The non-stationary influence of geography on the spatial agglomeration of production in the EU

Coro Chasco and Ana María López and Rachel Guillain

Universidad Autónoma de Madrid, Pôle d'Economie et de Gestion (Université de Bourgogne)

24. September 2008

The non-stationary influence of geography on the spatial agglomeration of production in the EU[1]

Coro Chasco
Dpto. Economía Aplicada
Universidad Autónoma de Madrid
28049 Madrid (Spain)
coro.chasco@uam.es

Ana Mª López
Dpto. Economía Aplicada
Universidad Autónoma de Madrid
28049 Madrid (Spain)
ana.lopez@uam.es

Rachel Guillain
Pôle d'Economie et de Gestion
Université de Bourgogne
21066 Dijon Cedex (France)
guillain@u-bourgogne.fr

ABSTRACT

In this paper, we investigate the relative importance of geographic features on the location of production in the EU. Specifically, we want to quantify how much of the spatial pattern of GDP can be attributed to only exogenous *first nature* elements (physical and political geography) and how much can be derived from endogenous *second nature* factors (man-made agglomeration economies). In order to disentangle both effects empirically, and to learn how they are interrelated, we control for second nature. We use a methodology based on an analysis of variance (ANOVA), which is applied to a panel of 1,171 European NUT-3 in 2006. We demonstrate that -due to a high degree of spatial non-stationarity present in the data- results can be biased if spatial autocorrelation and spatial heterogeneity, as well as multicollinearity and endogeneity, are not properly taken into account.

Key-words: Agglomeration, Geography, Spatial Heterogeneity, Endogeneity, EU Regions

JEL codes: C21, C51, C52, O18, O52, R12

1. INTRODUCTION

In recent years, there has been a growing interest in the geographic aspects of development. In fact, many economic activities are concentrated geographically and most people in advanced countries or regions live in densely populated metropolitan areas. The main issue is how to explain this concentration. Most of the references

[1] Coro Chasco acknowledges financial support from the Spanish Ministry of Education and Science SEJ2006-02328/ECON and SEJ2006-14277-C04-01. Rachel Guillain also acknowledges the financing support of the Programme 'Young researchers 2005' from the French ANR-CNRS ('Agence Nationale de la Recherche').

assume two approaches, *first nature* (Sachs 2000) and *second nature* (Krugman 1993; Krugman 1999; Venables 2003), which are also identified as Sachs' (first nature) and Krugman's approach (second nature). Krugman's New Economic Geography states that agglomerations can be explained by second nature alone (i.e. man-made agglomeration economies due to increasing returns to scale and transportation costs), which arises endogenously in the economic process. However, real world agglomeration is possibly caused by both first and second nature. In this case, it would be interesting to compute the exact influence of both types of agglomeration advantages on economic distribution across space.

In this paper, our aim is to examine the influence of geographic features on the location of production in Europe. In other words, we focus on quantifying how much of the geographic pattern of GDP can be attributed to only exogenous first nature elements (physical and political geography), how much can be derived from endogenous second nature factors (man-made agglomeration economies) and how much is due to the interaction of both effects. Specifically we disentangle the two net effects empirically, as well as their mixed effect, analyzing the spatial non-stationary distribution of these relationships across the European regions.

For this purpose, we apply a methodological approach based on Roos (2005) for Germany. He proposes to employ an analysis of variance (ANOVA) to infer the unobservable importance of first nature indirectly in a stepwise procedure. In order to disentangle first and second nature effects empirically, we control for second nature because every locational endowment will be reinforced and overlaid by second nature advantages. Controlling for spatial effects, we also estimate how much agglomeration can by explained by both gross and net second nature with the aim of isolating the importance of first nature alone. Whereas this method seems quite clear and direct, we demonstrate that results could be biased if some potential econometric questions are not properly taken into account; e.g. multicollinearity, relevant missing variables, endogeneity, spatial autocorrelation and spatial heterogeneity.

In fact, in many countries GDP density is strongly polarized on two subspaces, core and periphery, displaying spatial heterogeneity. Nevertheless in the particular case of the EU, we identify a 'Eurocenter' and three extensive peripheries located in the North, East and South edges of the continent. In addition, inside these principal spatial regimes we could also distinguish their own core and periphery.

The organization of the paper is as follows. Section 2 contains a description of the ANOVA model. In Section 3, we analyze the data and the distribution of economic agglomeration in the EU. The empirical results derived from the econometric process, and the concrete analysis of spatial non-stationarity, are presented in Section 4 and 5, respectively. The conclusions in Section 6 and the references put an end to our analysis.

2. THEORETICAL MODEL

Three forces operate in forming agglomerations: an unobservable direct effect of first nature, a first nature effect working through induced agglomeration economies and a direct effect of second nature, which would exist even without any first nature forces. In order to get a better knowledge of these effects, Roos (2005) states a methodology based on analysis of variance (ANOVA). The total variance V of the dependent variable can be decomposed into four parts:

$$V = V_u + V_f + V_s + V_{fs} \tag{1}$$

where V is the total variance of the dependent variable, V_u is the unexplained variance, V_f is the variance explained by first nature alone, V_s is the variance explained by second nature alone and V_{fs} is the variance explained by a combination of both forces.

ANOVA is employed to infer the unobservable importance of first nature alone indirectly, as well as to assess about the relative importance of first and second nature forces. It is a four-step process that proceeds as follows:

1. Since man-made agglomeration effects (second nature) are usually triggered by natural advantages (first nature), we must first identify the net from the gross second nature effect. For this purpose, we regress two gross second nature variables on first nature. These regressions explain how much of the gross second nature effects are caused by purely first nature. By mean of the residuals of the regressions, we filter the net from the gross second nature variables.

2. We estimate how much of GDP per area variance can be explained by gross ($V_s + V_{fs}$) and net (V_s) second nature advantages. These calculations can be derived from the results of two regressions of GDP density on both gross and net second nature variables.

3. We estimate how much of GDP per area variance can be explained jointly by first and second nature ($V_f + V_s + V_{fs}$**)**. The total effect of first and second nature can be obtained from a regression, using first and net second nature variables as explanatory variables.

4. We calculate the difference between the result in step 3 (total effect of first and second nature) and step 2 (total effect of second nature), which is the importance of first nature alone (V_f) on GDP per area.

   Next, we will explain the whole process in depth.

Since first and second nature are interrelated, in a *first step* it is necessary to disentangle the second nature variables (population and GDP per worker) empirically. For that purpose, we can regress them on geography and take the residuals $\hat{\pi}$ and $\hat{\rho}$ as variables of net second nature forces:

$$\begin{aligned} \log\left(pop_i\right) &= \gamma_0 + \sum_{k=1}^{K} \gamma_k f_{ki} + \pi_i \\ \log\left(prod_i\right) &= \rho_0 + \sum_{1}^{K} \rho_k f_{ki} + \delta_i \end{aligned} \tag{2}$$

where pop_i and $prod_i$ are total population and GDP per worker in region i, f_{ki} is the group of k geography variables, γ, ρ are coefficients and π, δ are the error terms of the regressions.

While variables s_{mi} = {log(pop_i), log($prod_i$)} are ‘gross’ second nature variables, the residuals of these regressions $\hat{s}_{mi} = \left\{\hat{\pi}_i, \hat{\delta}_i\right\}$ could be taken as geography-filtered or net second nature forces. The introduction of these sets of variables, s_{mi}, $\hat{s}_{mi}$, as explanatory variables will allow to evaluate the total influence of gross and net second nature variables on economic agglomeration, respectively.

In a *second step* we can compute the effects of total -both gross and net- second nature variables on GDP per area (*gd*), which is a proxy variable to measure economic agglomeration. In this fashion, the gross second nature variables influence is obtained with the estimation of the following equation:

$$\log\left(gd_i\right) = \alpha_0 + \sum_{m=1}^{M} \phi_m s_{mi} + \varepsilon_i \tag{3}$$

The resulting determination coefficient, R^2, indicates this gross effect of second nature on agglomeration:

$$R_{gs}^2 = \frac{\left(V_s + V_{fs}\right)}{V} \tag{4}$$

Regarding the net effect of second nature on GDP per area, it is derived from the estimation of the following equation:

$$\log\left(gd_i\right) = \alpha_0 + \sum_{m=1}^{M} \phi_m \hat{s}_{mi} + \varepsilon_i \tag{5}$$

The net effect of second nature on agglomeration can be expressed as:

$$R_{ns}^2 = \frac{V_s}{V} \tag{6}$$

Therefore, the mixed effect of the interaction between first and second nature on GDP density can be extracted as follows:

$$\frac{V_{fs}}{V} = R_{gs}^2 - R_{ns}^2 \tag{7}$$

In the *third step*, we measure the total effect of first and second nature on GDP per area. We could simply include, in another equation, the gross second nature variables as regressors together with a set of first nature indicators. However, this could bias the estimates of the first nature coefficients since first nature also has an effect on the second nature variables. In order to adjust the later for the former, we specify a regression of GDP per area on first and *net* second nature variables, which avoids the stochastic regressors problem:

$$\log\left(gd_i\right) = \alpha_0 + \sum_{k=1}^{K} \phi_k f_{ki} + \sum_{m=1}^{M} \phi_m \hat{s}_{mi} + \varepsilon_i \tag{8}$$

The joint importance of first and second nature is measured by the corresponding determination coefficient:

$$R^2_{f+s} = \frac{V_f + V_{fs} + V_s}{V} \tag{9}$$

In the *fourth step*, we derive the net importance of first nature on GDP density from the results of the previous estimations:

$$\frac{V_f}{V} = R^2_{f+s} - R^2_{gs} \tag{10}$$

The estimation of Eqs. (3), (5) and (8) by Ordinary Least Squares (OLS) could lead to biased results due to the presence of endogeneity on some of the explanatory variables and/or spatial effects on the residuals. Roos (2005) and Gallup et al. (1999) only consider the first problem but omit the second.

In effect, on the one hand endogeneity in a regressor can lead to a well-known simultaneity bias in the OLS estimates. Even in the pure-geography variables there could be different degrees of exogeneity. Physical geography variables (temperature, coast, etc.) can be considered as exogeneous since they do no depend on underlying economic forces. However political geography could have more endogeneous elements; e.g. the location of state capitals, though do not change very often, are possibly the result of the economic importance of the corresponding city. Moreover, the second nature variables (population and productivity) are much more endogenous and simultaneously determined with GDP density. On the other hand, spatial autocorrelation and/or spatial heterogeneity in the OLS residuals are also causes of misspecification problems in the regression (see Anselin 1988 for a complete view of this topic). They must be tested and corrected, as will be shown hereafter.

3. DATA

It is our aim to explain agglomeration from first and second nature elements. Hence, we must define first what we understand for agglomeration and geography to find the appropriate indicators.

3.1. The agglomeration measure

Regarding the endogenous variable, several measures have been used in the literature, e.g. population (Graves 1979; Cragg and Kahn 1997; Knapp et al. 2001), employment or GDP (Freeman 2001, Roos 2005), employment densities (Ciccone and Hall 1996, Rapaport and Sachs 2003), and area (Dobado 2004). In these last cases, agglomeration is conceived as the spatial concentration of not only production activities but also both workers/citizens.

We opt to use the relative GDP density –GDP per km^2- as the endogenous variable as in Delgado and Sánchez (1998). In spite of its potential drawbacks[2], this

[2] This indicator has potential drawbacks. On the one hand, it could not necessarily reflect in the true level spatial agglomeration of firms and workers, but only generated value added; i.e. GDP per km^2 can be possibly understated in regions where workers mostly commute to neighboring regions. On the other hand, aggregate GDP does not allow analysing the different effects of first and second nature factors in different industries (Alonso-Villar et al. 2004).

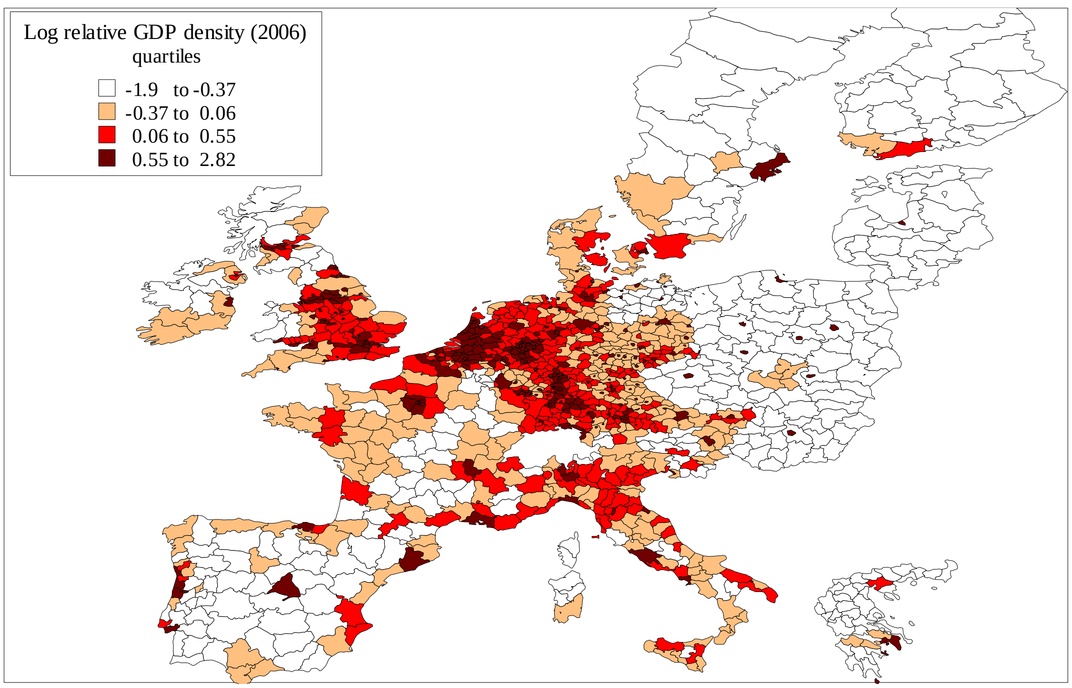
variable allows us to make direct comparisons with Roos's results for German regions. Formally, the endogenous variable is defined as follows:

$$\log\left(gd_i\right) = \log \frac{Y_i / A_i}{\sum_i Y_i / A_i} = \frac{\log\left[Y_i / \sum_i Y_i\right]}{\log\left[A_i / \sum_i A_i\right]} \tag{11}$$

where Y is GDP and A_i is the area of region i. The relative GDP density of a region is its GDP density relative to the average density of all regions or, equivalently, the ratio of its share of GDP relative to the share of the country's total area. If $\log\left(gd_i\right)$ is equal to zero, region i's GDP share is equal to its area share. If it is larger (smaller) than zero, the region has a concentration of economic activity above (below) the average.

In this section, we explore the geographic dimension of GDP per area for 1,171 NUTS 3 regions of the EU[3]. To be exact, the sample includes information about Austria (35 units), Belgium (44 units), the Czech Republic (14 units), Denmark (10 units), Estonia (5 units), Finland (19 units), France (96 units), Germany (429 units), Greece (42 units), Holland (40 units), Hungary (20 units), Ireland (8 units), Italy (103 units), Latvia (6 units), Lithuania (10 units), Luxembourg (1 unit), Poland (45 units), Portugal (28 units), Slovakia (8 units), Slovenia (12 units), Spain (47 units), Sweden (20 units) and United Kingdom (129 units).

Fig. 1 Choropleth map of relative GDP per Area (1=national GDP/km^2)



Source: Self elaboration

We have used the GDP series proposed by the Cambridge Econometrics database for the year 2006. This distribution is non-normally distributed; i.e. the Jarque-

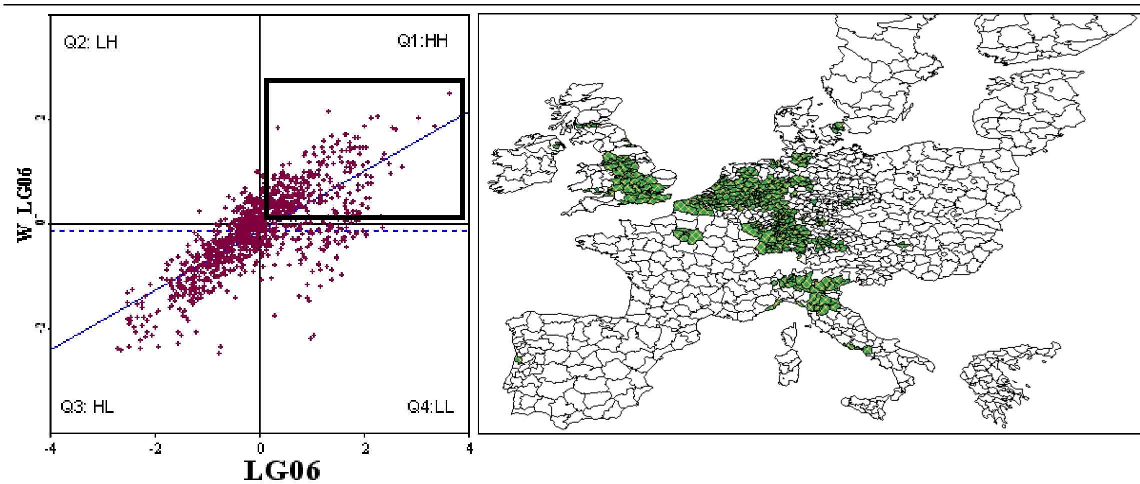
[3] From the total group of 1,232 NUTS 3 existent in the EU-25, we have omitted some units with missing data and the 'islands'; i.e. those regions without any spatial contiguous neighbour. Since Island and Malta belong to this group, actually, only 23 EU countries are represented in our dataset.

Bera normality test (6.18) rejects log-normality with more than 95% of confidence. As shown in Fig. 1, we can find some kind of general spatial trend -spatial autocorrelation- in this variable: from the center (high GDP density) to the peripheral regions (low GDP density), with some exceptions. In effect, log relative GDP per area displays a significant degree of spatial autocorrelation: the magnitude of the Moran's I tests[4] is high (0.57) and significant at $p = 0.001$, which is above its expected value under the null hypothesis of no spatial autocorrelation, $E(I) = -0.0009$. Inference is based on the permutation approach (999 permutations), since this series does not distribute normally (Anselin 1995a).

This result suggests that the production distribution appears to be somewhat clustered in nature. That is, regions with very relatively high/low production density levels tend to be located near other regions with high/low production density levels more often than would be expected as a result of purely random factors. If this is the case, then each region should not be viewed as an independent observation.

Fig. 2 provides a disaggregated view of the nature of spatial autocorrelation for production density by means of a Moran scatterplot (Anselin 1996), which plots the standardized log-relative production density of a region (LG06) against its spatial lag (also standardized), W_LG06. A region's spatial lag is a weighted average of the productions of its neighboring regions, with the weights being obtained from a row-standardized spatial weight matrix (**W**). The four different quadrants of the scatterplot identify four types of local spatial association between a province and its neighbors: HH ('High-High'), LL ('Low-Low'), LH ('Low-High') and HL ('High-Low').

Fig. 2 Moran scatterplot of log relative GDP per Area in 2006 (left). Map with the selection of regions located in Quadrant 1.



Source: Self elaboration

In Quadrant 1, the Moran scatterplot represents those high-GDP density regions that are surrounded by high-GDP density neighbors, which have been highlighted in the

[4] We have specified the spatial weights matrix, **W**, such that each element is set equal to 1 if region *j* has a common border with *i*, and 0 otherwise. Similar results have been observed with other specifications. These include an inverse distance matrix (such that each element w_{ij} is set equal to the inverse of the squared distance between provinces *i* and *j*), and a matrix obtained from a 350 km distance threshold to define a province's neighborhood set (as stated in Rey and Montouri 1999).

map. It can be appreciated that they are all mainly located in the center of the EU and UK. Quadrants 2 and 4 represent negative spatial dependence, while Quadrants 1 and 3 belong to positive forms of spatial dependence. Therefore, the Moran scatterplot reveals the presence of a big cluster of high production density in the EU, the Center, which comprehends from Weser to Seine Rivers, the Rhine basin, Northern Italy and England.

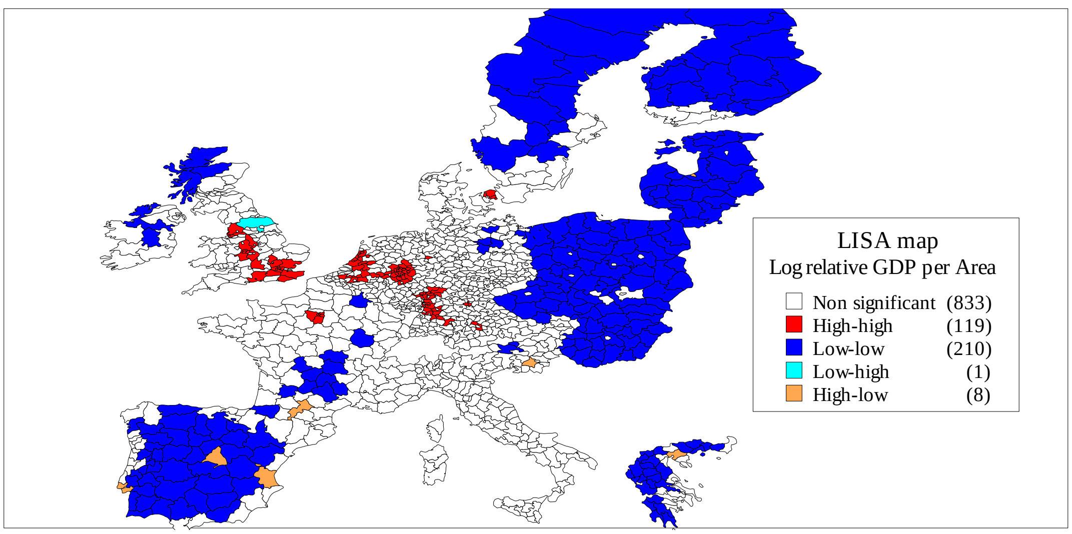
However, the Moran scatterplot does not provide any evidence on the statistical significance of the HH, HL, LH and LL links between one observation and its neighbor. This is the purpose of the Local Moran statistic I_i (Anselin 1995b), which for each region i gives an indication of significant spatial clustering of similar values around that region. A positive value for I_i indicates clustering of similar values (high or low), whereas a negative value indicates clustering of dissimilar values. It takes the following form:

$$I_i = \frac{z_i}{m_2} \sum_{j=1}^{n} w_{ij} z_j \qquad (12)$$

with $m_2 = \sum_{j=1}^{n} z_i^2$; and where z_j is the GDP density in region i (measured as a deviation from the mean value); and w_{ij} is an element of a spatial weights matrix **W**, such that each element is set equal to 1 if region j has a common border with i, and 0 otherwise.

In Fig. 3, a LISA Cluster Map is represented: it is a special choropleth map showing the locations associated with a significant Local Moran statistic classified by type of spatial correlation. The high-high and low-low locations suggest clustering of similar values, whereas the high-low and low-high locations indicate spatial outliers.

Fig. 3 LISA cluster map for log relative GDP per Area in 2006



Source: Self elaboration

The LISA Cluster Map also reveals the presence of a main core inside the central regions, which is mainly placed in the German polycentric urban regions Rhine-Ruhr and Rhine-Main-Neckar[5], as well as Central England and Ile-de-France. However,

[5] Concerning Germany, Roos (2005) found such important differences between the West and East regions that lead him to work only with the former (ignoring the later). This is another proof in favor of the need of considering different subspaces inside Europe when analyzing economic agglomeration.

the most highlighting finding is the presence of an extensive periphery, which is a big cluster of lower production density in the EU. This cluster is divided into a North-East and South Periphery. Though the South Periphery comprehends a more or less uniform area (the European Mediterranean countries), the North-East Periphery is clearly the superposition of two clusters. In effect, as stated in Disdier and Mayer (2004) or Borén and Gentile (2007), Central-Eastern countries constitute a compact area, quite different from Western Europe in terms of agglomeration and metropolitan development. This is why we will split this vast peripheral area into a North Periphery (Baltic countries, Scotland and Ireland) and an East Periphery (Austria and the rest of Central-Eastern Europe).

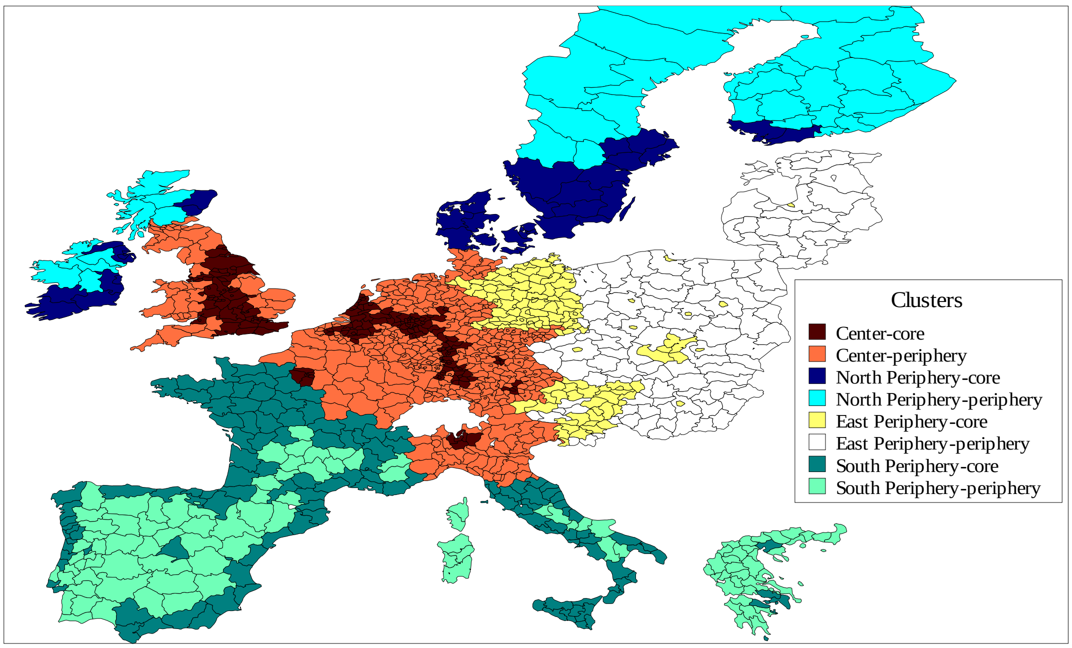
Both Moran scatterplot and LISA cluster map has shown the existence of spatial non-stationarity of NUTS 3 regional economies in the EU, which can be expressed as four main spatial regimes: Center, North Periphery, East Periphery and South Periphery. Nevertheless, inside these spatial regimes we could also distinguish their own core and periphery. In fact, in the 'Eurocenter' we find a Center-Core and the Center-Periphery. The former is depicted by the high-high cluster in Fig. 3 (the German Rhine-Ruhr and Rhine-Main-Neckar, Central England and Ile de France). Immediately closed to the Center-Core megacities, we can find the Center-Periphery, which is constituted by regional growth belts. These 'sunbelts' are networks of regions neighboring big metropolitan areas. Somewhat similarly to the US 'Boomburbs' (Lang and Simmons 2003), we can find in Europe some of these corridors such as the 'Megalopolis England', which is a super urban-agglomeration of 63 metropolitan areas in England and Wales (Hall 1974), or the 'Dorsale', which was the name given by Brunet (1989) to the urban area comprehended between Lancashire and Toscana. Therefore, the EU Center can be roughly described as a huge pentagon circumscribed by London, Paris, Milan, Munich and Hamburg (Sessa 2006).

Additionally, it is known that the peripheries, though suffering from the remote geographical location plus an underdeveloped transport infrastructure, enclose their own regional centers. This is the case of those cities that have good air connections, which has allowed them to develop and attract important ICT industry (e.g. some important Eastern countries capitals, Copenhagen, Helsinki, Dublin, Barcelona and Madrid) or even declining industrial regions, such as Belfast, Glasgow, Clermont-Ferrand, Asturias or Cádiz. To this hypothesis, it is particularly revealing the intuition shown in Cheshire and Hay (1989) about the 'peripheralization of the peripheries' in Europe. In effect, in peripheral regions, a gain in accessibility through a new motorway or rail line brings often more segregation if the new connections open a formerly isolated region to the competition of more efficient or cheaper suppliers in other regions.

In Fig. 3, the LISA cluster map patently illustrates the location of the peripheral subspaces: North Periphery, East Periphery and South Periphery. Nevertheless, a new computation of both Moran's scatterplot and Local Moran's *I* to these peripheral spatial regimes, allows us to detect a new core-periphery structure inside them. The final result is shown in Fig. 4. The North Periphery core is represented by Denmark, Southern Sweden and Finland, Belfast, and Southeastern Scotland and Ireland. The East Periphery core is represented by East Germany, most of Austrian and Slovenian regions and the Eastern metro areas. Concerning the South Periphery, the 'core' lies -with some exceptions- on the coastal edges (the Atlantic coast, Mediterranean Arc and Eastern maritime Greece) whereas the 'periphery' stands in the interior center of these countries (mainly rural hinterlands).

Exploratory Spatial Data Analysis (ESDA) has shown the existence of spatial autocorrelation as well as a spatial hierarchical polarization of economic agglomeration in the European NUTS 3 regions. This spatial heterogeneity -spatial instability- is articulated as four extensive spatial regimes which are also sub-divided into a core-periphery area leading to a final set of 8 clusters. In summary, spatial autocorrelation and spatial heterogeneity are two effects that must be tested when modeling GDP density since they could lead to biased results if they are not adequately taken into account.

Fig. 4 Map with the selection of the EU regional clusters



Source: Self elaboration

3.2. First and second nature indicators

Next, we define some good indicators to measure first and second nature effects. About first nature, we are interested in those geographical characteristics that are related to the distribution of economic activity. In general, this is the case of natural endowment, physical geography, relative location and political geography. Examples of **natural endowment** positively related to GDP density are agriculture, minerals, natural resources, good soil and water supply (Gallup et al. 1999, Rapaport and Sachs 2003, Dobado 2004, Ayuda et al. 2005, Roos 2005). In order to measure this feature, we have chosen the following indicators (see in Table 1 a full description of the variables): the average of the daily precipitation amount during the XX Century (*rain*) and three dummy variables for the presence of agricultural areas (*agric*), water bodies (*water*) and mineral extraction sites (*miner*). In this context, dummy variables are preferred to production quantities because we want to measure the exogenous endowment not the endogenous output. The first three variables can be considered as good proxies for agricultural potential. Nevertheless, despite its importance for most Mediterranean regions, agriculture is actually a small sector in the overall European economy. This is why it is not *a priori* clear either the statistical significance or the sign for the relationship between these indicators and economic agglomeration. Regarding the mineral extraction sites variable, we expect a strong positive effect on GDP density since the EU mineral industry is a considerable producer for both the EU domestic and export markets.

In addition, Gallup and Sachs (1998), Rappaport (2000) and Limão and Venables (2001) include certain kind of **physical geography** indicators, such as altitude, distance to the coast and waterways, lying to the seashore (or being landlocked), navigable rivers and climate. On our side, we have considered the average altitude in a region (*altit*), two dummy variables to indicate if a region lies at the seashore (*coast*) and have a navigable river (*navriv*), and three climate variables, such as mean temperature (*tave*), cloud cover (*cloud*) and sunshine (*sunsh*). As a general rule, we could expect a positive relationship between waterways (seashore and navigable rivers) and economic agglomeration, though negative values for high altitudes.

Table 1. Variable list for the NUTS 3 EU regions

Variable	Description	Units	Font	Period
gd	GDP per Area	Euros/sq. m.	Cambridge Econometrics Ltd.	1991-2006
rain	Daily Precipitation amount	mm	ECA&D and self elaboration**	XX Century*
agric	Agricultural area	0-1	LANMAP2 Dataset	2004
water	Water bodies site	0-1	LANMAP2 Dataset	2004
miner	Mineral extraction site	0-1	LANMAP2 Dataset	2004
altit	Altitude or elevation	Meters	LANMAP2 Dataset	2004
tave	Daily mean Temperature	ºC	ECA&D and self elaboration**	XX Century*
cloud	Daily Cloud cover	Octas	ECA&D and self elaboration**	XX Century*
sunsh	Daily Sunshine	# hours	ECA&D and self elaboration**	XX Century*
navriv	Navigable river	0-1	Self elaboration	-
coast	Maritime limit	0-1	Self elaboration	-
distm	Average Distance to all other regions	km	Self elaboration	-
xcoo	Longitude (X-coordinate)	grades	Self elaboration	-
ycoo	Latitude (Y-coordinate)	grades	Self elaboration	-
capreg	NUTS 2 regional capital	0-1	Self elaboration	-
border	Border region	0-1	Self elaboration	-
Pop	Population	people	Cambridge Econometrics Ltd.	1991-2006
Prod	GDP per employee	Euros	Cambridge Econometrics Ltd.	1991-2006

* Average of the period, ** Interpolation of the ECA&D original variables, *LANMAP2* Landscape of Europe Project (Klein Tank et al. 2002), *ECA* European Climate Assessment & Dataset.

Location is another geographical feature affecting agglomeration, which has been represented as relative distance to core -or other- regions or simply by the latitude-longitude Earth coordinates. Joint to the later (*xcoo*, *ycoo*), we have also considered the relative location of a region, which has been measured by the average distance to all other regions (*distm*). As we will prove further, long relative distance, southern and eastern latitudes are negatively related to economic agglomeration, in general terms.

Political geography has also been highlighted by Mathias (1980), McCallum (1995) and Roos (2005) who consider that agglomeration is positive or negatively affected by containing a capital city or being a border region, respectively. In this case, similarly to the German regions (Roos 2005), we have considered the NUTS 2 regional capitals (*capreg*) since they usually concentrate a lot of legislative and executive power and have better access to information about regional government investment and decision plans (Ades and Glaeser 1995; Funck 1995; Ayuda et al. 2005). In addition, we have also included a dummy for border regions (*border*) due to the still important differences existent among European countries in terms of language, culture and institutions.

The ECA&D climate four variables have been interpolated since they are only available from a reduced number of monitoring stations. As shown in Table 2, we have applied kriging, selecting the procedure with less mean square error (MSE).

Table 2. Interpolation methods

Variable	Description	Interpolation method (least MSE)	Monitoring stations
rain	Daily Precipitation amount	Universal Kriging	303
taver	Daily mean Temperature	Ordinary Kriging	235
cloud	Daily Cloud cover	Ordinary Kriging	76
sunsh	Daily Sunshine	Universal Kriging	85

Concerning second nature (man-made agglomeration economies), we have followed Roos (2005) because it allows us to make better comparisons with this case. He chose total population (*pop*) and labor productivity (*prod*) since on aggregate levels both variables can capture many agglomeration economies, i.e. informational spillovers and labor market economies. Population could be considered as an indirect measure of agglomeration economies. In effect, as stated in Henderson (1988) if agglomeration economies exist in an area, labor productivity should rise in the level of population (employment). Other indicators, such as population density (proposed in Gallup et al. 1999), provide not so clear relationship with GDP density (e.g. some densely/sparsely populated areas are rich whereas others are poor, which are the cases of Western Europe/New Zealand and Indonesia/African Sahel, respectively).

4. GEOGRAPHY AND THE LOCATION OF PRODUCTION IN THE UE

In this chapter, we apply the ANOVA methodology proposed in Roos (2005) for German regions in 2000. In our case, we present a static analysis for 2006 testing for not only endogeneity but also spatial effects in the residuals. As stated before, it is a four-step analysis that proceeds as follows: 1) we filter gross second nature indicators from first nature interrelations; 2) we estimate how much of GDP per area variance can be explained by gross (V_s+V_{fs}) and net (V_s) second nature advantages; 3) we estimate how much of GDP per area variance can be explained jointly by gross first and second nature (V_f+V_s+V_{fs}); and 4) we calculate the difference between the result in step three and two, which is the importance of first nature alone (V_f).

4.1. Filtering gross second nature from first nature elements

In order to disentangle empirically the second nature variables (population and GDP per worker) from first nature interactions, we proceed to regress them on geography and take the residuals as variables of net second nature forces (see Eq. 2). The regressions of population and productivity on the complete set of 15 first nature variables lead to high multicollinearity what inflate the determination coefficients. To avoid this problem, we opted for group 11 physical geography variables (excluding mineral extraction sites, navigable rivers, regional capitals and border regions) with factor analysis[6]. The rotated factors can be interpreted as follows: Factor 1 is a variable of Southern Latitude (direction North to South), which is called *FSLAT*. It contains high scores of mean temperature and sunshine (positive), cloud cover and latitude (negative). Factor 2 is called *FSEA* and it is mainly related to maritime limit and average distance to all other regions. Factor 3 is a variable of latitude, which is called *FALT*. It contains high scores of altitude and precipitation (positive), as well as agriculte (negative). Regarding Factor 4 (*FWEST*), it is mainly based on Western Latitude (East-West orientation), though it also contains high scores of water bodies and precipitation. The regressions of population and GDP per worker on the 4 geography factors, as well as mineral extraction sites (*miner*), navigable rivers (*navriv*), regional capital cities (*capreg*) and border regions (*border*) show much lower multicollinearity numbers (3.89 and 2.25, respectively), well below the acceptable limit of 20/30 (Anselin 1995a). Table 3 presents the results of the final regressions of the second nature variables on first nature.

Table 3. Second nature on first nature OLS regression results

Dependent variable	Log(*pop*)		Log(*prod*)	
		without *capreg*	2006	without *capreg*
constant	2.15^{***}	2.15^{***}	1.63^{***}	1.64^{***}
FSLAT	0.10^{***}	0.11^{***}	-0.05^{***}	-0.05^{***}
FSEA	0.06^{***}	0.07^{***}	-0.09^{***}	-0.08^{***}
FALT	-0.05^{***}	-0.04^{***}	0.03^{***}	0.03^{***}
FWEST	0.08^{***}	0.08^{***}	0.05^{***}	0.05^{***}
miner	0.30^{***}	0.35^{***}	-	-
navriv	0.14^{***}	0.17^{***}	-0.07^{***}	-0.06^{***}
capreg	0.27^{***}	-	-0.04^{**}	-
border	0.05^{**}	0.06^{***}	-0.15^{***}	-0.15^{***}
R2	0.43	0.38	0.27	0.27
Multic. #	3.89	3.66	2.25	2.10
Net 2nd	pi_{06}	-	del_{06}	-

*** significant at 0.01, ** significant at 0.05, * significant at 0.1, *Log(pop)* log population, *Log(prod)* log labor productivity, *Multic. #* multicollinarity number, *pi*, *del* residuals of Eq. 2.

Measured by R^2, the fit of both population and labor productivity equations is good, even higher -in the case of productivity- than that found in Roos' application for Germany (43% for population and 27% for productivity). In the population equation, the capital dummy has the largest influence (86%)[7] whereas this variable has much less power and is negatively related with labor productivity (-9%). It is as if regional capital cities are -in general- more capable of attracting people at cost to productivity.

Since there is a difference in the degree of exogeneity between the variables of the physical and the political geography, we have run regressions with and without the capital dummy (*capreg*). In fact, although the location of capitals does not change very often, they do change and are possibly the result of underlying economic forces. Cities

[6] The excluded variables are the ones less correlated with the others. Factors have been extracted using principal components and rotated with Varimax method.

[7] In semi-logarithmic equations, the dependent variable changes by [exp(b)-1]·100 percent if the explanatory variable changes from zero to one unit, where b is the explanatory variable coefficient.

might be capitals because they are economically important. Nevertheless, notice that excluding this variable hardly alters the regressions goodness-of-fit.

We find significant relations between second nature and geography what allow us to conclude that both forces interact. Therefore, we have filtered the residuals of these 10 regressions, *pi*, *del*, which will be considered as net second nature forces.

4.2. First and second nature effects on GDP per Area

In this step, we compute second nature effects on GDP per area with the estimation of two equations. Firstly, we regress the log-relative GDP per area on population and labor productivity. The resulting determination coefficient will indicate the second nature *gross* effect $R_{gs}^2 = (V_s + V_{fs})/V$. Secondly, the second nature *net* effect on GDP per area is obtained from the estimation of this variable on the residuals, *pi*, *del*, derived from the last estimations, with the help of the corresponding determination coefficient $R_{ns}^2 = V_s/V$. Finally, we estimate how much of GDP per area variance can be explained jointly by gross first and second nature (V_f+V_s+V_{fs}). As in Eq. 8, we include a set of first nature indicators together with the net second nature variables (*pi*, *del*) as regressors. The joint importance of first and second nature is then measured by $R_{f+s}^2 = (V_f + V_{fs} + V_s)/V$.

As stated in Roos (2005), one problem is that the second nature variables are endogenous and simultaneously determined with GDP. This might lead to the well-known simultaneity bias in the regressions violating the necessary conditions to obtain estimates with good properties. The instrumental variables estimation is the standard approach to overcome the consequences of simultaneity, i.e. biasness, inefficiency and inconsistency on OLS-estimators[8]. In our case, in order to decide whether we need IV estimation, we have first analyzed the potential system feedbacks between the dependent variable, log-relative GDP per Area, and the four second nature explanatory variables, i.e. population, labor productivity and the OLS residuals (*pi*, *del*) found in Table 3 estimations. For this purpose, we have used the *Durbin-Wu-Haussman* (DWH) test, which is an 'exogeneity test' (Anselin 1999) that compares the IV and OLS estimates assuming the former are consistent[9]. Since we need to estimate IV equations to perform this test, we must decide before the set of adequate instruments for each potential stochastic regressor. As already stated, they should be correlated to the original endogenous variables but asymptotically uncorrelated to the error term.

[8] The principle of the IV estimation is based on the existence of a set of instruments that are strongly correlated to the original endogenous variables but asymptotically uncorrelated to the error term. Once these instruments are identified, they are used to construct a proxy for the explanatory endogenous variables, which consists of their predicted values in a regression on both the instruments and the exogenous variables. However, it is very difficult to find such instruments because most socioeconomic variables will be endogenous as well. In the standard simultaneous equations framework, the instruments are the 'excluded' exogenous variables.

[9] This test reports the confidence level at which consistency of OLS estimates can be rejected. In fact, it is an F test with (k^*, n–k–k^*) degrees of freedom on the null hypothesis of exogeneity of a k^* subset of the total k explanatory variables, with n as the number of observations (for technical issues, see Davidson and McKinnon 1993). As shown in Anselin (1999), DWH test is consistent with spatially autocorrelated OLS residuals.

Roos proposes to use mainly time-lagged variables as instruments, since they are highly correlated with the actual variables but also non-contemporary correlated with the errors[10]. Besides, we have also considered other space and/or time lagged second nature variables as well as 'excluded' first nature explanatory variables. In all cases, we have selected only those instruments more correlated with the corresponding endogenous regressor but less correlated with OLS error terms. The goodness of the instruments can be proved with the help of the Sargan test, which contrasts the null hypothesis that a group of *s* instruments of *q* regressors is valid[11]. In Table 4, we have shown the instruments definitely used in each equation, the adjusted R^2 of the corresponding endogenous variable on each set of instruments, as well as the results of the *Sargan* and *Durbin-Wu-Haussman* (DWH) tests.

Table 4. Instruments and endogeneity tests in second nature effect regressions

		Instruments	Adjusted R^2 instruments	Sargan	DWH
Gross second (Eq. 3, Table 5)					
Log (*pop*)	2006	*wlpo06, pi06, falt, miner, capreg*	0.90	0.00	77.93***
Log (*prod*)	2006	*wlpr06, del06*	0.92	0.00	81.10***
Net second (Eq. 5, Table 5)					
pi	2006	*wpi06, lpo06, lpo05, lpo04*	0.61	0.00	93.38***
del	2006	*wdel06, lpr06, lpr05, lpr04*	0.84	0.00	121.11**
First & second nature joint effect (Eq. 8, Table 5)					
pi	2006	*wpi06, lpo06, lpo05, lpo04*	0.61	0.00	0.17
del	2005	*wdel06, lpr06, lpr05, lpr04*	0.84	0.00	0.02

Log(pop) log population, *Log(prod)* log labor productivity, *pi* residual of the regression of log population on first nature variables, *del*: residual of the regression of log labor productivity on first nature variables, *falt* latitude, *miner* mineral extraction sites, *capreg* regional capital cities, *DWH* Durbin-Wu-Haussman exogeneity test, *** significant at 0.01, ** significant at 0.05, * significant at 0.1.

Results show a high degree of simultaneity in all the second nature regressors with respect to log-relative GDP per area, with the exception of net second nature variables (*pi*, *del*) in Eq. 8. As a consequence, both Eqs. 3 and 5 must be estimated by IV whereas Eq. 8 can be estimated by OLS. In Table 5, we show the estimation results of Eqs. 3, 5 and 8. Specifically, in Eq. 3 and 5, log-relative GDP per Area is regressed on gross and net second nature variables, respectively. We have computed the so-called asymptotic *t*-tests as a ratio of the estimate to its asymptotic standard error.

As stated in Anselin (1988, pp. 244), in the IV estimation approach the residuals have a zero mean, so than the standard variance decomposition can be obtained and a determination coefficient can be computed in the usual manner (the ratio of the variance of the predicted values over the variance of the observed values for the dependent variable). Consequently, the regressions on gross/net second nature variables provide a determination coefficient (R^2_{gs}, R^2_{ns}) equal to 0.33 and 0.25, respectively, which is the

[10] Non-contemporary dependence between regressors and the error terms lead to asymptotically unbiased estimators only in absence of temporal autocorrelation. However, in our case it is difficult to suppose time independence between the error terms what could somewhat affect our results.
[11] This is a Chi-2 test with (*s*–*q*) degress of freedom that rejects the null when at least one of the instruments is correlated with the error term (Sargan 1964). In our case, we can clearly accept the null with a confidence level of 0.99. All the computations can be obtained upon request from the authors.

share of GDP density variance that is explained by gross and net second nature effects. Regarding the mixed effect of the interaction between first and second nature on GDP density (R^2_{fs}), it can be extracted as the difference between R^2_{gs} and R^2_{ns} (Eq. 7). These results are remarkably lower in comparison with those obtained by Roos for Germany. In effect, Roos found that a 65% of German GDP density in 2000 was caused by gross second nature, which can be decomposed into a mixed-indirect effect (29%) and a net-direct effect (36%).

Table 5. Regression results of GDP per area on second and first nature variables

	Gross 2nd nature	Net 2nd nature	First & second nature joint effect	
				Without *capreg*
Equation	(a)	(b)	(c)	(d)
Estimation	IV	IV	OLS	OLS
Constant	-3.87***	0.11***	-0.09***	-0.09***
FSLAT	-	-	-0.16***	-0.15***
FSEA	-	-	-0.26***	-0.25***
FALT			-0.001	0.001
FWEST	-	-	0.14***	0.13***
miner	-	-	0.50***	0.51***
navriv			-0.05	-0.02
capreg	-	-	0.23***	-
border	-	-	-0.46***	-0.45***
Log(*pop*)	0.57***	-	-	-
Log(*prod*)	1.65***	-	-	-
pi	-	0.88***	0.23***	0.23***
del	-	1.81***	1.27***	1.28***
R-squared	0.33	0.25	0.53	0.52
Multicollinearity #	21.97	1.22	3.90	3.66
Heter. tests K-Basset	28.89***	3.71	150.55***	146.48***
Heter. tests White	52.84***	77.95***	313.24***	297.41**
Spatial Chow test	1266.93***	1709.09***	18.84***	18.95***
Kelejian-Robinson	647.10***	580.40***	796.30***	689.33***
LM (spatial lag)	436.04***	507.12***	270.74***	248.30***
LM (spatial error)	467.57***	545.96***	330.19***	285.95***

Log(pop) log population, *Log(prod)* log labor productivity, *pi* residual of the regression of log population on first nature variables, *del* residual of the regression of log labor productivity on first nature variables, *K-Basset* Koenker-Bassett heteroskedasticity test, *Spatial Chow test* spatial Chow test for 6 clusters, *LM* Lagrange Multiplier test for spatial autocorrelation, ** significant at 0.01, * significant at 0.05

The final lines of diagnostics in Table 5 reports three asymptotic tests for spatial autocorrelation[12] (Anselin 1999), which are highly significant. In addition, we have also tested for spatial heterogeneity in the errors, in the form of eight subspaces, as detected before for GDP density distributions (Fig. 4). For this purpose, we use the spatial Chow test proposed by Anselin (1990)[13]. In Table 5, the null hypothesis on the joint equality

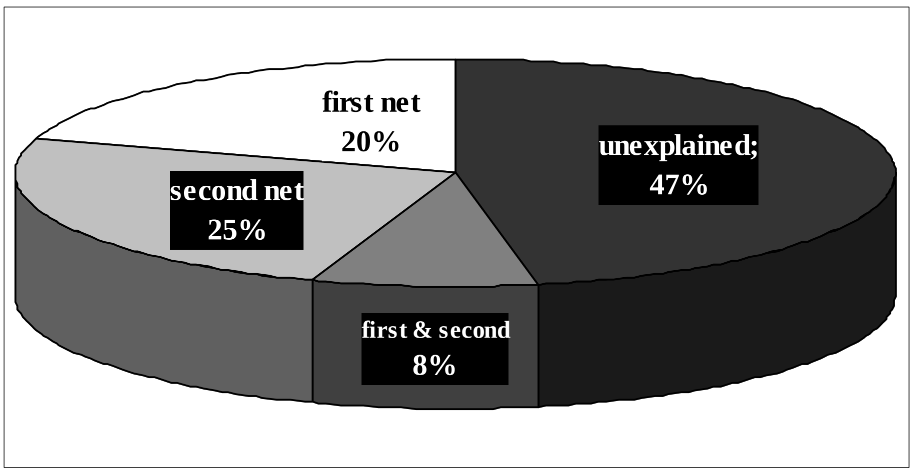
[12] The spatial weight matrix is specified as in footnote 7.

[13] In this test, the null hypothesis states that the coefficients are the same in all regimes. It is based on an asymptotic Wald statistic, distributed as a χ^2 distribution with $[(m–1)\cdot k]$ degrees of freedom (m being the number of regimes).

of coefficients is clearly rejected by the Chow-Wald test in all the regressions. Therefore, when regressing GDP density on first and second nature variables, both spatial effects are present demonstrating the existence of non-randomness in the error terms of every equation. It is known that sometimes, spatial autocorrelation in the residuals may be induced by a strong spatial heterogeneity that is not correctly modeled by spatial dependence specifications (Brunsdon et al. 1999). This is something that will be tested hereafter.

When measuring how much of GDP per area variance can be explained jointly by gross first and second nature (V_f+V_s+V_{fs}), we find that all coefficients of Equations (a) and (b) are significant except in altitude (*FALT*) and navigable river (*navriv*) variables. Results show the great importance of net second nature variables (population and productivity) on GDP density. Among physical geography, the variable *miner* (mineral extraction sites) has the largest influence (216%). Regional capital is also a very influential variable (70%), though its omission does not alter significantly the results.

Fig. 5 ANOVA results for log relative GDP per Area in 2006



If we calculate the difference between the determination coefficient in Equation (c) and (a), in Table 6, we obtain the importance of first nature alone (V_f) for the whole Europe: $V_f / V = R^2_{f+s} - R^2_{gs}$. In Fig. 5 we show the complete ANOVA decomposition. The total variation that can be assigned to the net effect of first nature ascends to 20%, which is almost three times that found by Roos' for Germany (7.1%); i.e. after controlling for agglomeration economies and the interaction effect of first-second nature, the net influence of natural geography is 20%. This could support the idea that the complete continent -Europe- is a much more heterogeneous area with very different climatic zones, which there are places more or less favorable to live in. On the other hand, notice that the important variance reserved to unexplained factors (almost a 50% of GDP density) could be the result of the unattended spatial effects present in the error terms. In it well-known that omitting substantive spatial autocorrelation and/or heterogeneity usually biases the regression results. This is why we next proceed to test and correct for these effects in order to know the real influence -in Europe- of geography on GDP density.

5. SPATIAL INSTABILITY IN THE INFLUENCE OF GEOGRAPHY ON AGGLOMERATION

Consequently, in order to capture the polarization pattern previously observed in the distribution of GDP density among the European NUTS 3 regions, we allow cross-region parameter variation. We estimate a *spatial regimes model*[14] by IV (for equations 3 and 5) and OLS (for equation 8), with the eight subspaces depicted in Fig 4. In Table 7, we show the estimation results of log-relative GDP per Area on gross and net second nature variables (Eqs. 3 and 5), respectively.

Table 6. Regression results of the spatial regimes models of GDP per area on gross second nature

Spatial cluster (IV estimation)	Center		North Periphery		East Periphery		South Periphery	
	Core	Periph.	Core	Periph.	Core	Periph.	Core	Periph.
Observations	203	432	35	33	128	106	126	108
	Regressions on gross second nature variables							
Constant	-1.57***	-1.97***	-3.32*	-3.05***	-1.82***	-4.08***	-0.62**	-2.46***
Log(*pop*)	0.23**	0.14**	0.90*	0.88***	0.72***	0.49***	0.88***	0.68***
Log(*prod*)	1.17***	1.15***	0.59	0.06	0.19	0.97***	0.95***	0.65***
R-squared	0.12	0.08	0.18	0.24	0.21	0.73	0.58	0.63
LM (sp.er.)	0.19	8.73***	8.43***	3.70*	2.50	15.32***	26.50***	16.12***
	Regressions on net second nature variables							
Constant	0.95***	0.23***	-0.07	-0.91***	0.15***	-0.37***	-0.14***	-0.60***
pi	0.22*	0.05	0.28	0.35*	0.83***	1.09***	1.40***	0.83***
del	1.20***	1.54***	1.59	0.26	0.24	1.77***	0.01	0.69***
R-squared	0.10	0.06	0.05	0.12	0.20	0.68	0.49	0.55
LM (sp.er.)	0.36	10.18***	6.32**	2.06	1.38	4.18**	2.05	8.66***

Log(pop) log population, *Log(prod)* log labor productivity, *pi* residual of the regression of log population on first nature variables, *del* residual of the regression of log labor productivity on first nature variables, *LM (sp.er.)* Lagrange Multiplier test for spatial error autocorrelation *** significant at 0.01 ** significant at 0.05 * significant at 0.10.

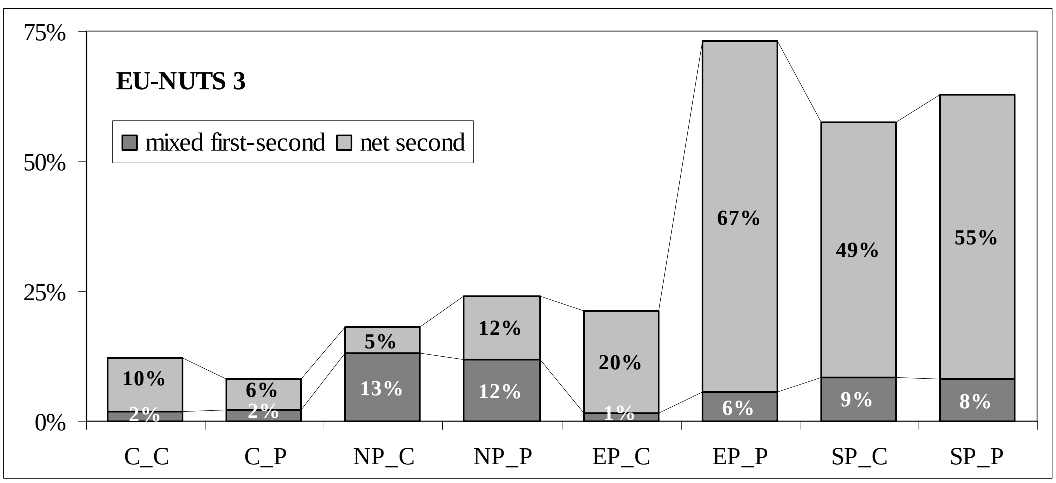
Whereas in the global estimation R^2_{gs} is equal to 0.33 (see Table 5), the spatial regimes regressions of GDP per area on gross second nature variables provide an R^2 that ranges from 0.08 (Center-periphery) to 0.73 (East Periphery-periphery). This result confirms our initial hypothesis about the importance of taking into account spatial instability in GDP density distributions. As regards net second nature elements, the share of GDP density variance that is explained for the complete set of European regions is 0.25, though it varies from 5-10% in the Center and North Periphery-core to 49-67% in the South and East Periphery-periphery.

[14] We have also estimated a groupwise heteroskedastic error model (GHE). In general, both GLS and LM estimations produce significant variance coefficients in each subspace, but cannot absorb all the heteroskedasticity and spatial dependence still present in the residuals; i.e. their performance is worse than the spatial regimes model. In addition, it must be said that both GLS and LM estimations produce determination coefficients (R^2) which are not comparable with the calculated in OLS/IV. Therefore, the ANOVA methodology proposed in this paper could not be applied with GHE models.

In general terms, second nature forces exert the highest influence in the East and South Periphery (Fig. 6). It is particularly highlighting the scarce effect across the so-called Western European regions (Center-North Periphery-core clusters). Maybe the exceptionally high degree of investment and development in this area motivates a less impact of relative changes in man-made agglomeration economies. This reasoning is also in accordance with the superior effect of second nature variables in the peripheral regions. Note that our results for Western Germany (which is inside the Center cluster) are remarkably lower in comparison with those obtained by Roos for Germany (65%).

These differences could be explained by the fact that Roos worked with the 97 planning regions of West Germany (which are in between the levels NUTS 3 and NUTS 2), while we use a group of 345 West Germany NUTS 3. It could easily be a good example of what is called the "ecological fallacy" –in social sciences- or modifiable unit problem (MAUP), in geography. It occurs when inference based on data aggregated to a particular set of geographical regions changes if the same data are aggregated to a different set of geographical regions (Arbia 1989). This phenomenon is particularly evident when spatial heterogeneity (non-stationarity) is present (Peeters and Chasco 2006), which is the case here.

Fig. 6 Impact of second nature forces on GDP in the 8 EU regional clusters



C_C Center-core, *C_P* Center-periphery, *NP_C* North Periphery-core, *NP_P* North Periphery-periphery, *EP_C* East Periphery-core, *EP_P* East Periphery-periphery, *SP_C* South Periphery-core, *SP_P* South Periphery-periphery.

Pertaining to the interaction effect of physical geography and agglomeration economies, it registers lower values than net second nature, with the exception of the North-Periphery regions (12-13%). Anyway, these values are three times lower than Roos' results (37%). This result shows the greater importance -in Northern Europe- of the interaction between economic agents and first nature as determinants of GDP per Area. It seems logical since in Northern latitudes, physical geography can constitute a severe barrier for human settlement and economic activity.

In Table 6, the final lines of diagnostics for second nature regressions report the asymptotic LM test for remaining spatial error autocorrelation. Notice that it provides lower values than in the global case (Table 6), though in some cases, they are still quite significant (what could affect the estimator efficiency). This event points out the need of an explicit quantification of spatial externalities in the influence of geography on agglomeration. This is what we tackle next.

In Table 7, we present the OLS estimation results of Eq. 8, which show how much of GDP per area variance can be explained jointly by gross first and second nature (V_f+V_s+V_{fs}**)**. We include as regressors the same set of first nature indicators than in Table 5, together with the net second nature variables (*pi*, *del*). The joint importance of first and second nature is then measured by $R^2_{f+s} = \left(V_f + V_{fs} + V_s\right) / V$. With the exception of the East-Periphery-periphery cluster, in all the cases there are some remaining spatial dependence in the errors. One way to both capture this effect and measure its importance on GDP density is estimating a mixed cross-regressive spatial model, which is defined as follows:

$$\log\left(gd_i\right) = \alpha_0 + \sum_{k=1}^{K} \phi_k f_{ki} + \sum_{m=1}^{M} \phi_m \hat{s}_{mi} + \sum_{l=1}^{L} \phi_l W r_{li} + \varepsilon_i \qquad (13)$$

where variables $r_{li} = \left\{ f_{ki}, \log\left(pop_i\right), \log\left(prod_i\right), \hat{\pi}_i, \hat{\delta}_i \right\}$ are the spatially lagged first and second nature variables. Since they are all exogenous (even the second nature as shown in Table 4), this model can also be estimated by OLS and its R^2_T is comparable with the estimated in the rest of regression (Anselin 1988). The difference between R^2_T and R^2_{f+s} could be interpreted as the influence than spatial externalities of geography exerts on agglomeration. We have only selected those regressor spatial lags that being statistically significant imply a reduction of the remaining spatial autocorrelation in the errors.

All coefficients have the expected signs. Results show the greater importance of minery extraction sites, regional capitals and net productivity, since they are positively significant in almost all the spatial clusters. Nevertheless, there are some noteworthy differences among the spatial clusters. Among physical geography, Southern latitude (*FSLAT*) is only (negatively) significant in the North and South Peripheries, though with positive sign in the very North regions. The existence of navigable rivers in a region (*navriv*) is mainly determinant across the Center regions (Rhine-Ruhr, Rhine-Main-Neckar, Seine and Thames) whereas in the North Periphery there is no navigable river (this variable has been excluded from their corresponding regressions). Regarding the Western longitude (*FWEST*), it has a diverse influence on agglomeration depending on the spatial orientation; i.e. it is significantly negative in both Center-core and South Periphery but positive in the North and East Periphery cores. A similar opposite effect exerts the existence of maritime limit (*FSEA*), which can also be interpreted as a relative location (mean distance to the rest of the regions) variable. It is positive in the peripheral North Periphery but also in the South Periphery-core, which highlights the importance -for these regions- of lying on the Baltic or the Mediterranean seas. Nevertheless, either the Atlantic or the Baltic seem to have negative effects in the peripheral East and Center regions, though it is probably due to the isolation that represents -in these areas- the vicinity to the coast.

As already stated, regional capital (*capreg*) is also a very influential variable with the exception of the North Periphery most likely because of two reasons. On the one hand, in some countries (e.g. United Kingdom and Ireland) the NUTS 3 division did not coincide with the administrative structure what led to an artificial construction of these regions by aggregating smaller pre-existing units. On the other hand, in some countries, like Finland, the fundamental administrative divisions are not the regions but the municipalities, which account for half of public spending.

Table 7. Regression results of the spatial regimes models of GDP per area on first and second nature

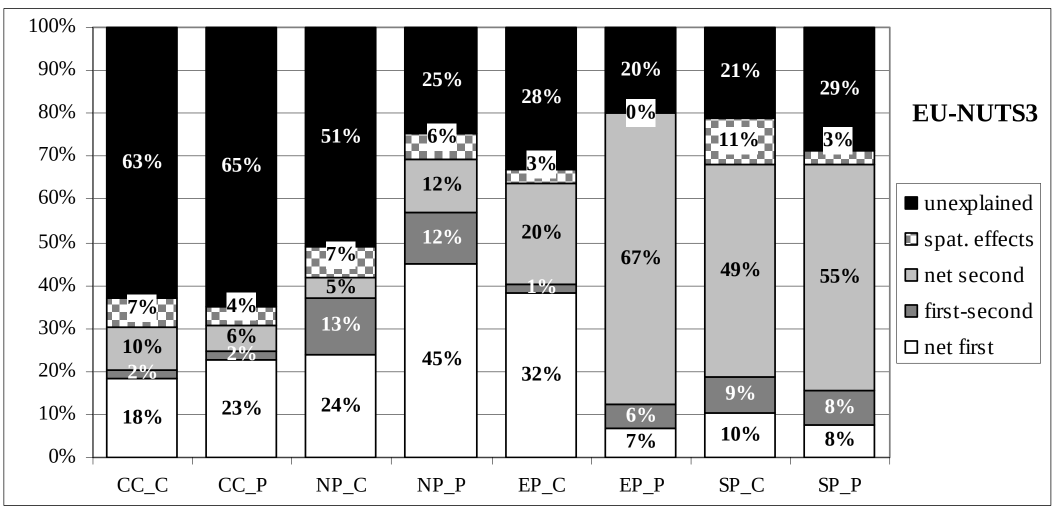
Spatial cluster (IV estimation)	Center		North Periphery		East Periphery		South Periphery	
	Core	Periph.	Core	Periph.	Core	Periph.	Core	Periph.
Observations	203	432	35	33	128	106	126	108
Constant	0.88^{***}	0.24^{***}	-2.14^{**}	0.68	-0.11	-0.50^{***}	-0.11	-0.45^{***}
FSLAT	-0.10	-0.05	-1.74^{**}	0.82^{***}	0.00	-0.01	-0.06^{**}	-0.10^{***}
FSEA	0.12	-0.11^{***}	-0.52	0.33^{**}	0.01	-0.24^{***}	0.09^{***}	0.01
FALT	0.10^{*}	0.01	-0.16	-0.15	0.00	-0.01	-0.12^{***}	-0.02
FWEST	-0.22^{***}	0.05	0.31^{*}	-0.03	0.23^{**}	0.01	-0.12^{***}	-0.04^{**}
miner	0.41^{***}	0.39^{***}	0.69^{*}	0.13	0.47^{***}	0.17^{***}	0.25^{***}	0.26^{***}
navriv	0.14^{*}	-0.13^{**}	-	-	-0.13	-0.01	0.09	0.18^{*}
capreg	0.36^{***}	0.20^{***}	0.58^{*}	0.14	0.55^{***}	0.16^{***}	0.15^{***}	0.14^{***}
border	-0.39^{***}	-0.27^{***}	-0.44	-0.06	-0.35^{***}	-0.16^{***}	0.00	-0.14^{***}
pi	0.16	-0.17^{**}	0.29	0.94^{***}	0.20	0.39^{***}	0.87^{***}	0.69^{***}
del	0.17^{**}	1.39^{***}	3.90^{*}	-0.42	1.51^{***}	1.07^{***}	0.73^{***}	0.46^{**}
WFSLAT	-	-	-	0.74^{**}	-	-	-	-
WFALT	-	-	0.81^{*}	-	-	-	0.19^{***}	-
WFWEST	-	-		-	0.30^{*}	-	-	-
Wnavriv	-1.02^{**}	-	-	-	-	-	-	-0.63^{**}
Wcapreg	-	-2.65^{***}	-	-	-	-	-0.33^{**}	-
Wdel	-	-	-	-	-0.74^{**}	-	-1.19^{***}	-
R2 (+sp. eff.)	0.37	0.35	0.42	0.75	0.57	0.80	0.79	0.74
Spatial eff.	0.07	0.04	0.07	0.06	0.03	0.00	0.11	0.03
Jarque-Bera	1.80	19.0^{***}	12.8^{***}	2.25	8.04^{**}	9.85^{***}	1.86	1.75
Multicol. #	17.53	8.41	33.97	46.00	10.73	11.40	11.71	11.67
KB/*BP* tests	*14.35*	35.29^{***}	16.13^{*}	*8.61*	30.97^{***}	32.37^{***}	*14.35*	*10.24*
Kelejian-R.	-	7.60	4.67	-	3.95	5.70	-	-
LM-lag	1.66	-	-	2.17	-	-	4.05^{**}	8.37^{***}
RLM-lag	-	-	-	-	-	-	0.08	1.45
LM-err	0.00	-	-	0.20	-	-	8.33^{***}	9.75^{***}
RLM-err	0.00	-	-	0.20	-	-	4.36^{**}	2.83^{*}

FSLAT Southern latitude factor, *FSEA* maritime limit (or relative average distance) factor, *FALT* altitude factor, *FWEST* Western latitude, *miner* mineral extraction sites, *navriv* navigable rivers, *capreg* regional capitals, *border* border regions, *pi* residual of the regression of log population on first nature variables, *del* residual of the regression of log labor productivity on first nature variables, *R2 (+sp.eff)*: R^2 of the mixed spatial cross-regressive models, *Multicol. #* multicollinearity number, *KB/BP tests* Koenker-Bassett/ Breusch Pagan heteroskedasticity tests, *Kelejian-R* Kelejian-Robinson test for spatial error autocorrelation (robust to non normality in the errors), *LM-lag* Lagrange Multiplier test for spatial lag autocorrelation, *RLM-lag* LM test for spatial lag autocorrelation robust to the presence of spatial error autocorrelation, *LM-err* LM test for spatial error autocorrelation, *RLM-err* LM test for spatial error autocorrelation robust to the presence of spatial lag autocorrelation, *** significant at 0.01, ** significant at 0.05, * significant at 0.10

Concerning net first nature effect, it plays a main role in the Center, North Periphery and East Periphery-core, whereas it is smaller in the South Periphery and the peripheral East Periphery (Fig. 7). In some sense, there are some similarities between these two macro-clusters. In the traditional Western Europe plus the East Periphery-core (mainly Austria, Slovenia, Eastern Germany and other ex-communist metro areas), first nature (physical and political advantages/disadvantages) represents the mayor influence on agglomeration. This is principally true in the Eastern periphery and very North regions where natural conditions are more abrupt. On its side, GDP density disparities in the peripheral East Periphery and the whole South Periphery regions are mainly due to differences existent in man-made agglomeration economies, which represent about 6 times first nature influence. In effect, Southern latitudes are climatologically more even or -at least- not so unfavorable for life conditions.

In spite of having controlled for spatial autocorrelation, this effect is still present in the South Periphery regressions, though as spatial error autocorrelation. As shown in Anselin (1988), it could produce inefficient -though unbiased- estimators. This situation also takes place in other regressions due to the presence of heteroskedasticity in the error terms. Both heteroskedasticity and spatial error autocorrelation, though much more controlled than in the global regression (Table 5) could somewhat affect the R^2 measure. In our opinion, this is one of the weaknesses of Roos' methodology. Nevertheless, in general terms the situation exposed in Fig. 7 is very reliable and it constitutes a first attempt of measuring the real impact of geography on economic agglomeration in Europe.

Fig. 7 ANOVA results for log relative GDP per Area in the 8 EU regional clusters



SP_0: South Periphery, periphery, SP_1: South Periphery, core, NP_0: North Periphery, periphery, NP_1: North Periphery, core, MC_0: Main Core, periphery, MC_1: Main Core, core Source: Self elaboration

The ANOVA allows disentangling -in each spatial cluster- the contribution of each net component of geography with the aim of isolating the importance of first nature alone. First nature is particularly relevant in the Center, North and East Periphery core. That is to say, existing agglomerations in these areas are more stable, since they are tied to specific places. At the same time, attempts to create new agglomerations at places without geographic advantages, might possibly fail. On the contrary, in the South Periphery and peripheral regions of Eastern Europe, first nature matters much less; i.e. regional policies could have more success in creating agglomerations anywhere.

6. CONCLUSIONS

In this paper, we examine the influence of geographic features on the location of production in the EU. In other words, we quantify how much of the geographic pattern of GDP can be attributed to only exogenous first nature elements (physical and political geography) and how much can be derived from endogenous second nature factors (man-made agglomeration economies), in which first nature also operates as a mixed effect.

For this purpose, we follow a methodological approach based on Roos (2005) for the German planning regions. He proposes to employ an analysis of variance (ANOVA) to infer the unobservable importance of first nature indirectly in a stepwise procedure. We demonstrate that results could be biased if some potential econometric questions are not properly taken into account; e.g. multicollinearity, relevant missing variables, endogeneity, spatial autocorrelation and spatial heterogeneity.

The main outcome of our study reveals that production is not randomly distributed across the European regions. In an exploratory spatial data analysis we find that GDP density is bifurcated on one core (Central Europe plus Paris and London) and three (North, East, South) peripheries. Therefore, we have estimated our models testing for and considering explicitly these spatial regimes.

Considering the EU territory as a whole we find that only 53% of GDP's spatial variation can be explained by direct and indirect effects of geography. After controlling for agglomeration economies and the interaction effect of first-second nature, the net influence of natural geography rises to 20%. We find that almost a 50% of GDP's variance is reserved to unexplained factors, what we attribute to the high degree of unattended spatial effects present in the error terms.

However, the influence of geography varies significantly from one spatial regime to the other. For example, in the traditional Western Europe plus the East Periphery-core (mainly Austria, Slovenia, Eastern Germany and other ex-communist metro areas), first nature (physical and political advantages/disadvantages) represents the mayor influence on agglomeration. This is principally true in the Eastern periphery and very North regions where natural conditions are more abrupt. That is to say, existing agglomerations in all these areas are more stable, since they are tied to specific places. At the same time, attempts to create new agglomerations at places without geographic advantages, might possibly fail. On its side, GDP density disparities in the peripheral East Periphery and the whole South Periphery regions are mainly due to differences existent in man-made agglomeration economies, which represent about 6 times first nature influence. In these last regions, first nature matters much less; i.e. regional policies could have more success in creating agglomerations anywhere.

Even so, this methodology has some important limitations. On the one hand, the definition of the endogenous variable as aggregated 'GDP per Area' implies more a concept of agglomeration of value added rather than the spatial concentration of workers/citizens. On the other hand, the cited potential econometric problems are not always easy to solve.

7. REFERENCES

Ades AK, Glaeser, EL (1995) Trade and circuses: explaining urban giants. Quarterly Journal of Economics 110: 195–227

Alonso-Villar O, Chamorro-Rivas JM, González-Cerdeira X (2004) Agglomeration economies in manufacturing industries: the case of Spain, Applied Economics 36: 2103–2116

Anselin L (1988) Spatial Econometrics: Methods and Models. Kluwer Academic Publishers. Dordrecht, Germany and Boston, MA, USA

Anselin L (1990) Spatial dependence and spatial structural instability in applied regression analysis. Journal of Regional Science 30: 185–207

Anselin L (1995a) Space Stat Version 1.80: Users' Guide. Regional Research Institute, West Virginia University, Morgantown, WV, USA

Anselin, L. (1995b), 'Local indicators of spatial association-LISA', *Geographical Analysis* 27 (2), 93–115

Anselin L (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer M, Scholten H, Unwin D (eds) Spatial Analytical Perspectives on GIS. Taylor and Francis, London, UK, pp. 111-126

Anselin L (1999) Spatial Data Analysis with SpaceStat™ and ArcView®. Workbook, 3rd edn, Department of Agricultural and Consumer Economics. University of Illinois at Urbana-Champaign., Urbana, IL, USA

Arbia G (1989) *Spatial data configuration in statistical analysis of regional economics and related problems*. Kluwer Academic Publishers, The Netherlands

Ayuda MI, Collantes F, Pinilla V (2005) From locational fundamentals to increasing returns: The spatial concentration of population in Spain, 1787-2000. Documento de Trabajo 2005-05, Facultad de Ciencias Económicas y Empresariales, Universidad de Zaragoza (Spain)

Borén T, Gentile M (2007) Metropolitan processes in post-communist states: an introduction, *Geografiska Annaler Series B, Human Geography* 89-2, pp. 95-110

Brunet R (1989) *Les villes «européennes»*, Datar

Brunsdon C, Fotheringham AS, Charlton M (1999) Some notes on parametric significance tests for Geographically Weighted Regression. Journal of Regional Science 39: 497-524

Ciccone A, Hall RE (1996) Productivity and the density of economic activity. American Economic Review 86: 54–70

Cragg M, Kahn M (1997) New estimates of climate demand: evidence from location choice. Journal of Urban Economics 42: 261–284

Cheshire P, Hay DG (1989) *Urban Problems in Western Europe: an Economic Analysis*, Unwin Hyman, London

Davidson R, Mckinnon JG (1993) Estimation and Inference in Econometrics. Oxford University Press, New York, NY, USA

Delgado M, Sánchez J (1998) Las desigualdades territoriales en el Estado Español: 1955-1995. Revista de Estudios Regionales 51: 61-89

Disdier A-C, Mayer T (2004) How different is Eastern Europe? Structure and determinants of location choices by French firms in Eastern and Western Europe 32, pp. 280-296

Dobado R (2004) Un legado peculiar: la geografía. In: Llopis E (ed) El legado económico del Antiguo Régimen en España. Editorial Crítica. Barcelona (Spain)

Freeman DG (2001) Sources of fluctuations in regional growth. Annals of Regional Science 35: 249–266

Funck RH (1995) Competition among locations: objectives, instruments, strategies, perspectives. In: Giersch H (ed.) Urban agglomeration and economic growth. Springer, New York (USA) and Heidelberg (Germany), 227-255

Gallup JL, Sachs JD (1998) Geography and economic development. Consulting Assistance on Economic Reform II Discussion Paper 39, March. Available for download from http://www2.cid.harvard.edu/cidpapers/caer/paper39.pdf.

Gallup JL, Sachs JD, Mellinger AD (1999) Geography and economic development. International Regional Science Review 22: 179-232

Graves PE (1979) A life-cycle empirical analysis of migration and climate, by race. Journal of Urban Economics 6: 135–147

Hall P (1974) The containment of urban England, *The Geographical Journal* 140-3, pp. 386-408

Henderson JV (1988) Urban Development – Theory; Fact, and Illusion. Oxford University Press, Oxford, UK

Kanbur R, Venables AJ (2007) Spatial disparities and economic development. In Held D, Kaya A (eds) Global Inequality. Polity Press, London (UK): 204-215

Klein Tank, AMG and 35 coauthors (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. International Journal of Climatology 22: 1441-1453. Data and metadata available at http://eca.knmi.nl

Knapp T A, White N E, Clark DE (2001) A nested logit approach to household mobility. Journal of Regional Science 41: 1–22

Krugman P (1993) First nature, second nature, and metropolitan location. Journal of Regional Science 33: 129-144

Krugman P (1999) The role of geography in development. International Regional Science Review 22: 142–161

Krugman P, Livas R (1996) Trade policy and third world metropolis. Journal of Development Economics 49: 137-150

Lang RE, Simmons, PA (2003), *Edge counties: Metropolitan growth engines*, Fannie Mae Foundations Census Note 11, June. Available from http://www.fanniemaefoundation.org/programs/pdf/census/notes_11.pdf

Limão N, Venables AJ (2001) Infrastructure, geographical disadvantage, transport costs, and trade, The World Bank Economic Review, 15(3): 451-479

Mathias K (1980) *Wirtschaftsgeographie des Saarlandes*. Buchverlag Saarbrücker Zeitung

McCallum J (1995) National borders matter: Canada–US regional trade patterns. American Economic Review 85: 615–623

Peeters L, Chasco C (2006) Ecological inference and spatial heterogeneity: an entropy-based distributionally weighted regression approach, *Papers in Regional Science* 85, pp. 257-276

Rappaport, J (2000) Is the speed of convergence constant? Federal Reserve Bank of Kansas City Working Paper 99-13 (August)

Rappaport J, Sachs J (2003) The United States as a coastal nation. Journal of Economic Growth 8: 5-46

Rey S, Montouri B (1999) US regional income convergence: a spatial econometric perspective. Regional Studies 33: 143–56

Roos MWM (2005) How important is geography for agglomeration? Journal of Economic Geography 5: 605-620

Sachs J (2000) Tropical underdevelopment. CID Working Paper 57

Sargan JD (1964) Wages and prices in the United Kingdom: A study of econometric methodology. In Hart PE, Mill G, Whitaker JK (eds.) Econometric analysis for national economic planning, Butterworths, London, UK

Sessa C (2006), Core-periphery issues, Discussion paper for Forum Meeting 3: New needs for Europe: the renewed Lisbon Strategy & transport policy. Available from http://www.transforum-eu.net/IMG/pdf/Core-periphery-2.pdf

Venables AJ (2003) Spatial disparities in developing countries: cities, regions and international trade". Available for download from:

http://www.econ.ox.ac.uk/members/tony.venables/unpub_papers.html#spatrktv