdi et al. (2013).

[7] 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 1.

the factors. However, as the correlation coefficients between factors and the yearly versions of each of the variables have a stable sign, we present the average correlation with factors for each variable. The most relevant correlations are shown in bold in Table 3. We also display in Table 3 the squared cosines of each variable to check for the quality of its projection on the four factors.

Factor 1, named *economic development* (ECO-DEV) is associated with a high level of the economic indicators presented in Table 1 (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 value on this factor are those with attractive and competitive employment centers.

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.

4.1. Distribution dynamics and spatial pattern: a global assessment

To start, we display in choropleth maps using a quintile classification (see Figure C1 in the appendix), 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

Table 3: Correlations between factors and the original variables and quality of the representation of original variables (squared cosines) – (averages)

	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

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 the 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 D4 in the appendix), 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.

This pattern of spatial clustering and its dynamics are explored in more detail using two global indexes: Moran 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 $\phi = \iota W \iota^\top / \iota(W + \overline{W})\iota^\top$ represents the share of all pairs that involve geographic neighbors. W and $\overline{W} = \iota\iota^\top - W - I_{n\times n}$ 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 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[8] and avoid a block-diagonal structure of the weights matrix (Le Gallo & 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). 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 D5 in the appendix) 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.

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

Table 4: Spatial autocorrelation and spatial concordance, 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

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), 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 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., 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

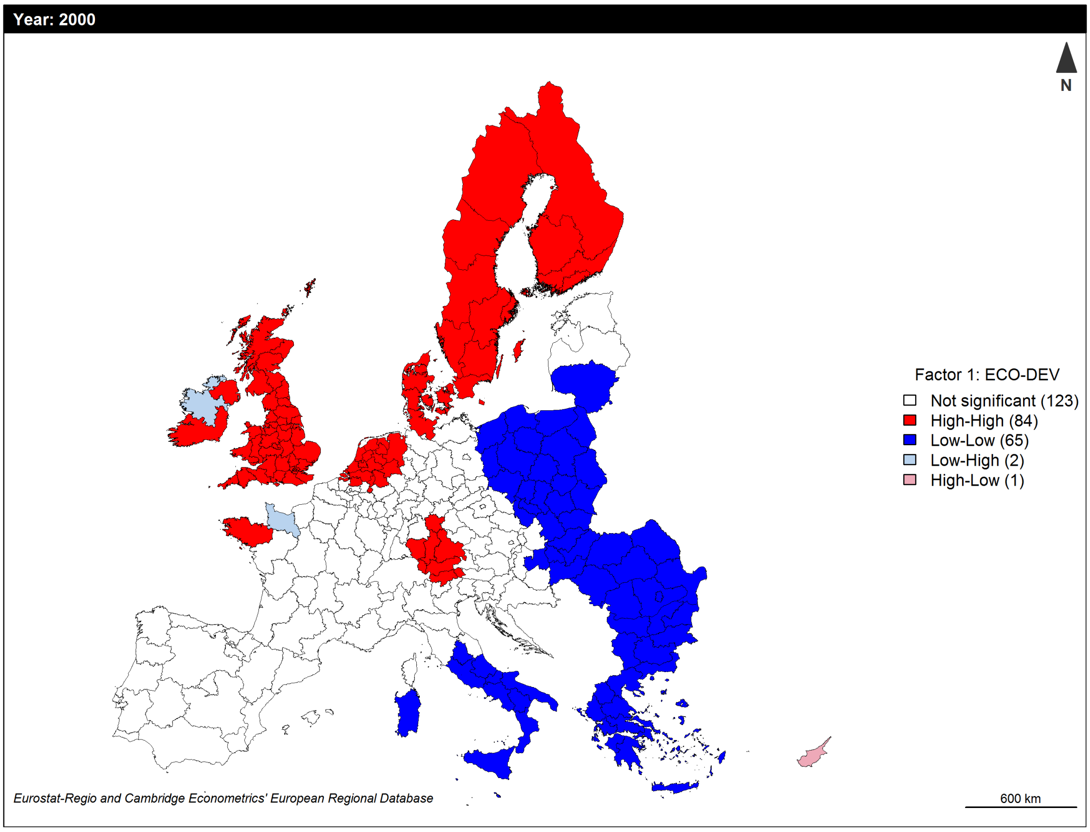
$\sum_{j=1}^{n} w_{ij} z_{j,t}$.[9] Inference is based on 9,999 random permutations and we use the Bonferroni p-value correction to deal with the multiple comparison problem. Specifically, we set to 0.05 the overall significance associated with multiple comparisons. Then since each observation (region) 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 1.

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 1 shows four big clusters of rich regions. The first one is located in the 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.[10] 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.[11] While this picture is overall close to the one provided by GDP p.c. (see Figure D5 in the appendix), and that found by Ertur & 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 the 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] 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).

[10] In total, more than 96.42% of regions in high-high clusters in 2000 remain in the same cluster in 2015. 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.

[11] More than 87% of regions in low-low clusters in 2000 remain in the same cluster in 2015. 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.



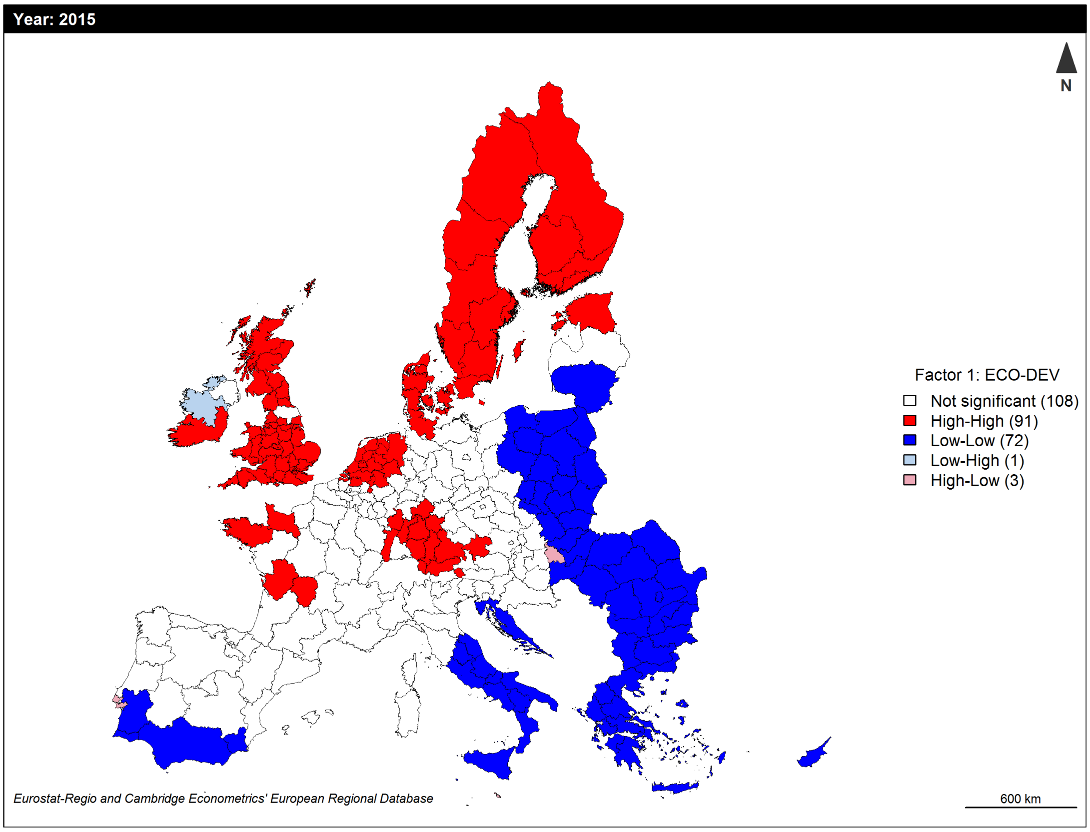


Figure 1: Local Moran clusters for 2000 and 2015 (ECO-DEV) – p-value = 0.05 with Bonferroni adjustment

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 2. 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 D6 in the appendix). 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%).

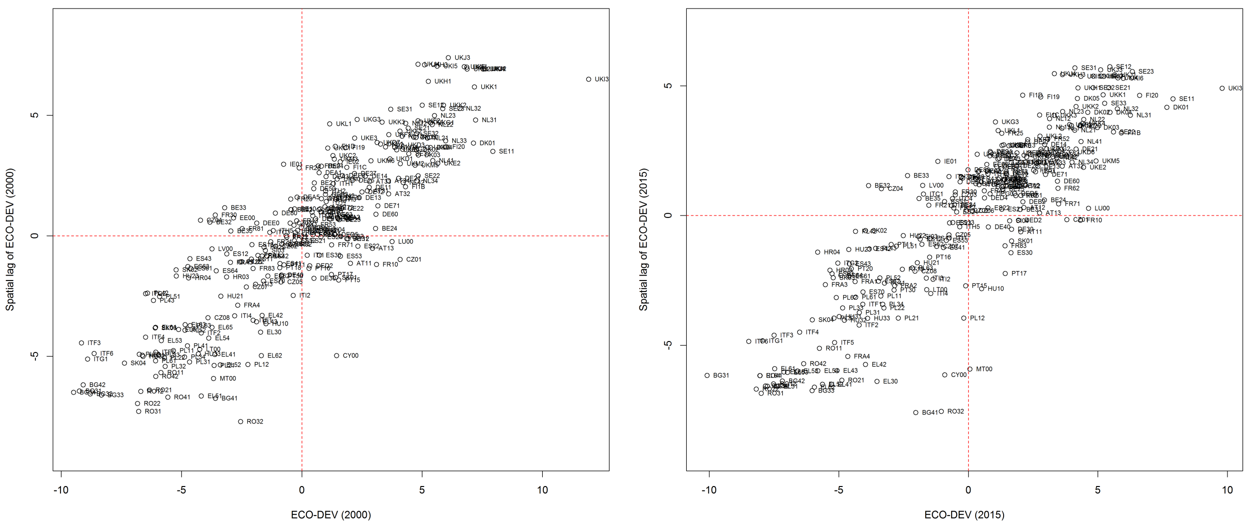


Figure 2: Moran Scatterplots for 2000 and 2015, ECO-DEV

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 3 (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 3 (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 3, we also provide in Figure C2 in the appendix their dynamics, depending on the location of regions at the beginning of the period. For example, in the first panel of Figure C2 (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. 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 D7 and D8 in the appendix) 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.

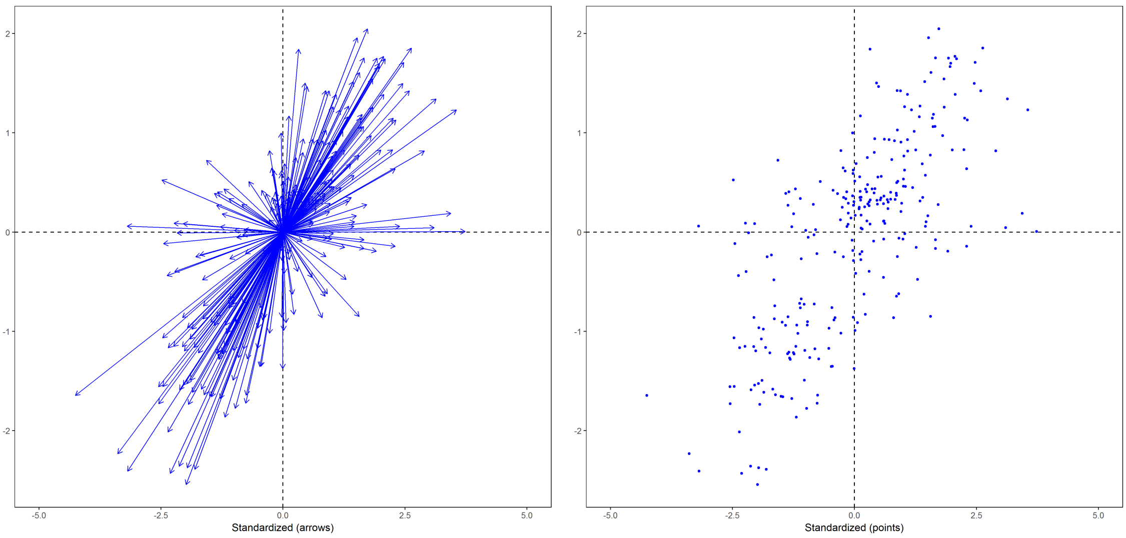


Figure 3: Standardized Moran scatterplots for 2000-2015, ECO-DEV

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. 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. 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 Figures D9 and D10 in the appendix): the position of the vast majority of regions worsens during the period.

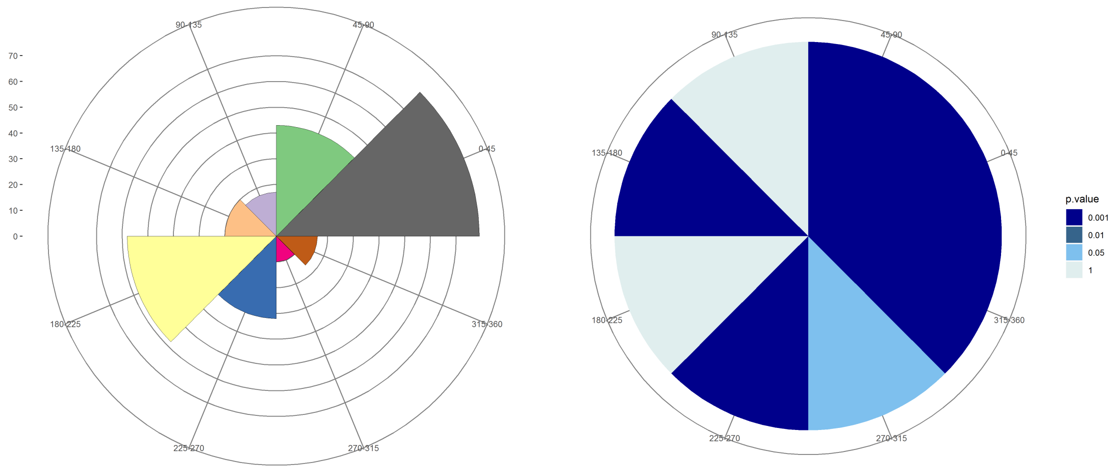


Figure 4: Rose diagram and p-values for $k = 8$, ECO-DEV

For identification purposes, we plot the regions displaying the direction of moves in Figure 5. 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 D11 in the appendix). 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.

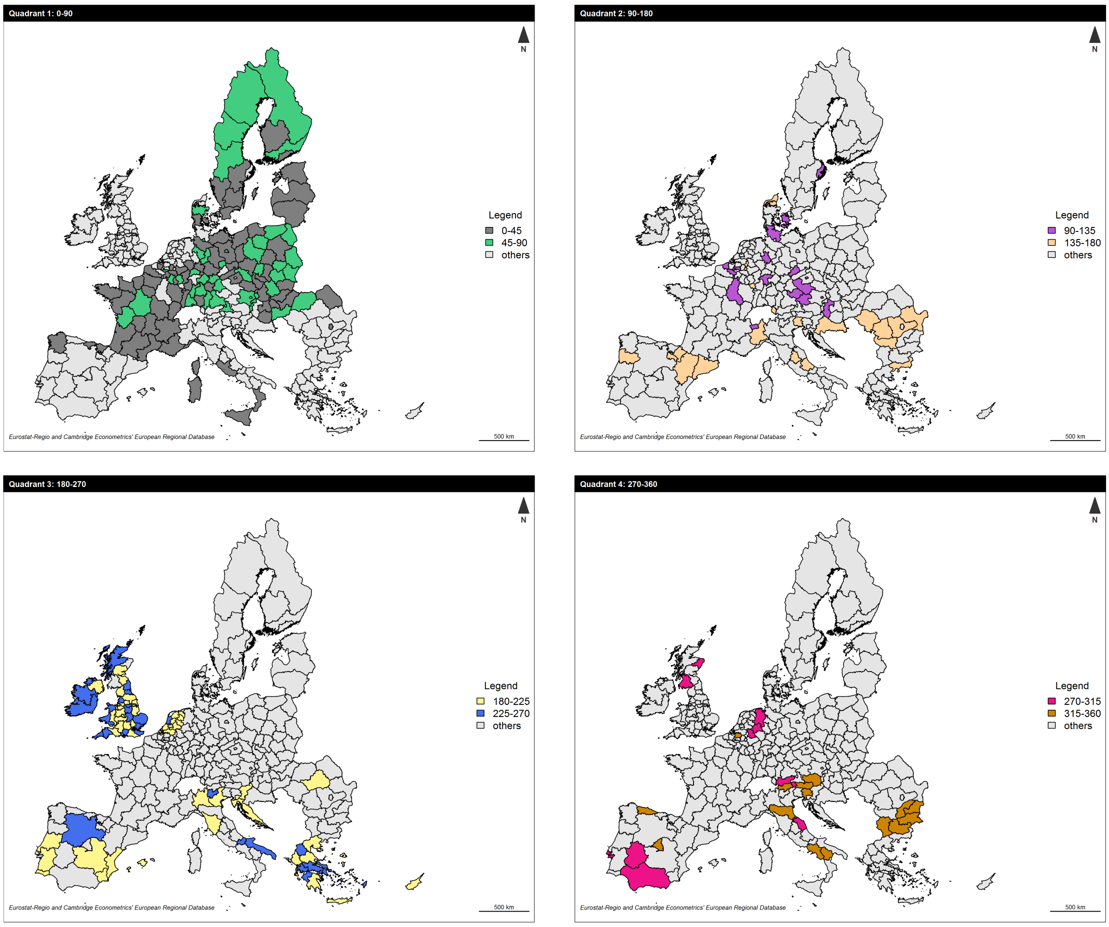


Figure 5: Identification of regions from the 8-class rose diagram, ECO-DEV

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 5. 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[12] is relatively small (0.1652), confirming the persistence of spatial dependence also highlighted by GIMA. The last row of Table 5 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.

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 D6 in the appendix– 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 (see Figure C2), stay in their starting quadrant over time.

Table 5: 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

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 6. 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).

[12]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 $k/(k-1)$.

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. When we compare these results to those obtained with GDP p.c. (see Table D7 in the appendix), 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 6: LISA Markov transition probabilities (5 classes), ECO-DEV

		End				
		HH	LL	LH	HL	non sign.
	HH	0.9581	0.0000	0.0054	0.0000	0.0365
	LL	0.0000	0.9749	0.0000	0.0029	0.0222
Beginning	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

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 the HH and LL quadrants. 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 factors suggest spatial association of similar values, more pronounced for LOW-EDUC and POP-DYN compared to ACT-POP (see Figure E12 in the appendix). Regarding the factor 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 Rodríguez-Pose & Tselios (2011) who found that Portuguese, followed by the Spanish, French, Italian and British are the least educated in Western Europe in 2010, 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.

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 E8 in the appendix). 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.

Figure 6 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 factor. Moreover, unlike the previous factors 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.

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 E13 in the appendix. 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.

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 E14 in the appendix), 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 E15 in the appendix), 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 E16 in the appendix), 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 E17 in the appendix) 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 E13 remain valid.

We finally estimate LISA Markov chains over the study period (from 2000 to 2015). The obtained results are reported in Table 7. 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

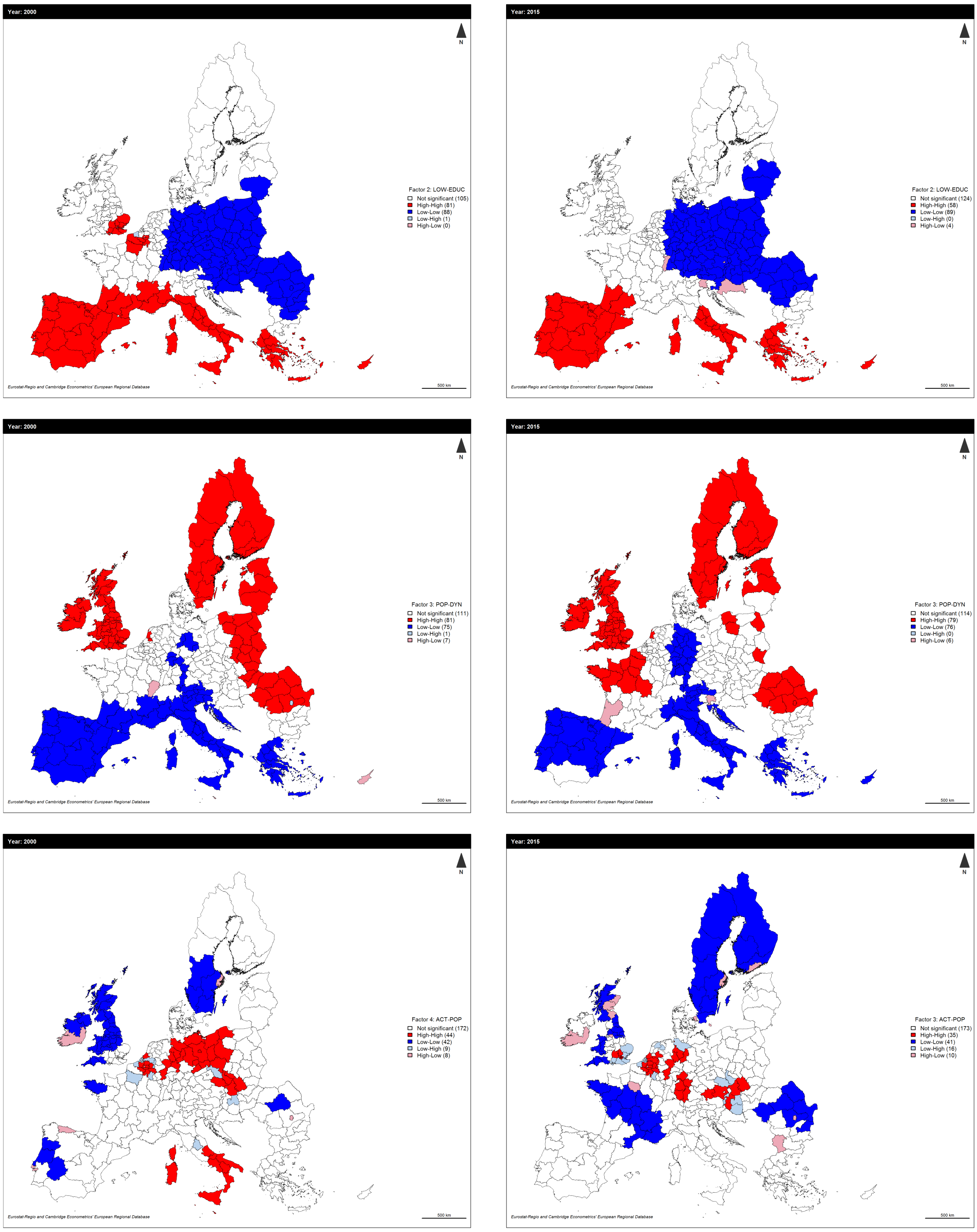


Figure 6: Local Moran clusters for 2000 and 2015 (factors 2 to 4) – p-value = 0.05 with Bonferroni adjustment

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. The five-class LISA Markov are in Table E9 in the appendix. 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 7: LISA Markov transition probabilities (4 classes), 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

6. Conclusion

We analyze in this paper 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 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 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 losing grounds on the competitiveness race.

All these results point to the limits of GDP p.c. 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, Díaz Dapena et al. (2019) 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.

References

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis: principal component analysis for multitable and multiblock data sets. *Wiley Interdisciplinary reviews: computational statistics*, *5*, 149–179.

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

Barro, R. J., Sala-I-Martin, X., Blanchard, O. J., & Hall, R. E. (1991). Convergence across states and regions. *Brookings Papers on Economic Activity*, *1991*, 107–182.

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 neighbourhood income disparities in canadian cities. *Journal of Geographical Systems*, *20*, 1–25.

Chocholatá, 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 R. Crescenzi, & M. Percoco (Eds.), *Geography, Institutions and Regional Economic Performance* Advances in Spatial Science (pp. 39–58). Springer.

Díaz Dapena, A., Rubiera-Morollon, F., & Paredes, D. (2019). New approach to economic convergence in the EU: a multilevel analysis from the spatial effects perspective. *International Regional Science Review*, *42*, 335–367.

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

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.

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*.

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.

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

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. (2014). Spatial dynamics and space-time data analysis. In *Handbook of Regional Science* (pp. 1365–1383). Springer.

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 T. Mills, & K. Patterson (Eds.), *The Palgrave Handbook of Econometrics Volume II: Applied Econometrics* (pp. 1251–1292). Palgrave-MacMillan.

Rey, S. J., & Montouri, B. (1999). US regional income convergence: A spatial econometric perspective. *Regional Studies*, *33*, 143–156.

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

Rey, S. J., & Ye, X. (2010). Comparative spatial dynamics of regional systems. In *Progress in spatial analysis* (pp. 441–463). Springer.

Robert, C., Kubiszewski, I., Giovannini, E., Lovins, H., McGlade, J., Pickett, K., Vala Ragnarsdóttir, 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.