LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Growth,

Innovation & Spatial Dynamics

A spatial analysis of innovation in regional Europe

Kristofer Rolf Söderström

kristofer_rolf.soderstrom.064@student.lu.se

Abstract: The analysis of spatial dependence can provide a useful perspective in the complicated analysis of the determinants of innovation in the regional level. Perhaps it is beneficiary to have geographical proximity towards other innovative regions. This study attempts to analyse and describe this patterns for 238 regions in two time periods of analysis (2000-2003 & 2007-2010). Visual analysis and statistical testing for spatial dependence and spatial regression models are used as tools for the study. It is possible to find some evidence for such dependence, although there are some key issues that have to be addressed beforehand.

Key words: **Innovation, spatial dependence, spatial regression models, region.**

EKHM51

Master thesis, first year (15 credits ECTS)

June 2015

Supervisor: Josef Taalbi

Examiner: Michaela Trippl

Word Count: 11,150

Table of Contents

1. Introduction

Spatial analysis can provide useful information when using models to measure the relationship of dependent and independent variables. If there exists spatial dependence in the variables and is ignored, it can lead to model misspecification and estimation biases. This study focuses on the analysis of this spatial dimension of innovation in Europe, at the regional level, in the hopes of finding evidence of said spatial dependence and measuring its effects on patent applications, which is used as a proxy for innovation.

The analysis of spatial dimensions can provide a useful perspective in the complicated analysis of the determinants of innovation in the regional level. Perhaps it is beneficiary to have geographical proximity towards other innovative regions. This study attempts to analyse and describe the patterns of spatial dependence in innovation at the regional level in Europe. This is done along a sample of 238 regions and two time periods of analysis. The periods of interest are 2000-2003 & 2007-2010. Analysing the potential changes along the time periods will hopefully provide a welcome addition of change over time, which may address some interesting questions about the increase or decrease of spatial dependence over time. Of course, there are several limitations to the study which will be addressed accordingly.

2. Aim & research question

There exists an ongoing debate on the actual role of geographical proximity in innovative activities. Empirical research regarding this topic is necessary in order to provide more information on the actual relationships. Although a complicated subject with a high number of potential errors in the empirical side of the study, an attempt will be made to provide some further information on the importance, if any, of spatial dependence and its effects on innovation. This is done in the regional level (NUTS 2) in Europe. Thus, the main research question is the following:

What are the effects of spatial dependence in Europe, at the regional level?

However, in order to provide an answer for the research question, it is first necessary to provide some evidence for the actual existence of spatial dependence on the variables used for the analysis and in the model. It is possible to provide some initial information through visual analysis via the use of maps. However, statistical testing is also necessary in order to provide some certainty in the results. Only if the results are satisfactory and are to provide evidence for the existence of spatial dependence on the variables, then the actual research question will be explored. Measuring the differences in the results from the two time periods in the analysis could provide some information on how this dependency, if any, changes over time.

It should be noted that the current study has to make some assumptions into consideration when constructing the empirical study which could provide biases in the results. An attempt will be made in addressing them and if not possible, they will be mentioned in the results and will be taken into consideration for the interpretation of the results.

3. **Previous Research**

A great number of literature has been written about innovation and its determinants, all with different approaches, scopes, methodologies and results. As the focus of this study is on the regional level and deals with the effect of spatial dependence on innovation, relevant research within these topics has been selected for review.

In their study of the role of geographical proximity in innovation, Sonn & Storper (2003) attempt to measure the tendency of an inventor to cite patents in their same geographical area. Patent citations are used in order to investigate if this dependence exists and if there is a significant change over time. In their study, they suggest that a difference exists between information and economically-useful knowledge and attempt to capture the latter in their analysis for patent citations, arguing that the effective transmission of this type of knowledge is affected not only still affected by geographical proximity, but that this effect is increasing over time. The authors analyse the period 1975-1997 and find an increasing tendency of inventors citing local patents at three different geographical levels: national, state and metropolitan levels. They find that inventors tend to rely much more in local, rather than foreign knowledge, suggesting an apparent increase in the role of geographical proximity in the creation of economically-useful knowledge. (Sonn & Storper, 2003).

In their study, Asheim & Gertler (2009) argue that the geographical configuration of economic actors is fundamentally important in shaping the innovative capabilities of firms and industries. The authors introduce the concept of a regional innovation system in order to describe this effect on the regional level. They argue that the synthetic and analytical knowledge bases are influenced by spatial proximity, although the mechanism in which this influence works is different depending on the knowledge base. The synthetic knowledge base, in innovation related activities, tends to be oriented towards the modification of existing processes and products with most of the modifications take place in existing firms. The analytical knowledge base is related to activities where scientific knowledge, formal modelling and codified science is highly important. Basic and applied research, along with development of new products and processes are main activities in this type of knowledge.

One of the most important basis for innovation-based value creation seems to be constituted by tacit knowledge, which is present in both knowledge bases. (Asheim & Gertler, 2009). The authors argue that this type of knowledge is what provides a key determinant of the geography of innovative activity. Tacit knowledge, the authors argue, is difficult to exchange over long distances and is better suited for face-to-face interaction with partners of equal backgrounds and commonalities, such as language, conventions and norms. Another aspect for the importance of geography in this type of knowledge, as explained by the authors, is related to the changing nature of the innovation process itself.

In her exploratory study on the role of geographical proximity in innovation, Gust-Bardon (2012) downplays the importance of geography, claiming that geographical proximity alone is not a sufficient factor to encourage collaboration between agents and enhancing knowledge transfer. The author also argues that the role of ICT has been crucial in the facilitation of knowledge transfer across long distances and works as a means of replacing face-to-face relations, in which virtual proximity can serve as surrogate for physical proximity. In this line of thought, it seems that he author expects a lower dependency on geographical proximity over time. However, the author does mention that geographical proximity still plays an important role, especially in the early stages of innovation processes and in local companies, which may have easier access to close knowledge networks and institutional support.

Audretsch & Feldman (1996) perform a geographical analysis of innovation and production for the United States. The authors use a database of some 8,000 commercial innovations which were introduced to the United States in 1982. They are able to find a concentration in the number of innovations along the coasts of the country and a seemingly inexistence of innovative activity in some Midwestern states. In the formal testing, they attempt to measure GINI indexes for different types of industry, university research and industry R&D research through OLS and 3SLS methods of estimation. The main hypothesis of the study is that innovative activity will tend to cluster in industries where new economic knowledge plays an essentially important role. The authors find some evidence to support the hypothesis, although they mention that based on the results, it would appear that this clustering may be more attributable to knowledge spillovers, rather than pure geographic concentration of production.

4. Theoretical framework

The first law of geography, attributed to Waldo Tobler states that "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970 pp.236). This idea is useful to understand the concept of spatial autocorrelation or dependence, which attempts to measure the degree to which one object is similar to other nearby objects. A formal definition would be "the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface, introducing a deviation from the independent observations assumptions of classical statistics". (Griffith, 2009 pp. 1).

The concept of spatial dependence is an interesting one because it is expected that real-world phenomena is more likely to interact in an orderly manner, in terms of spatial distribution, rather than its alternative of random spatial distribution. There are a number of useful examples for this. Mineral deposits tend to cluster in few locations around the Earth, housing prices tend to be influenced by the prices of nearby housing areas, and disease tends to concentrate across space due to contagion capabilities, amongst others (Griffith, 2009). It is of interest then, to be able to provide some evidence for another example of spatial dependence; innovation, as measured by its proxy of patent applications.

Technology and knowledge flow across borders. Firms interact with foreign firms and universities. Markets are global and communication technologies have greatly enhanced opportunities for communication and business across countries (Frascati Manual, 2002). The possibility of interaction between both national and regional borders is not necessarily an indication that geography does not play a role in the innovation process. On the contrary, it could signify that the concentration transcends the national level and it is in the regional level in which this interactions are more easily detected. According to the Oslo Manual (2005), although much knowledge can be accessed without direct interaction with the source, it is more likely that the codified knowledge will serve as a type of barrier, resulting in the process of finding new information a very costly one. Perhaps then, it is possible to assume that this codification could be somewhat easier to understand if there is a close interaction to the source, having the possibility of readily accessing the source of information, due to spatial proximity, may serve as a catalyst for the decoding of new information.

The Frascati Manual (2002) states that access to knowledge and technology would most likely depend on the connections between firms and organizations, particularly for tacit knowledge. The argument here is that the innovation process has been increasingly dependent on interactions and knowledge flows between economic entities, research organizations and public agencies. In consequence, it could be reasonable to assume that spatial proximity could play a crucial role, but perhaps is not the only determinant, in the effective transmission and production of tacit knowledge. This reinforces the relevance of innovative clusters, districts and, the main focus of this thesis; regions.

In his assessment on the role of proximity in innovation, Boschma (2005) argues that geographical proximity is not the only type of proximity, but only one of the different mechanisms in which closeness can play a role in innovative activity. The author mentions the different types of proximity as cognitive, organizational, social, institutional and finally, geographical. A short definition will be given for each one. Cognitive proximity is related to the firm's knowledge base, capabilities and skills, amongst others and how differences between actors affect their interaction. It is more likely that two firms with similar cognitive structures will interact with each other. Organizational proximity is defined as the extent of relationships in an organizational arrangement, closer relationships offer better and more efficient solutions to problems. Social proximity is defined as the degree of embedded relationships and ties between agents at the micro-level, the better the social structure within a firm, the higher the capacity of interactive learning and innovative performance. Institutional proximity is defined as the degree of shared formal and informal institutions. Formal institutions can be defined as the ones influenced by rule of law while informal institutions can be related to norms, culture and habits within society. Geographical proximity is defined as the spatial or physical distance between actors, both in relative and absolute meanings.

A large number of literature claims that agents that suffer spatial concentration usually benefit from knowledge externalities. In theory, Boschma (2005) claims that geographical proximity combined with some level of cognitive proximity is sufficient for interactive learning to take place. However, the author claims that geographical proximity can be substituted by any of the previously mentioned types of proximity, particularly because of the advancement of information and communication technologies. Boschma (2005) further states that while geographical proximity may facilitate learning and subsequently innovative performance, it is not a necessary nor sufficient condition. The author concludes that some level of proximity is required to benefit learning and innovation, but too much or too little can be detrimental.

On the challenges of empirical work in this topic, Boschma (2005) mentions that it is quite challenging to define the concepts of proximity such that there exist no overlap between them. This could present a problem for the actual study. The method for this study measures the relationship between patenting applications between a region and its neighbours through the use of a distance matrix. Although geographical proximity is the desired measurement, it is not possible to completely eliminate the aforementioned effect in the study. As Boschma (2005) states, the impact of geographical proximity can only be assessed in empirical studies when controlling for the other dimensions of proximity, as they may act as a powerful substitute. This is an important limitation in the empirical section of the study that needs to be understood.

The role of geographical proximity is also studied by Malmberg and Maskell (2006), in contrast with Boschma (2005), the authors suggest that geographical proximity is not a substitute, but a complement to the innovative process. In their study, the authors attempt to address some of the critiques on components behind the role of localized learning, as well as sorting out some misunderstandings to its use. They attempt to disentangle two elements of this concept: how localized capabilities enhance learning and the possible benefits that firms with similar activities may accumulate by locating in spatial proximity to one another. In respect to the role of spatial proximity, the authors conclude that this concept may be analysed along three dimensions of a local economic setting: the vertical, the horizontal and the social dimensions. The vertical dimension can be explained as the relatedness of firms in input-output relations. This dimension needs interaction between firms in order to develop. The horizontal dimension however, does not need interaction and is related to the competition between firms in which observation and comparison is useful in order to provide superior solutions than competitors. The social dimension refers to neighbourhood effects and how interaction in everyday life is responsible to the learning process as an unintended side effect of spatial proximity. The authors conclude that geographical proximity serves as a complementary effect in the process as it strengthens common settings between agents.

In regards to the mentioned theory, the following hypotheses are constructed with the same rejection structure. They will be addressed in the empirical analysis.

Hypothesis 1: Evidence of spatial dependence in the model for innovation.

- H0: No evidence for spatial dependence in the model.
- H1: Evidence for spatial dependence in the model.

Hypothesis 2: Evidence of a positive effect of spatial dependence in the model for innovation.

- H0: No evidence of a positive effect of spatial dependence in the model.
- H1: Evidence of a positive effect of spatial dependence in the model.

Hypothesis 3: Evidence for an increase of the positive effect of spatial dependence in the model for innovation over time.

- H0: No evidence for increasing effect over time.
- H1: Evidence for increasing effect over time.

In order to formally test the hypotheses, it will be necessary to construct a simple OLS model which attempts to measure some (but not all) determinants of innovation. This initial model will then be tested for spatial dependence and thus, the results will not attempt to explain what determines innovation. It rather attempts to measure the potential differences in the model when the spatial dimension is included in the model, while the rest is left constant. This will hopefully relax the necessary heavy assumptions behind the original model. As with most research, this study was subject to data availability, which is highly irregular at the regional level, particularly if one attempts to measure the relationships of large samples over a significant time period.

5. Estimation method

5.1. Spatial models and dependence

The use of cross-sectional tools is extremely useful when attempting to measure the relationship between variables. However, it is not sufficient when there is a possibility of spatial dependence and there may be a need to account for it in the models. Spatial regression models are useful when spatial dependency might be expected in the analysis.

In its simplest form, it is possible to see that a spatial regression model is an extension of a cross-sectional model which includes one or more spatially lagged terms on the right hand side of the equation. This spatial lag is achieved by associating the variable results with a spatial weights matrix *(W),* which will be addressed accordingly. There are two basic types of models with different assumptions in the structure of the spatial dependence: the spatial lag model and the spatial error model. The following presents the models in matrix notation:

- OLS model, simple form:

$$Y = X\beta + \epsilon \quad (1)$$

Where $\boldsymbol{Y}$ is the $N \times 1$ vector of the dependent variable is, $\boldsymbol{X}$ is the $N \times k$ matrix of the independent variable and $\boldsymbol{\epsilon}$ represent the error terms.

- Spatial lag model, with spatially lagged dependent variable:

$$Y = X\beta + \rho W y + \epsilon \quad (2)$$

Where $\boldsymbol{Y}$ is the $N \times 1$ vector of the dependent variable is, $\boldsymbol{X}$ is the $N \times k$ matrix of the independent variable, $\boldsymbol{W}$ is the $N \times N$ spatial-weights matrices that account for the geographical proximity component for the regions, $\boldsymbol{\epsilon}$ represent the error terms and $\boldsymbol{\rho}$ represents the spatial lag value.

- Spatial error model, with spatially correlated errors:

$$y = X\beta + \lambda W\xi + \epsilon \qquad (3)$$

Where most of the model is defined similarly to the spatial lag model. However, spatial dependence is seen primarily as a nuisance, much like statistical approaches often treat temporal serial correlation as an estimation problem. The error is then decomposed in two components, $\boldsymbol{\epsilon}$ a spatially uncorrelated error term and $\boldsymbol{\xi}$ which indicates the spatial component of the error term. The parameter $\boldsymbol{\lambda}$ indicates the extent to which the spatial component of the errors $\boldsymbol{\xi}$ are correlated with one another for nearby observations (Ward & Gleditsch, 2007).

The spatial lag model is useful when focusing on the spatial interactions of the dependent variable, when the structure of the spatial relationship is known. A useful example is how the price of a house will be affected by the price of neighbouring houses. The spatial lag model is a spatial autoregressive model that includes a spatially lagged dependent variable. The spatial lag is a representation of the weighted average of its neighbours. A spatial error model, on the other hand, is useful when the focus is correcting for spatial autocorrelation in the model and the structure of the spatial relationship is not well known. The spatially correlated errors are included due to unobservable features or omitted variables associated with allocation. An example for this type of model is when the technology adoption of a farmer may be influence by their neighbours. In this model, the error terms have the spatial structure. The multipliers in the dependent and independent variables represent the variation that cannot be explained by the neighbours' values. (Katchova, 2013)

It is difficult to differentiate the models in purely statistical grounds. Although there are some formal tests for comparing the models, the results are often inconclusive and will unlikely be able to provide strong support for either model (Ward & Gleditsch, 2007). According to the authors, the spatial error model is likely to be less interesting for the social sciences as this model is appropriate when the researchers are unwilling or unable to make assumptions about the origin of the spatial pattern, but otherwise suspect to find it in the error terms. Because of this difficulties in discriminating between the models, both results will be presented and compared. However, based on the descriptions of the models and its assumptions, one can be tempted to be inclined towards the spatial lag model.

The extension of Moran's I statistic to the regression context is one of the methods to measure spatial dependence in the variables, following the notation from Varga (1988):

$$I = \frac{e\prime We}{e'e} \qquad (4)$$

Where e is an Nx1 vector of regression residuals from the OLS estimation on a sample with N observations. W is an NxN spatial weights matrix. The result from Moran's I is global. This means that it is only able to provide evidence for spatial dependence for the whole sample. The null hypothesis of the test corresponds to spatial dependence, the alternative hypothesis states that the results are randomly distributed across the weights matrix.

The estimations of the spatial dependence test and spatial models are done with tools provided by the user commands for spatial data analysis (*spatwmat, spatdiag* and *spatreg*) written by Pisati (2001) for use in STATA. The models may be subject to estimation issues and it may be preferable to estimate them using the maximum likelihood estimator, this is automatically done in the STATA module. A good analysis of the issues is provided by Ward & Gleditsch (2007) and Varga (1988).

5.2. Spatial weights matrix

A spatial weight matrix provides the structure of the spatial relationship across the observations, providing the necessary information on neighbouring regions and serves as the spatial lag for the variables. The spatial weight matrix is usually defined as *W* and contains elements which indicate the spatial proximity of observations *i* and *j*. Two types of matrices exist, contiguity and distance based matrices. Contiguity matrices indicate if observations share a border or a vertex. Distance based matrices can be elaborated based in distance decay (1/d) or in distance bands. In spatial regression analysis, spatial weight matrices usually need to undergo some type of standardization, which depends on the actual method for analysis. In this case a distance based matrix with "row-standardization" is necessary. This implies that the sum of the weights is one in every row of the matrix, which will provide with the average value of the neighbouring regions. (Katchova, 2013)

There can be some degree of personal preference in the choice and development of spatial weight matrices. However, certain properties are necessary for spatial regression and spatial dependence analysis, both of which are used in the analysis

1. The diagonal elements of the spatial matrix are set equal to zero, so that results from own regions are not taken into consideration. The non-diagonal elements represent a binary value when being close to a region (1) or being far from a region (0), depending on the band specification.
2. When the distance band approaches zero, the spatial regression results approximates the results of Ordinary Least Squares.
3. Spatial weight matrices are row standardized. The sum of the row elements equals to 1. Each unit is a weighted average of the neighbours, which depends on the distance band.
4. The dimensions of the spatial matrix *NxN* have to be equal to the variable matrix Nx1.
    a. No missing data on the variables.

5. The distance band is selected by the researcher. Although it has to be high enough to avoid "islands" (regions without neighbours) as this prevents the calculation of results.

6. Data

6.1. Geographical data

As the interest of this study is to describe and measure the spatial dependence of innovation at the regional level, it is necessary to capture where the regions are located in geographical space. There are many different ways of capturing this characteristic in the regional level. However, only two types are necessary for this study: polygons and points in two dimensional space. In vector data models and maps, polygons are defined as a closed shape interconnected by a sequence of x and y coordinates, where the first and last pairs of coordinates are equal and every other pairs of coordinates are unique. Points on the other hand, are simply defined as a single pair of x and y coordinates (Huisman & de By, 2001).

The two types of geographical representation of space will be used in the different analyses, polygon data is used in the development of maps for the visual analysis and point data is used in the development of spatial weight matrices for use in the statistical analysis because of its simplicity and relative ease of development and use. Polygonal coordinates are available from Eurostat (2015b) and contain the x and y coordinates for all the European countries in different regional levels. Point data is publicly available at European Transport Policy Makers (2015) for all region levels and all European countries. Through some manipulation and merging of the original geographical databases, it was possible to construct a master database which included all of the necessary geographical data for the two types of analyses. This master database includes the polygon coordinates and regional centroid coordinates for 281 regions and 30 countries across Europe[1].

As mentioned earlier, geographical proximity can be measured by different methods. Both measurements in this study are similar in their assumptions and are subject to the same set of advantages and disadvantages. However, it is only necessary to address them for the point data, as it is the one that will be used in the statistical analysis and such limitations are not an issue for visual data, which uses polygon coordinates. Because of the nature of point data, which uses x and y coordinates of the regional centroids, geographical distance is measured as the straight line distance between two points in space.

One can think of some disadvantages and advantages to this type of geographical measure of distance. Several disadvantages are mentioned. Firstly, it is a measurement of two dimensional space, which by itself is a contrast to reality. Secondly, it ignores any boundary, physical or otherwise, between regions. One can think of mountains, hills, valleys, bodies of water, amongst others. Lastly, it measures distance in a straight line, which ignores roads, train tracks or other types of transportation methods. The advantages to using this type of geographical measure are mostly related to the relative ease of interpretation and use. Calculating the necessary spatial weights matrix for the statistical analysis is relatively easier to do when using this method, particularly for such a massive scale as the one in this study. When considering the size of the regions, one can assume that regional centroid points could be a relatively accurate representation of their location in two dimensional space and their distance with other regional centroids, with few exceptions that will be addressed accordingly. One could expect to have increasing biases in the accuracy of the data with bigger area sizes, very large region or countries for example. Overall, perhaps a straight line is not the most accurate measurement of geographical distance but its benefits, especially regarding the ease of use and interpretation when constructing a spatial weights matrix, greatly outweighs the disadvantages.

6.2. Variables

The variables for analysis are patent applications, education attainment at the tertiary level and GDP per capita, all of which will be explained in a moment. The data consists of cross-sectional data for 264 regions in 27 European countries[2]. As one of the objectives of the study is to analyse possible changes in results over time, two time periods are selected. The first time period contains the years 2000-2003 and the second period contains the years 2007-2010. The regions correspond

[1] See annex for the complete list of countries and regions.
[2] Data availability for Patent applications and GDP. Availability for Tertiary education is lower, see annex for details.

to the Nomenclature of Territorial Units for Statistics (NUTS) 2 level, which refers to medium sized regions (Eurostat, 2015b). As it is the interest of this study to measure spatial dependence in innovation, patent applications is selected as the dependent variable. The independent variables are GDP per Capita and Tertiary education attainment[3]. The patent data is obtainable as Patent Applications to the European patent office (EPO) by priority year by NUTS 2 region, the units are number of applications and its denominator is per million inhabitants. This indicator is used as a proxy for innovation. The data for education attainment is obtainable by as Population Aged 24-65 with Tertiary Education Attainment units are in percentage of population and are also in the NUTS 2 level. The data for GDP per capita is obtainable as Gross domestic product (GDP) at current market prices, units are in Purchasing Power Standard (PPS) per inhabitant and the data is on the NUTS 2 regional level. All data was obtained through Eurostat (2015c), database of Regional Statistics by NUTS Classification.

As patent application data is used as a proxy for innovation, it is important to recognize it potential benefits and drawbacks. Measuring innovation is not without its controversy. There is an ongoing debate about the effectiveness of different methods and proxies and although no definitive consensus has been achieved. According to Fagerberg & Mowery (2009) innovation is sometimes suggested as impossible to fully quantify and that although this is true for some aspects of innovation, the overall characteristics allow for the measurement of key dimensions of processes and outputs. Patent data is one indicator that is able to capture some of these dimensions, most likely related to output.

Fagerberg & Mowery (2009) provide some advantages of using patent data as an innovation indicator. Firstly, the patent system records important information about the inventions. Secondly, the patent system collates the technologies according to a detailed, slow to change, classification system which could result in measurement consistency, at least in the short term. Thirdly, the patent system relates the invention to relevant technologies and provides citations to relevant technical and scientific literature, which makes it easier to track knowledge flows. Fourthly, the patent system is an old institution, which allows to the indicator to extend back for longer time periods, relative to other indicators. Lastly, this type of data is usually free and easy to obtain.

[3] As mentioned earlier, the choice of independent variables was subject of availability. It would have been preferred to use R&D data, in place of GDP per capita. However, this was not possible due to heavily missing data. It is hoped that including this indicator will account for economical size of the regions, which may affect innovation output. However, it is completely possible that the relationship is in the reverse. This has to be addressed as a major limitation in the study.

Some of the disadvantages related to the use of patent applications as an innovation indicator are also provided by Fagerberg & Mowery (2009). Firstly, patent data is usually an indicator for invention, rather than innovation. Secondly, patents mark emergence of a new technical principle and not necessarily a commercial innovation. Thirdly, many patents are not of technological and economical significance and lastly, the patent system misses many non-patented inventions and innovations.

With all its pitfalls and measurement issues, patent data has proven very useful in the empirical analysis of innovation[4]. It should be noted that for the purposes of this study, the decisive factor for the selection of patent data as the innovation indicator was related to availability. Out of the other possible indicators to choose from, patent application data was the most complete for the selected time period and was the one with the most availability across the regional level.

7. Results

7.1. Visual analysis

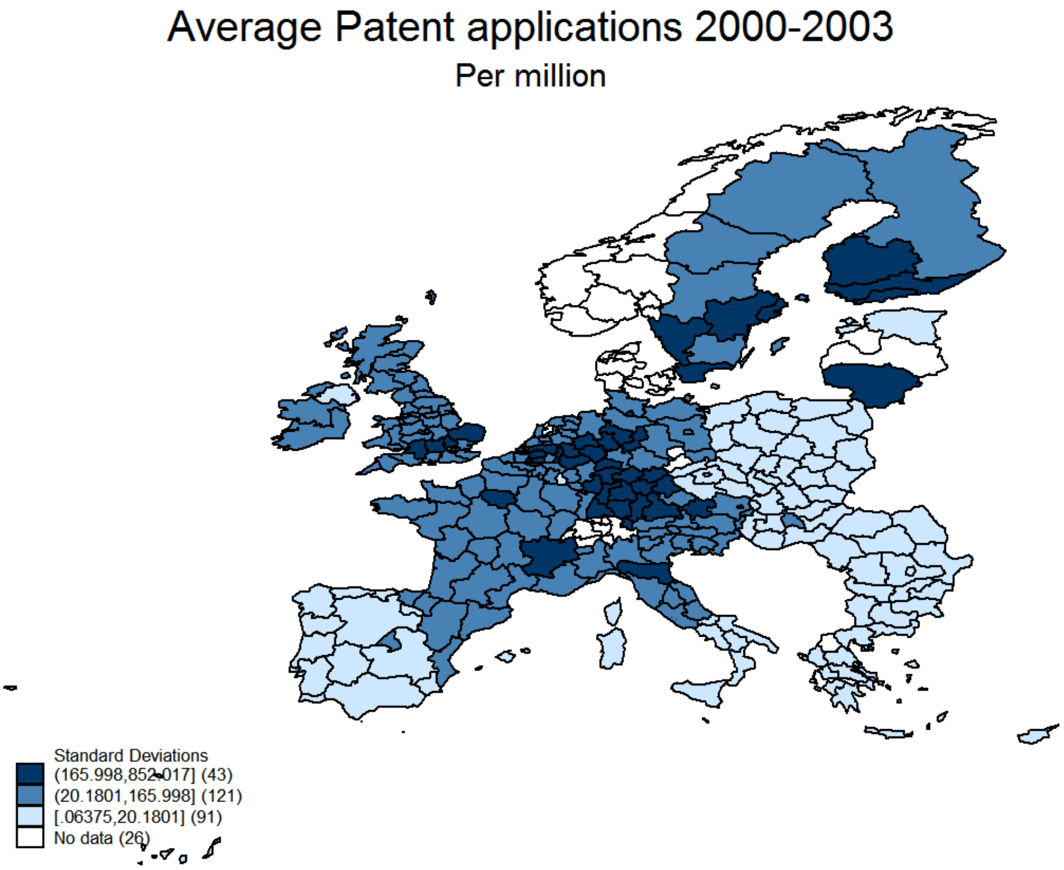
Visual analysis is an important aspect of any study. Perhaps it is even more important in a study that attempts to measure effects of geographical proximity as it would display more accurately the existence, if any, of geographical patterns in the data. For this reason, it was decided that presenting the data in map format is perhaps the most accurate method for a visual analysis, as it will provide with information relevant to the proposed hypotheses.

The map construction required the merging of the geographical database and the variables database. Because data is not available for all regions, some will appear as blank spaces in the map. In order to provide some kind of consistency within the measurements, standard deviations were selected as cut-offs for the scale in all the variables and no data transformations were made.

[4] Fagerberg & Mowery (2009) provide a useful summary of empirical works related to the analysis of innovation with patent data.

7.1.1. Patent applications to the EPO, per million

Figure 1: Patent applications, period 2000-2003



Note: Adapted with data from Eurostat © EuroGeographics for the administrative boundaries

Figure 1 shows average patent applications per million for the period 2000-2003 by region. From a visual aspect, it is possible to see an overall geographical concentration of high levels of average patent applications around centre-north and central Europe. It is also possible to see some concentration of high levels of patent applications in the southern regions of Sweden and Finland as well, even though they are some distance away from this central area of high patent applications. It would seem that going outwards of these apparent geographical concentrations tends to a decrease in regional patent applications. Perhaps this provides some evidence for geographical concentration.

Table 1: Countries with high patent regions, period 2000-2003

Country	Regions		
	High Patents	**Total**	**% high patents**
Lithuania	1	1	100.00%
Finland	3	5	60.00%

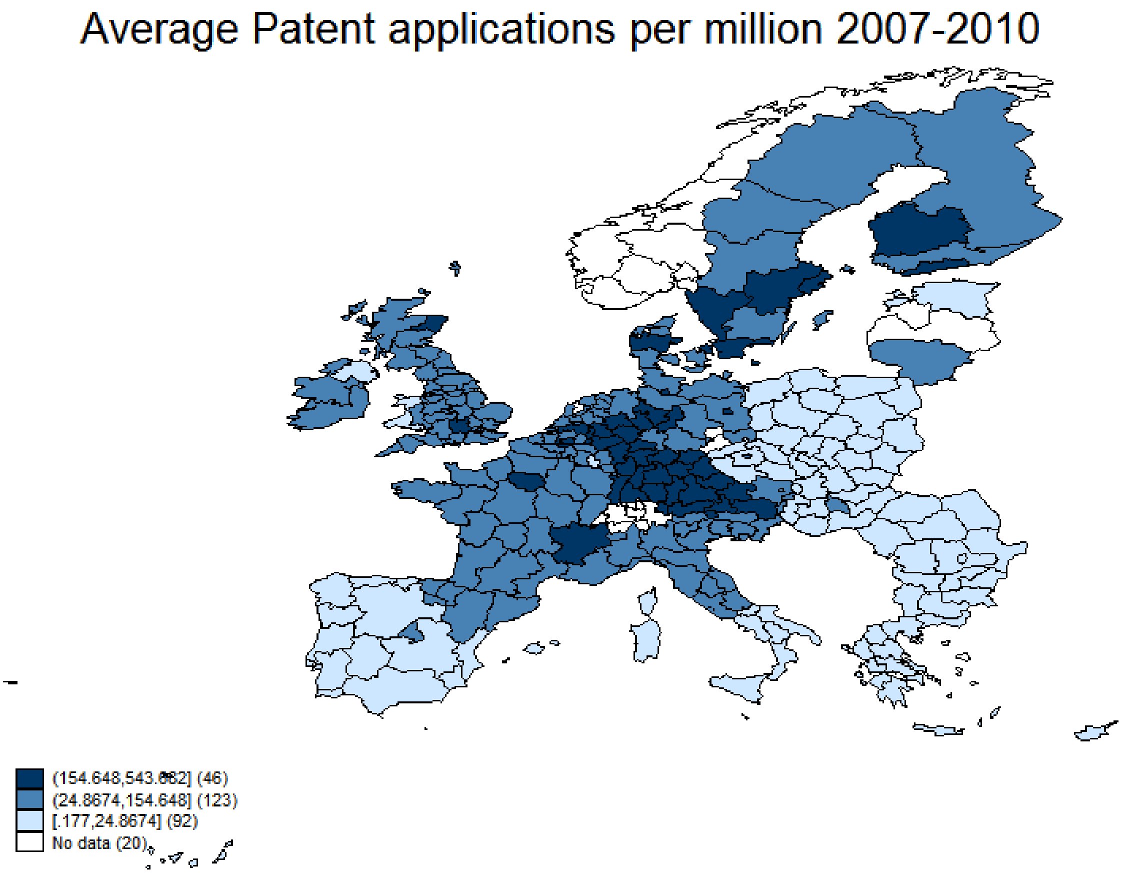
Germany	21	38	55.26%
Sweden	4	8	50.00%
Belgium	3	11	27.27%
Austria	2	9	22.22%
Netherlands	2	12	16.67%
United Kingdom	5	37	13.51%
France	2	22	9.09%
Italy	1	21	4.76%

Note: Adapted with data from Eurostat

Table 1 complements the information from the first map. It shows the number of regions within the highest scale of average patent applications by country, for the first period. Showing the information in this manner will hopefully make it easier to identify possible geographical patterns of high levels of patent applications. Making a comparison with the maps, it may be possible to identify even more characteristics of these patterns.

It is possible to see some kind of geographical agglomeration within the higher performers. Germany is the country with the most innovative regions in Europe and it is possible to see how some of its neighbouring countries also contain regions with a high average patent applications. The results of Sweden and Finland seem to be concentrated in the southern regions and interestingly enough, have some of the highest percentages of regions with high patent applications. Lithuania is an interesting result as well, it only contains one region in the NUTS 2 level but it still is one of the countries with the highest results in patent applications for the period. Although separated, it is interesting to note that out of the 37 regions in the United Kingdom, the five regions with high patent applications in the period seem to be closer to continental Europe, although it is possible that being close to the capital has an equal effect.

Figure 2: Patent applications, period 2007-2010



Note: Adapted with data from Eurostat © EuroGeographics for the administrative boundaries

Figure 2 shows average patent applications per million for the period 2007-2010 by region. The second period retains the previous apparent geographical concentration in centre-north and central Europe. However, the degree of the concentration appears to be higher in the second period. It is interesting to note that this concentration appears to be encompassing some of the regions towards east Austria. The high levels of patent applications per million in the southern regions of Sweden and Finland remain and with the now available data for Denmark, it is possible to see some apparent similarities in levels of patent applications with regions close to Sweden.

Table 2: Countries with high patent regions, period 2007-2010

Country	Regions		
	High Patents	**Total**	**% high patents**
Germany	24	38	63.16%
Austria	5	9	55.56%
Sweden	4	8	50.00%
Denmark	2	5	40.00%

Finland	2	5	40.00%
Belgium	2	11	18.18%
Netherlands	2	12	16.67%
France	3	22	13.64%
United Kingdom	2	37	5.41%

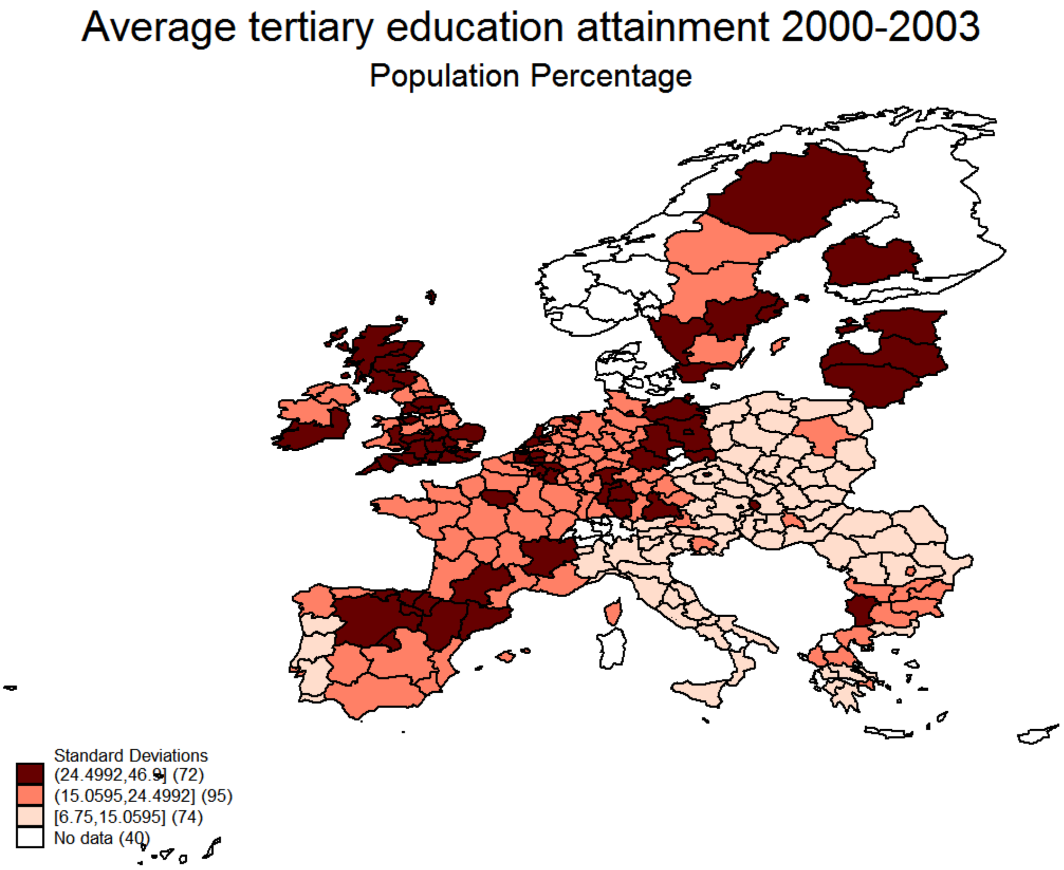
Note: Adapted with data from Eurostat

Table 2 complements the information from the map in figure 2. According to this figures, Germany is still the country with the most regions with a high average number of patent applications per million in Europe. It is also possible to see an increase in the number of regions relative to the previous period. Most of its neighbours retain similar levels of patent applications, although Belgium loses one region, relative to the previous period. It is interesting to see an increase in the number of regions with high patent applications in Austria, this reflects the eastward tendency shown in the map, relative to last period. With the now available data from Denmark, it is possible to identify an interesting apparent agglomeration of regions with high patent applications in the Nordic countries. These regions would appear to form some type of belt along the x axis of the map. Perhaps this says something about the relationship between these countries at the regional level. The United Kingdom has lost a number of regions, relative to last period, but it should be noted that the region of North Eastern Scotland is now in the top level of patent applications, which is relatively distant to any other high level regions. Both the northern region of Italy and Lithuania no longer hold a position in the top level of patent applications and there seems to be an overall tendency towards centralization.

According to the visual and descriptive information, it is possible to see some kind of geographical concentration in the data for patent applications per million. Overall, there seems to be a relatively high concentration around Germany and some kind of mutual influence between the Nordic countries. Regarding the changes over the time periods, it may be possible to see some changes in the geographical patterns. More regions in the central area of Europe have higher levels of patent applications, some isolated regions in the past period no longer take part in the highest performers and the availability of data for Denmark reinforces the idea of geographical concentration in the Nordic countries. However, this type of analysis is not enough to claim for sure the presence of either spatial dependence or its potential increase over time. Formal testing is a welcome addition for this, and will be performed accordingly.

7.1.2. Percentage of the population aged 25-64 with tertiary education attainment

Figure 3: Tertiary education, period 2000-2003



Note: Adapted with data from Eurostat © EuroGeographics for the administrative boundaries

Figure 3 shows the average tertiary education attainment, as percentage of the population, for the period 2000-2003 by region. From a visual aspect, it may be possible to see some small and somewhat scattered pockets of geographical concentration amongst some of the regions. However, it is not as evident as the data for patent applications. It is possible to see some geographical concentration amongst the northern regions of Spain and southern France. There are isolated pockets of high education levels in the Netherlands and Belgium, northern and southern United Kingdom, the regions in eastern Germany, Lithuania and Estonia, and a similar configuration of southern Sweden and Finland[5], with northern Sweden also having high levels of education.

[5] Finland has missing data on three out of the five regions for this period.

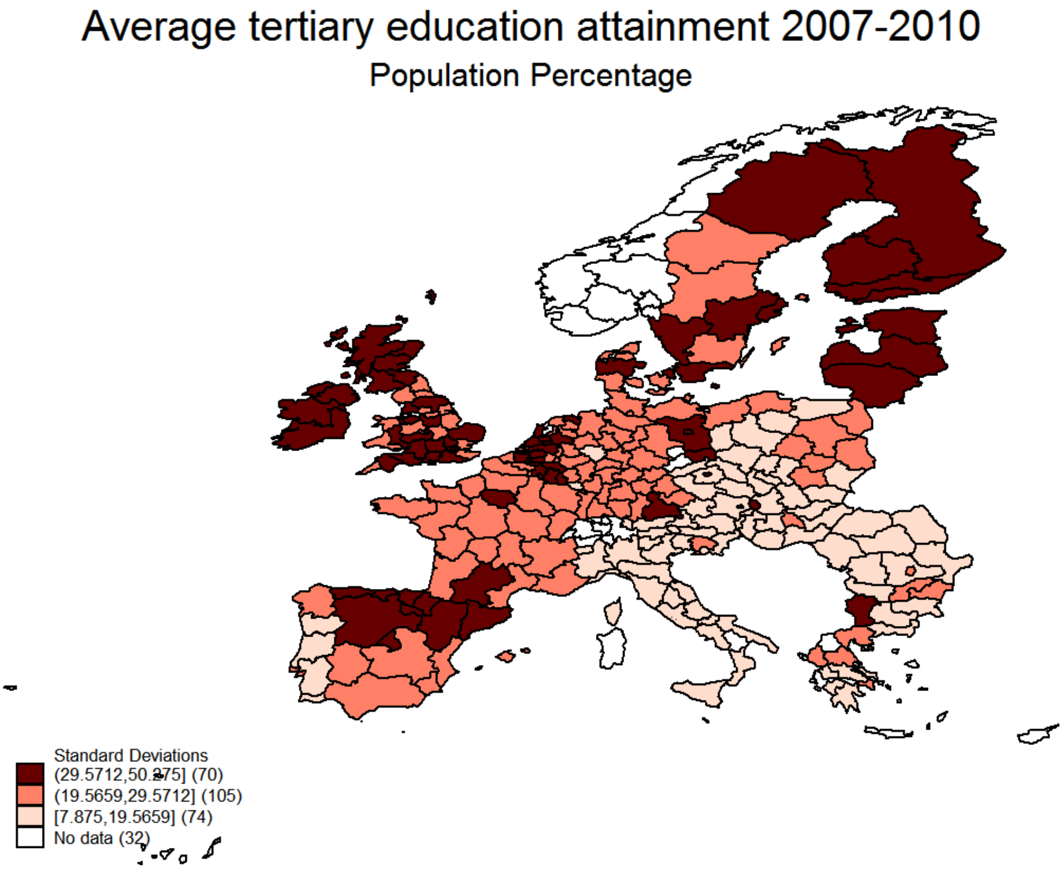
Table 3: Countries with high education regions, period 2000-2003

Country	**Regions**		
	High Education	**Total**	**% high education**
Cyprus	1	1	100.00%
Estonia	1	1	100.00%
Lithuania	1	1	100.00%
Belgium	8	11	72.73%
Sweden	5	8	62.50%
United Kingdom	23	37	62.16%
Ireland	1	2	50.00%
Spain	8	19	42.11%
Finland	2	5	40.00%
Netherlands	4	12	33.33%
Germany	12	38	31.58%
Slovakia	1	4	25.00%
Bulgaria	1	6	16.67%
France	3	22	13.64%
Czech Republic	1	8	12.50%

Note: Adapted with data from Eurostat

Table 3 complements the information from the map in figure 3. According to this figures, the United Kingdom is the country with the most number of regions with high tertiary education attainment. However, in terms of percentage it is surpassed by Belgium and Sweden (not taking into account one-to-one relationships between country-regions like Estonia and Lithuania). Reinforcing the information from the map, it appears that although there may be some kind of geographical agglomeration, the evidence is not as clear and it mostly resembles isolated pockets rather than a larger scale concentration.

Figure 4: Tertiary education, period 2007-2010



Note: Adapted with data from Eurostat © EuroGeographics for the administrative boundaries

Figure 4 shows the average tertiary education attainment, as percentage of the population, for the period 2007-2010 by region. There appears to be some kind of consistency amongst the high level regions, although with some apparent short range shifts. For one, the regions of northern Spain and southern France appear to not suffer much change, relative to the last period. The data for the United Kingdom would suggest the same consistency. Interestingly, there seems to be a high increase in the education levels in Ireland and the UK region of Northern Ireland, which share geographical boundaries. Similar pockets of high education levels can be seen in the Netherlands and Belgium, and in eastern Germany. However, both pockets seem more concentrated, relative to the last period. The Nordic countries also appear to retain some consistency in the data, although it should be noted that there was some missing data in the first period which would make this comparison no more than an assumption.

Table 4: Countries with high education regions, period 2007-2010

Country	Regions		
	High Education	Total	% high education
Cyprus	1	1	100.00%
Estonia	1	1	100.00%
Ireland	2	2	100.00%
Lithuania	1	1	100.00%
Luxembourg	1	1	100.00%
Finland	4	5	80.00%
Belgium	8	11	72.73%
Sweden	5	8	62.50%
United Kingdom	20	37	54.05%
Netherlands	6	12	50.00%
Spain	9	19	47.37%
Denmark	2	5	40.00%
Slovakia	1	4	25.00%
Bulgaria	1	6	16.67%
Germany	5	38	13.16%
Czech Republic	1	8	12.50%
France	2	22	9.09%

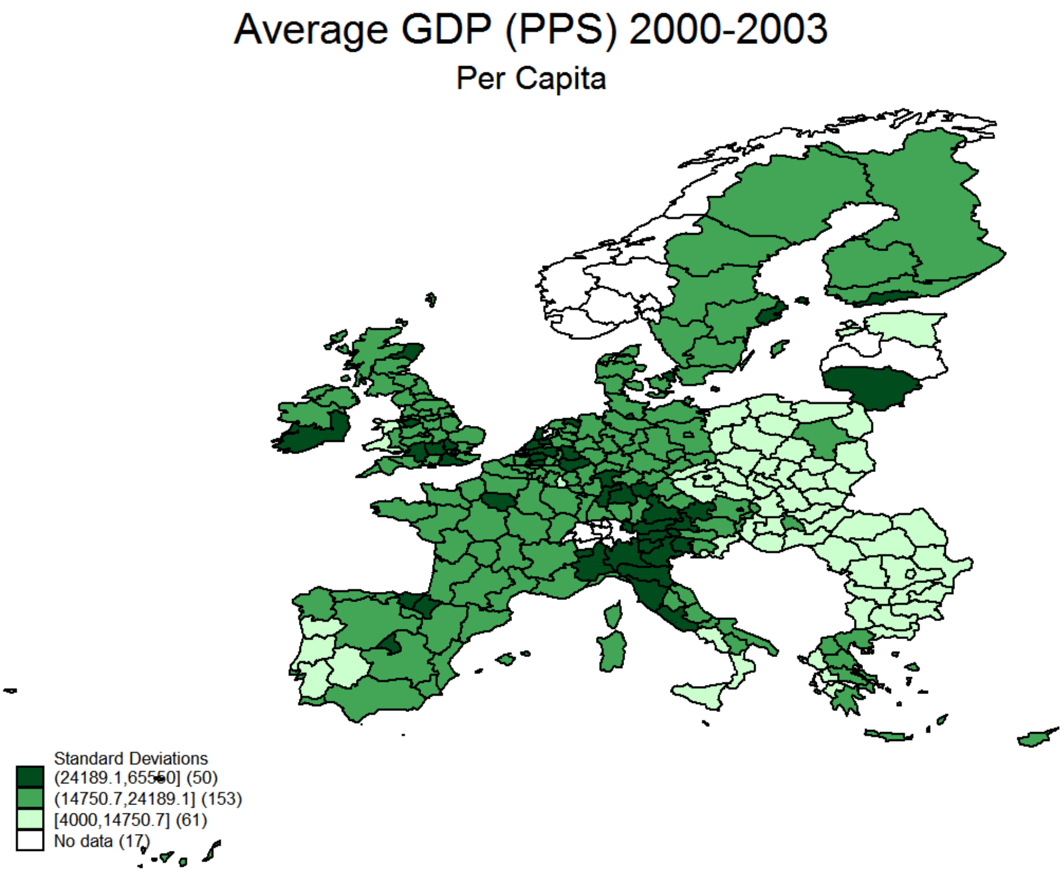
Table 4 complements the information from the map in figure 4. According to the figures, the United Kingdom still leads with the higher number of regions with high levels of education. It is surpassed in percentage levels by Ireland, Finland, Belgium and Sweden (again, without including one-to-one relationships between countries and regions). There is a large decrease in the number of regions with high levels of tertiary education in Germany, relative to the previous period. The United Kingdom also reduces its number of regions, although not in the same manner as Germany.

According to the visual and descriptive information, is possible to see some kind of geographical concentration in the data for tertiary education attainment. However, the data seems to behave more like scattered pockets of agglomeration, rather than the larger scale that was seen in patent application data. There seems to be some consistency over time, with what appears to be a higher concentration of said pockets, with some interesting decrease of regions with high levels of tertiary in unexpected countries, Germany being the most extreme case. The Nordic countries appear to retain similar characteristics over time, from what was possible to assume with the available data.

As with the data for patent applications, more testing is necessary before claiming any type of geographical concentration in the data.

7.1.3. Gross domestic product (PPS) per capita

Figure 5: GDP per capita, period 2000-2003



Note: Adapted with data from Eurostat © EuroGeographics for the administrative boundaries

Figure 5 shows the average levels of GDP (PPS) per capita for the period 2000-2003. It is possible to see some degree of geographical concentration of high levels of GDP per capita in in northern Italy and southern Austria, along a geographical "belt" that stretches from regions in western Austria, through Germany an up to the Netherlands. It is also possible to see some geographical concentration in southern United Kingdom, although it does not seem to be a high degree of concentration, relative to other examples in the visual analysis.

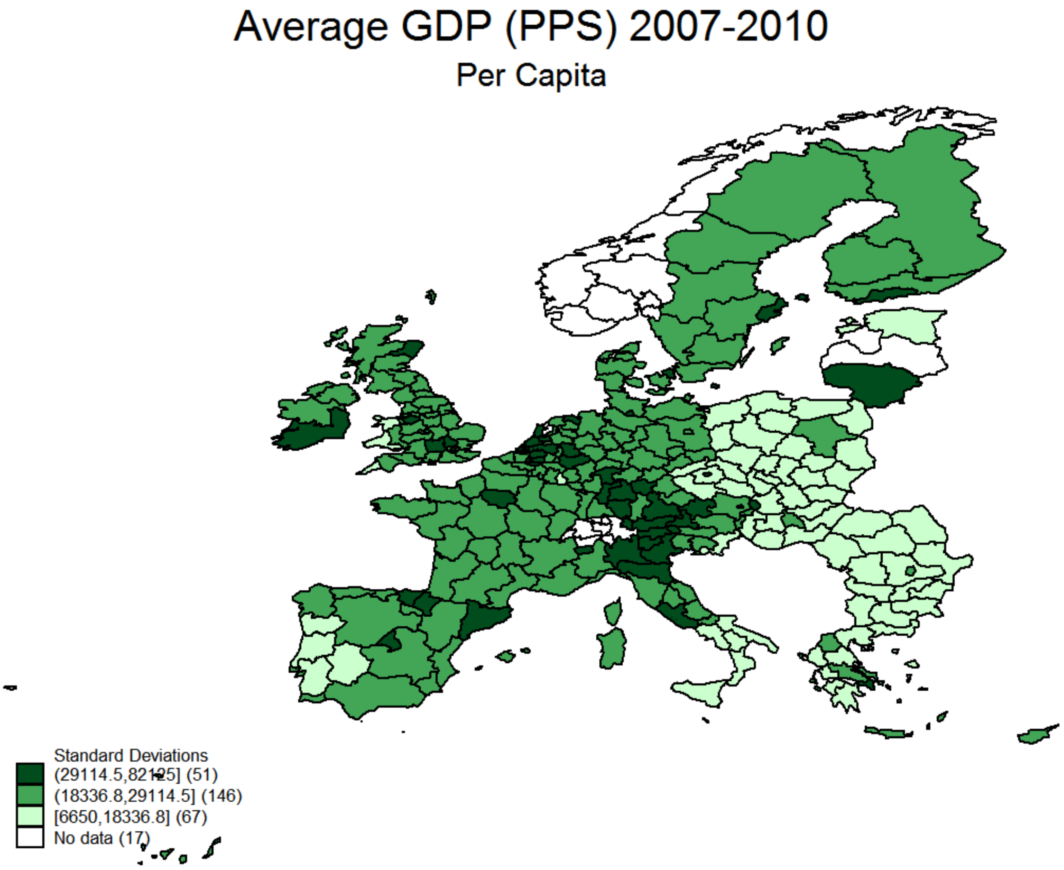
Table 5: Countries with high GDP per capita, period 2000-2003

Country	Regions		
	High GDP	**Total**	**% high GDP**
Lithuania	1	1	100.00%
Austria	5	9	55.56%
Ireland	1	2	50.00%
Italy	10	21	47.62%
Netherlands	5	12	41.67%
Finland	2	5	40.00%
Belgium	3	11	27.27%
Germany	9	38	23.68%
Denmark	1	5	20.00%
United Kingdom	7	37	18.92%
Spain	3	19	15.79%
Czech Republic	1	8	12.50%
Sweden	1	8	12.50%
France	1	22	4.55%

Note: Adapted with data from Eurostat

Table 5 complements the information from the map in figure 5. According to the figures, Italy is the country with the highest number of regions with high levels of GDP per capita, with Germany coming as a close second place, when accounting for purchasing power. It is surpassed in percentage levels by only Austria and Ireland. Visual evidence for this period is relatively weak when evaluating geographical concentration of the variable. The exception seems to be the aforementioned area around northern Italy and its connecting regions over the geographical "belt".

Figure 6: GDP per capita, period 2007-2010



Note: Adapted with data from Eurostat © EuroGeographics for the administrative boundaries

Figure 6 shows the average levels of GDP (PPS) per capita for the period 2007-2010. The second period does not seem to suffer much change, relative to the first period. The concentration of GDP per capita around the area of northern Italy appears to remain in the second period. Although, it should be noted that the concentration around the geographical "belt" of the previous period seems to be higher from a visual perspective, it seems to contain more regions around it.

Table 6: Countries with high GDP per capita regions, 2007-2010

Country	Regions		
	High GDP	Total	% high GDP
Lithuania	1	1	100.00%
Austria	5	9	55.56%
Ireland	1	2	50.00%
Netherlands	6	12	50.00%
Finland	2	5	40.00%
Belgium	4	11	36.36%

Italy	7	21	33.33%
Germany	10	38	26.32%
Slovakia	1	4	25.00%
Spain	4	19	21.05%
Denmark	1	5	20.00%
United Kingdom	5	37	13.51%
Czech Republic	1	8	12.50%
Sweden	1	8	12.50%
Greece	1	13	7.69%
France	1	22	4.55%

Note: Adapted with data from Eurostat

Table 6 complements the information from the map in figure 6. According to the figures, Germany is the country with the highest number of regions with high levels of GDP per capita, trading places with Italy relative to the last period. It is surpassed by Austria, Ireland, the Netherlands, Finland, Belgium and Italy in terms of percentage levels.

It seems, from the descriptive information on average GDP per capital levels appear to have some slight geographical concentration. This concentration appears to not only remain over the two time periods but seems to be of a somewhat higher concentration in the second period along the aforementioned geographical "belt". However, it is not possible as clear as in the visual representation of patent data and perhaps even less clear than education data. Nevertheless, geographical concentration is not possible to claim with absolute certainty from a visual analysis and further testing is required, as is the case for the previous variables.

The visual representation of the variables implies some interesting results towards the potential role of spatial dependence in the variables. Although with different behaviours across variables, there seems to be an overall geographical pattern regarding the data. It is difficult to conclude in a concise manner as there are different relationships across the variables, regions and time periods. In order to provide more information for the hypotheses, formal testing is required.

7.2. Spatial regression models

As mentioned earlier, spatial regression analysis can be performed in spatial lag models (equation 2) and spatial error models (equation 3). This section will provide the results from the estimations and their interpretation. In order to be able to perform this estimations, observations with missing data had to be deleted from the database. This was primarily because of missing data for the variable of Tertiary education attainment for the period 2000-2003.

7.2.1. Descriptive statistics

In order to better understand the selected variables, it is necessary to first present their descriptive statistics. This will provide additional information for data interpretation and will allow for a more proper evaluation in the model analysis. The variables in the descriptive analysis are in levels, without any kind of transformation.

Table 7: Descriptive statistics, period 2000-2003

Variable	Obs.	Mean	Std. Dev.	Min	Max
Patent Applications	238	94.26	121.65	0.06	852.02
Tertiary Education	238	19.62	7.84	6.75	46.90
GDP per Capita	238	19,528.05	8,120.42	4,000.00	65,550.00

Note: Adapted with data from Eurostat

Table 7 shows the descriptive statistics for the period 2000-2003. At first glance, it is possible to see an unequal distribution of the data for Patent applications and GDP per Capita, as the mean seems to be on the lower side of the minimum and maximum range of the data. The distribution for tertiary education seems to follow a more normal path, relative to the other variables. It is possible to expect this behaviour in the data as the sample is quite large and contains regions with very different characteristics. Perhaps it is necessary to make note of the relatively large standard deviation in patent applications, which is higher than the mean. This implies and even larger variation across the regions in terms of patent applications, which is to be expected.

Table 8 Descriptive statistics, period 2007-2010

Variable	Obs.	Mean	Std. Dev.	Min	Max
Patent Applications	238	90.37	109.19	0.18	543.68
Tertiary Education	238	24.14	8.22	7.88	50.28
GDP per Capita	238	23,750.84	9,292.14	6,650.00	82,125.00

Note: Adapted with data from Eurostat

Table 8 shows the descriptive statistics for the period 2007-2010. Similar to the previous period, it is possible to see a somewhat related distribution amongst the variables with patent applications and GDP per capita being consistent in their apparent unequal distribution amongst regions, and the relative normality in the data for tertiary education attainment. A similarly high standard deviation is seen in patent applications, relative to the previous period, in which it is higher than the mean.

In terms of change over the time period, it is possible to see a substantial decrease in the descriptive statistics for patent applications. On the negative side, the maximum and mean levels of the data have decreased in the second period. On the positive side, the standard deviation has decreased as well, which perhaps shows some lesser degree of variation of the data across the regions. A slight increase in the minimum number of patent applications is also present, although it is minimal. In regards to the changes of tertiary education attainment, it is possible to see an increase in the descriptive statistics. On the positive side, the maximum and mean levels have increased in the second period. On the negative side, the standard deviation also increases, possibly signifying a higher degree of variation in the data. In terms of GDP per capita, it is possible to see an overall increase in the statistics. On the positive side, the mean and maximum have increased in the second period. On the negative side, the standard deviations have also increased, which could signify increasing variation in the data[6].

7.2.2. Constructing the spatial weights matrix

Based on the required properties and characteristics mentioned earlier in the study, a distance matrix based on the distance between x and y coordinates was elaborated. This required an additional revision of the sample data and ultimately resulted in the elimination of regions with the following characteristics:

1. Regions with missing data in any of the variables or the time periods.
2. Regions without neighbours.
    a. Geographical Islands (i.e. Madeira, Portugal; Canarias, Spain)
    b. Very large regions with relative isolation (i.e. Northern Sweden)[7].

After this initial selection, the distance was calculated from the x and y coordinates in a Cartesian plane, from the regional centroids of the sample. The distance was calculated with the Euclidean distance equation between two points in a Cartesian plane:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

This resulted in a 238x238 distance matrix between every region of the sample. The first property was implemented, in which diagonal elements are equal to zero. A significantly reduced example of the matrix is shown in table 9:

[6] Although not the main focus of the study, perhaps it is important to note the possible effects of the economic crisis of 2008/09 in the variables. The only important decrease in the data seems to be associated with patent applications. However, without a deeper analysis, it is not possible to assure this is the reason for the decrease.

[7] See annex for deleted regions.

Table 9: Distance matrix

	AT11	*AT12*	...	*UKN0*
AT11	0	1.106938	...	24.45635
AT12	1.106938	0	...	23.46073
⋮	⋮	⋮	⋱	⋮
UKN0	24.45635	23.46073	...	0

Note: Adapted Euclidean distance matrix with data from ETIS

After this initial calculation of distance, a distance band of three was implanted. This initial distance band was selected as it is the minimum distance between two regions in the study. The distance band is used to produce a binary matrix in which 1 represents that the regions are within the distance band and 0 represents that the regions are outside of the distance band.

The binary matrix is now standardized by row weight. In other words, the non-zero values are divided by the number of regions per row. This results in the row standardized matrix in table 10:

Table 10: Binary matrix, row standardized

	AT11	*AT12*	...	*UKN0*
AT11	0	0.09091	...	0
AT12	0.06667	0	...	0
⋮	⋮	⋮	⋱	⋮
UKN0	0	0	...	0

Note: Adapted distance band matrix with data from ETIS

In order to better understand the distance, the calculation of the Haversine Formula[8] was performed, after the fact, in order to avoid further modifications to the distance matrix. The initial distance band of three implies that regions are considered neighbours if they are within approximately 215 kilometres.

[8] The Harvesine Formula calculates the great circle distance between a set of longitude and latitude coordinates: $a = sin^2\left(\frac{\Delta lat}{2}\right) + \cos(lat1) \cdot \cos(lat2) \cdot sin^2\left(\frac{\Delta long}{2}\right)$, $c = 2 \cdot atan2(\sqrt{a}, \sqrt{1-a})$, $d = R \cdot c$, Where $R = 6{,}371km$, the radius of the Earth

7.2.3. OLS model specification

It is necessary to construct an initial model in which the existence of spatial dependence will be tested. The original model is a cross-section for each time period of analysis with the following specification.

$$\ln Y_i = \beta_0 + \ln X1_i\, \beta_1 + \ln X2_i\, \beta_2 + e_i$$

Where Y is Patent applications per million, $X1$ is GDP (PPS) per Capita and $X2$ is Tertiary education attainment[9], for every i region in the sample. One regression is done per time period.

7.2.4. Testing for spatial dependence

With the specification of the OLS model and the calculation of the spatial weights matrix, it is now possible to construct the spatial models. However, it is first necessary to perform spatial diagnostics in order to identify if it is necessary to include the spatial dimension into the model. As mentioned earlier, this is possible to test with the STATA command *spatdiag* (Pisati, 2001), which provides the results for the Moran's I test and the Lagrange Multiplier test associated with the spatial lag and spatial error models. The statistic for Moran's I test the null hypothesis for global spatial autocorrelation. The statistics test the null hypothesis of no spatial autocorrelation in the dependent variable (in the spatial lag model) and no spatial autocorrelation in the error terms (in the spatial error model). The spatial dependence test is done in a simple OLS regression, in which the spatial weights matrix is included in order to test for the existence of spatial dependence in the model.

Table 11: Spatial dependence tests, period 2000-2003

Test	Statistic	p-value
Spatial Error		
Moran's I	13.163	0.00
LM	152.93	0.00
Robust LM	58.73	0.00
Spatial Lag		
LM	127.65	0.00
Robust LM	33.45	0.00

[9] Discussion about model specification in conclusions.

Table 12: Spatial dependence tests, period 2007-2010

Test	Statistic	p-value
Spatial Error		
Moran's I	12.97	0.00
LM	149.42	0.00
Robust LM	35.02	0.00
Spatial Lag		
LM	167.768	0.00
Robust LM	53.366	0.00

Tables 11 and 13 show the results for the different tests for the first and second period, respectively. It is possible to see that the null hypothesis for no spatial autocorrelation is rejected for every test and for both time periods. In this case, ignoring the spatial dependence of the variables could lead to estimation errors, as described earlier.

7.2.5. Estimation of results

Table 14: Estimation results for Patent applications, period 2000-2003

Variable	OLS	Spatial Lag	Spatial Error
GDP per Capita (log)	3.0489***	1.8426*9***	1.9625***
	(0.162813)	(0.1605)	(0.1909)
Tertiary education attainment (log)	0.9085***	0.7086***	1.0700***
	(0.1772)	(0.1383)	(0.1897)
Constant	-29.0385***	-18.3866***	-18.957***
	(1.3932)	(1.3895)	(1.6690)
Rho (ϱ)		0.5049***	
		(0.0414)	
Lambda (λ)			0.7955***
			(0.0452)
Observations	238	238	238
R-squared	0.7385	0.841	0.727

Standard errors in parenthesis. *, **, *** indicate significance at the 90%, 95% and 99% level, respectively.

Table 15: Estimation results for Patent applications, period 2007-2010

Variable	OLS	Spatial Lag	Spatial Error
GDP per Capita (log)	3.08353***	1.8486***	1.7694***
	(0.1929)	(0.1672)	(0.1888)
Tertiary education attainment (log)	0.4918**	0.3902**	0.7462***
	(0.2120)	(0.3902)	(0.2198)
Constant	-28.7888	-18.1543***	-16.5025***
	(1.6718)	(0.1558)	(1.6318)
Rho (ρ)		0.5641***	
		(0.0406)	
Lambda (λ)			0.7940***
			(0.0430)
Observations	238	238	238
R-squared	0.6422	0.805	0.628

Standard errors in parenthesis. *, **, *** indicate significance at the 90%, 95% and 99% level, respectively.

As with any type of modelling, a certain degree of caution is suggested when interpretation of results is made. For this reason, the interpretation will only address the hypotheses of this thesis. Certain other analyses can certainly be made with this type of modelling. However, they are beyond the scope of this study. Firstly, it is possible to see that all the coefficients show the expected signs in all the model types and most are significant to the 1% level (with tertiary education being the only exception in the OLS and Spatial Lag model, in the second period). The standard errors seem to be generally lower for the spatial lag model. However, they all seem small, relative to the coefficient size. The results from the r-squared also seem to be relatively higher for the spatial lag model. If one would have to choose one model in terms of the test statistics, one would prefer the spatial lag model. However, as was pointed out earlier, it is not recommend to select the models based on statistical results. It is for this reason that interpretation is somewhat limited to the overall results and the finer details will have to remain undiscovered.

With all the limitations however, it is possible to describe some overall results regarding the statistics. In terms of GDP per capita, the results show that if the spatial dependence between the variables is ignored it can lead to an overestimation of the coefficient, as seen from the results from the OLS regression. The results for tertiary education attainment are not as consistent as the ones for GDP. It is possible to see an overestimation of the coefficient in the OLS regression, when compared with the spatial lag model. However, this is reversed when the results are compared to the spatial error model, in which the results in the OLS are underestimated. Although there is a difference in the behaviour between the models, there is at least some consistency between the time periods.

One interesting interpretation between the different coefficients sizes can be inferred from an example provided in Ward & Gleditsch (2007). As specified earlier, the spatial lag model assumes that the spatial dependence comes from the spatial structure of the dependent variable, which in this case is patent applications. The region is influenced by the neighbours, which in turn influences the neighbours in some kind of feedback effect. In contrast, the spatial error model corrects for the positive spatial correlation of the dependent and independent variables (including the variables of GDP and education into the spatial structure). This difference in the mechanisms of the spatial structure is responsible for generally smaller coefficients of the spatial lag models, relative to the error models. As the spatial lag is more likely measuring immediate effects.

In the spatial lag model, the Rho (ϱ) estimate indicates that there is a positive, statistically significant and somewhat large spatial dependence. This provides support that a region's level of patent applications co-varies with the level of patent applications of its geographical neighbours (within 215 km). The value measures the average influence on the observations by the neighbouring observations. It is possible to see a slight increase in the result in the first period from 0.5049 to the result of the second period of 0.5641. This may provide some evidence to an increasing influence in the dependent variable.

In the spatial error model, the Lambda (λ) estimate indicates that there is also a positive, statistically significant, although larger spatial dependence, relative to the spatial lag model. However, this measure of spatial dependence is related to the error terms and thus includes effects from both dependent and independent variables. This provides further support for spatial dependence amongst geographical neighbours, although with a somewhat different structure. It should be noted that unlike the results from Rho (ϱ) estimate, the results stay constant over time, even decreasing in a relatively small manner from 0.7955 in the first period, to 0.7940 in the second period. This does not provide evidence of increasing spatial dependence in the model, although it may be explained by the method in which the spatial structure is defined, or more specifically, is not.

8. Conclusions

The aim of the study was to analyse the geographical patterns of innovation through means of visual and statistical analysis. The following can be concluded for the analysis with regards to the proposed hypotheses in the theoretical framework. In general terms, it has been possible to provide information for the rejection of the null hypotheses. A more detailed description follows.

It is possible to see some visual evidence of spatial dependence of the variables in the periods 2000-2003 & 2007-2010. Although with some different characteristics amongst themselves, the use of maps provides some compelling evidence for the geographical concentration for patent applications. The results for GDP per capita are less conclusive from a visual analysis, while the results for tertiary education appear to be even less favourable towards spatial dependence. However, with the use of several tests for spatial dependence in the OLS estimation provides sufficient information to safely assume that a spatial component, as defined by the spatial weights matrix, has to be taken into consideration in order to properly estimate the regression model. This appears to provide enough evidence to reject the first null hypothesis of the no spatial dependence in the regression model for innovation.

Comparing the results from the OLS model, the spatial lag model and the spatial error model, it is possible to assume that there indeed exists an effect in the model when including the spatial component. According to the theory of spatial regression analysis, the existence of a difference between the coefficients of the non-spatial and spatial models is an indication of an effect of spatial dependence. In other words, there seems to be enough evidence to suggest that spatial dependence affects the measurements of the original regression model. This appears to provide enough evidence to reject the second null hypothesis of no positive effect of spatial dependence in the regression model.

When doing the model comparisons between the time periods, the results differ depending on the spatial model. In the spatial lag model, there appears to be a slight increase in the positive effect of spatial dependence. However, this does not seem to be the case for the spatial error model. The difference may be stem from the methods in which the spatial structure is constructed, as stated before. From an a priori perspective, it would seem preferable to use the spatial lag model as it seems to more closely resemble the objective of the study. However, this type of spatial model makes strong assumptions about the spatial structure. As the original model already makes some strong assumptions about the use of independent variables it would not seem wise to make further, unnecessary assumptions. For this reason, the difference in the results would suggest that although there could be some increasing positive effect of spatial dependence in patent applications, as proxy for innovation, it is not possible to fully commit to this conclusion. Thus, the final null hypothesis of no increasing positive effect of spatial dependence cannot be rejected properly.

As Boschma (2005) suggested, perhaps it is not possible to fully measure the sole effect of geographical proximity as described in this study. It may be possible to accidentally include other types of proximity in the analysis and making the distinction would provide with different results. However, if one understands the limitations of the study, it still provides some interesting results with regards of the role of spatial dependence in innovation.

9. References

Asheim, B.T. & Gertler, M.S. (2009), The Geography of Innovation: Regional Innovation Systems, Chapter in the Oxford Handbook of Innovation (Online), edited by Fagerberg, J. & Mowery, D.C.

Audretsch, D.B. & Feldman, M.P. (1996), R&D Spillovers and the Geography of Innovation and Production, The American Economic Review, Vol. 86, No. 3, pp. 630-60.

Boschma, R. (2005), Proximity and Innovation: A critical Assesment, Regional Studies, Vol. 39, No. 1, pp. 61-74.

European Transportation Policy Makers (2015), Public downloads, available at: http://www.etisplus.eu/data/Public/Downloads.aspx

Eurostat (2015a) Administrative and Statistical Units © EuroGeographics for the administrative boundaries, Available at: http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units

Eurostat (2015b), NUTS Overview, Available at: http://ec.europa.eu/eurostat/web/nuts/overview

Eurostat (2015c), Regional statistics by NUTS classification, available at: http://ec.europa.eu/eurostat/web/regions/data/database

Griffith, D.A. (2009), Spatial Autocorrelation, Elsevier Inc, Available at: http://booksite.elsevier.com/brochures/hugy/SampleContent/Spatial-Autocorrelation.pdf

Gust-Bardon, N. (2012), The Role of Geographical Proximity in Innovation: Do Regional and Local Levels Really Matter?, Working Papers Firms and Regions, No.4

Huisman, O. & de By, R.A. (2001), Principles of Geographic Information Systems: An introductory textbook, The International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands

Katchova, A. (2013), The Econometrics Academy, Spatial Econometrics, Available at: https://sites.google.com/site/econometricsacademy/econometrics-models/spatial-econometrics

Malmberg, A. & Maskell, P. (2006), Localized Learning Revisited, Growth and Change, Vol. 37, No.1, pp. 1-18

Organization for Economic Cooperation and Development (OECD), Frascati Manual 2002: The measurement of scientific and technological activities - Proposed Standard Practice for Surveys on Research and Experimental Development", OECD, Paris, 2002, available at: http://ec.europa.eu/eurostat/ramon/statmanuals/files/Frascati_Manual_2002_EN.pdf

Oslo Manual (2005), Guidelines for Collecting and interpreting Innovation Data, 3rd edition, Available at: http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/oslo-manual_9789264013100-en

Pisati, M. 2001. sg162: Tools for spatial data analysis. Stata Technical Bulletin 60: 21–37. In Stata Technical Bulletin Reprints, vol. 10, 277–298. College Station, TX: Stata Press.

Sonn, J.W. & Storper, M. (2003), The Increasing Importance of Geographical Proximity in Technical Innovation: An Analysis of U.S. Patent Citations, 1975-1997, Paper for Conference: What Do we know about Innovation?, Sussex, 13-15 November 2003

Tobler, W.R. (1970), A Computer Movie Simulating Urban Growth in the Detroit Region, Economic Geography, Vol. 46, pp. 234-240

Ward, M.D. & Gleditsch, K.S. (2007), An Introduction to Spatial Regression Models in the Social Sciences

Varga, A. (1998), Spatial Analysis, chapter in University Research and Regional Innovation: A Spatial Econometric Analysis of Academic Technology Transfers, Kluwer Academic Publishers, Vienna

Vespargen, B. (2009), Innovation and Economic Growth, Chapter in The Oxford Handbook of Innovation (Online), edited by Fagerberg, J. & Mowery, D.C.

10. Annex

10.1. List of all regions and countries from the geographical database

Country	Region	NUTS ID
Austria	Vorarlberg	AT34
Austria	Tirol	AT33
Austria	Salzburg	AT32
Austria	Kärnten	AT21
Austria	Oberösterreich	AT31
Austria	Steiermark	AT22
Austria	Niederösterreich	AT12
Austria	Wien	AT13
Austria	Burgenland (A)	AT11
Belgium	Prov. West-Vlaanderen	BE25
Belgium	Prov. Oost-Vlaanderen	BE23
Belgium	Prov. Hainaut	BE32
Belgium	Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest	BE10
Belgium	Prov. Vlaams Brabant	BE24
Belgium	Prov. Brabant Wallon	BE31
Belgium	Prov. Antwerpen	BE21
Belgium	Prov. Namur	BE35
Belgium	Prov. Limburg (B)	BE22
Belgium	Prov. Luxembourg (B)	BE34
Belgium	Prov. Liège	BE33
Bulgaria	Yugozapaden	BG41
Bulgaria	Severozapaden	BG31
Bulgaria	Yuzhen tsentralen	BG42
Bulgaria	Yugoiztochen	BG34
Bulgaria	Severen tsentralen	BG32
Bulgaria	Severoiztochen	BG33
Cyprus	Cyprus	CY00
Czech Republic	Severozápad	CZ04
Czech Republic	Jihozápad	CZ03
Czech Republic	Praha	CZ01
Czech Republic	Strední Cechy	CZ02
Czech Republic	Severovýchod	CZ05
Czech Republic	Jihovýchod	CZ06
Czech Republic	Strední Morava	CZ07
Czech Republic	Moravskoslezsko	CZ08
Denmark	Syddanmark	DK03
Denmark	Midtjylland	DK04
Denmark	Nordjylland	DK05
Denmark	Sjælland	DK02
Denmark	Hovedstaden	DK01
Estonia	Estonia	EE00
Finland	Åland	FI20
Finland	Länsi-Suomi	FI19
Finland	Pohjois-Suomi	FI1D

Finland	Etelä-Suomi	FI1C
Finland	Itä-Suomi	FI1B
France	Bretagne	FR52
France	Pays de la Loire	FR51
France	Basse-Normandie	FR25
France	Aquitaine	FR61
France	Poitou-Charentes	FR53
France	Haute-Normandie	FR23
France	Midi-Pyrénées	FR62
France	Centre	FR24
France	Limousin	FR63
France	Île de France	FR10
France	Picardie	FR22
France	Nord - Pas-de-Calais	FR30
France	Languedoc-Roussillon	FR81
France	Auvergne	FR72
France	Bourgogne	FR26
France	Champagne-Ardenne	FR21
France	Rhône-Alpes	FR71
France	Provence-Alpes-Côte d'Azur	FR82
France	Franche-Comté	FR43
France	Lorraine	FR41
France	Alsace	FR42
France	Corse	FR83
Germany	Düsseldorf	DEA1
Germany	Trier	DEB2
Germany	Köln	DEA2
Germany	Saarland	DEC0
Germany	Münster	DEA3
Germany	Koblenz	DEB1
Germany	Weser-Ems	DE94
Germany	Rheinhessen-Pfalz	DEB3
Germany	Arnsberg	DEA5
Germany	Freiburg	DE13
Germany	Bremen	DE50
Germany	Darmstadt	DE71
Germany	Detmold	DEA4
Germany	Karlsruhe	DE12
Germany	Gießen	DE72
Germany	Hannover	DE92
Germany	Kassel	DE73
Germany	Tübingen	DE14
Germany	Stuttgart	DE11
Germany	Schleswig-Holstein	DEF0
Germany	Unterfranken	DE26
Germany	Hamburg	DE60
Germany	Lüneburg	DE93
Germany	Braunschweig	DE91

Germany	Schwaben	DE27
Germany	Mittelfranken	DE25
Germany	Thüringen	DEG0
Germany	Oberfranken	DE24
Germany	Sachsen-Anhalt	DEE0
Germany	Oberbayern	DE21
Germany	Oberpfalz	DE23
Germany	Mecklenburg-Vorpommern	DE80
Germany	Leipzig	DED5
Germany	Niederbayern	DE22
Germany	Chemnitz	DED4
Germany	Brandenburg - Südwest	DE40
Germany	Berlin	DE30
Germany	Dresden	DED2
Greece	Ionia Nisia	EL22
Greece	Ipeiros	EL21
Greece	Dytiki Makedonia	EL13
Greece	Dytiki Ellada	EL23
Greece	Thessalia	EL14
Greece	Peloponnisos	EL25
Greece	Sterea Ellada	EL24
Greece	Kentriki Makedonia	EL12
Greece	Attiki	EL30
Greece	Kriti	EL43
Greece	Anatoliki Makedonia, Thraki	EL11
Greece	Voreio Aigaio	EL41
Greece	Notio Aigaio	EL42
Hungary	Nyugat-Dunántúl	HU22
Hungary	Dél-Dunántúl	HU23
Hungary	Közép-Dunántúl	HU21
Hungary	Közép-Magyarország	HU10
Hungary	Dél-Alföld	HU33
Hungary	Észak-Magyarország	HU31
Hungary	Észak-Alföld	HU32
Iceland	Iceland	IS00
Ireland	Southern and Eastern	IE02
Ireland	Border, Midlands and Western	IE01
Italy	Valle d'Aosta/Vallée d'Aoste	ITC2
Italy	Piemonte	ITC1
Italy	Sardegna	ITI4
Italy	Liguria	ITC3
Italy	Lombardia	ITC4
Italy	Emilia-Romagna	ITF5
Italy	Toscana	ITF6
Italy	Provincia Autonoma Trento	ITF2
Italy	Provincia Autonoma Bolzano-Bozen	ITF1
Italy	Veneto	ITF3
Italy	Umbria	ITG1

Italy	Lazio	ITH1
Italy	Marche	ITG2
Italy	Friuli-Venezia Giulia	ITF4
Italy	Abruzzo	ITH2
Italy	Sicilia	ITI3
Italy	Molise	ITH3
Italy	Campania	ITH4
Italy	Basilicata	ITI1
Italy	Calabria	ITI2
Italy	Puglia	ITH5
Latvia	Latvia	LV00
Liechtenstein	Liechtenstein	LI00
Lithuania	Lithuania	LT00
Luxembourg (Grand-Duché)	Luxembourg (Grand-Duché)	LU00
Malta	Malta	MT00
Netherlands	Zeeland	NL34
Netherlands	Zuid-Holland	NL33
Netherlands	Noord-Holland	NL32
Netherlands	Noord-Brabant	NL41
Netherlands	Utrecht	NL31
Netherlands	Flevoland	NL23
Netherlands	Friesland (NL)	NL12
Netherlands	Limburg (NL)	NL42
Netherlands	Gelderland	NL22
Netherlands	Overijssel	NL21
Netherlands	Drenthe	NL13
Netherlands	Groningen	NL11
Norway	Vestlandet	NO05
Norway	Agder og Rogaland	NO04
Norway	Sør-Østlandet	NO03
Norway	Hedmark og Oppland	NO02
Norway	Oslo og Akershus	NO01
Norway	Trøndelag	NO06
Norway	Nord-Norge	NO07
Poland	Lubuskie	PL43
Poland	Zachodniopomorskie	PL42
Poland	Dolnoslaskie	PL51
Poland	Wielkopolskie	PL41
Poland	Opolskie	PL52
Poland	Pomorskie	PL63
Poland	Kujawsko-Pomorskie	PL61
Poland	Slaskie	PL22
Poland	Lódzkie	PL11
Poland	Malopolskie	PL21
Poland	Swietokrzyskie	PL33
Poland	Warminsko-Mazurskie	PL62
Poland	Mazowieckie	PL12
Poland	Podkarpackie	PL32

Poland	Podlaskie	PL34
Poland	Lubelskie	PL31
Portugal	Região Autónoma dos Açores (PT)	PT20
Portugal	Região Autónoma da Madeira (PT)	PT30
Portugal	Lisboa	PT17
Portugal	Algarve	PT15
Portugal	Centro (PT)	PT16
Portugal	Alentejo	PT18
Portugal	Norte	PT11
Romania	Vest	RO42
Romania	Nord-Vest	RO11
Romania	Sud-Vest Oltenia	RO41
Romania	Centru	RO12
Romania	Bucuresti - Ilfov	RO32
Romania	Nord-Est	RO21
Romania	Sud - Muntenia	RO31
Romania	Sud-Est	RO22
Slovakia	Bratislavský kraj	SK01
Slovakia	Západné Slovensko	SK02
Slovakia	Stredné Slovensko	SK03
Slovakia	Východné Slovensko	SK04
Slovenia	Zahodna Slovenija	SI02
Slovenia	Vzhodna Slovenija	SI01
Spain	Canarias (ES)	ES70
Spain	Galicia	ES11
Spain	Extremadura	ES43
Spain	Principado de Asturias	ES12
Spain	Ciudad Autónoma de Ceuta (ES)	ES63
Spain	Andalucia	ES61
Spain	Castilla y León	ES41
Spain	Cantabria	ES13
Spain	Comunidad de Madrid	ES30
Spain	Castilla-la Mancha	ES42
Spain	Ciudad Autónoma de Melilla (ES)	ES64
Spain	Pais Vasco	ES21
Spain	La Rioja	ES23
Spain	Comunidad Foral de Navarra	ES22
Spain	Región de Murcia	ES62
Spain	Aragón	ES24
Spain	Comunidad Valenciana	ES52
Spain	Cataluña	ES51
Spain	Illes Balears	ES53
Sweden	Västsverige	SE23
Sweden	Sydsverige	SE22
Sweden	Norra Mellansverige	SE31
Sweden	Småland med öarna	SE21
Sweden	Mellersta Norrland	SE32
Sweden	Östra Mellansverige	SE12

Sweden	Stockholm	SE11
Sweden	Övre Norrland	SE33
Switzerland	Espace Mittelland	CH02
Switzerland	Région lémanique	CH01
Switzerland	Nordwestschweiz	CH03
Switzerland	Zentralschweiz	CH06
Switzerland	Zürich	CH04
Switzerland	Ticino	CH07
Switzerland	Ostschweiz	CH05
United Kingdom	Northern Ireland	UKN0
United Kingdom	Cornwall and Isles of Scilly	UKK3
United Kingdom	Highlands and Islands	UKM6
United Kingdom	South Western Scotland	UKM3
United Kingdom	West Wales and The Valleys	UKL1
United Kingdom	Devon	UKK4
United Kingdom	Eastern Scotland	UKM2
United Kingdom	East Wales	UKL2
United Kingdom	Cumbria	UKD1
United Kingdom	North Eastern Scotland	UKM5
United Kingdom	Dorset and Somerset	UKK2
United Kingdom	Merseyside	UKD7
United Kingdom	Lancashire	UKD6
United Kingdom	Cheshire	UKD3
United Kingdom	Shropshire and Staffordshire	UKG2
United Kingdom	Greater Manchester	UKD4
United Kingdom	Gloucestershire, Wiltshire and Bristol/Bath area	UKK1
United Kingdom	Herefordshire, Worcestershire and Warks	UKG1
United Kingdom	Northumberland, Tyne and Wear	UKC2
United Kingdom	West Midlands	UKG3
United Kingdom	West Yorkshire	UKE4
United Kingdom	Tees Valley and Durham	UKC1
United Kingdom	North Yorkshire	UKE2
United Kingdom	Derbyshire and Nottinghamshire	UKF1
United Kingdom	South Yorkshire	UKE3
United Kingdom	Hampshire and Isle of Wight	UKJ3
United Kingdom	Berkshire, Bucks and Oxfordshire	UKJ1
United Kingdom	Leicestershire, Rutland and Northants	UKF2
United Kingdom	East Yorkshire and Northern Lincolnshire	UKE1
United Kingdom	Bedfordshire, Hertfordshire	UKH2
United Kingdom	Lincolnshire	UKF3
United Kingdom	Inner London	UKI1
United Kingdom	Surrey, East and West Sussex	UKJ2
United Kingdom	Outer London	UKI2
United Kingdom	East Anglia	UKH1
United Kingdom	Essex	UKH3
United Kingdom	Kent	UKJ4

2. List of removed regions for GDP and Patent data

No data for GDP and Patent applications

Country	Region	NUTS ID
Greece	Dytiki Makedonia	EL13
Greece	Ionia Nisia	EL22
Greece	Voreio Aigaio	EL41
Iceland	Iceland	IS00
Latvia	Latvia	LV00
Norway	Vestlandet	NO05
Norway	Agder og Rogaland	NO04
Norway	Sør-Østlandet	NO03
Norway	Hedmark og Oppland	NO02
Norway	Oslo og Akershus	NO01
Norway	Trøndelag	NO06
Norway	Nord-Norge	NO07
Spain	Ciudad Autónoma de Ceuta (ES)	ES63
Spain	Ciudad Autónoma de Melilla (ES)	ES64
Switzerland	Espace Mittelland	CH02
Switzerland	Région lémanique	CH01
Switzerland	Nordwestschweiz	CH03
Switzerland	Zentralschweiz	CH06
Switzerland	Zürich	CH04
Switzerland	Ticino	CH07
Switzerland	Ostschweiz	CH05

2. List of removed regions for Education data

No data for Education

Country	Region	NUTS ID
Denmark	Syddanmark	DK03
Denmark	Midtjylland	DK04
Denmark	Nordjylland	DK05
Denmark	Sjælland	DK02
Denmark	Hovedstaden	DK01
Finland	Pohjois-Suomi	FI1D
Finland	Etelä-Suomi	FI1C
Finland	Itä-Suomi	FI1B
Germany	Leipzig	DED5
Germany	Chemnitz	DED4
Iceland	Iceland	IS00
Italy	Emilia-Romagna	ITF5
Italy	Marche	ITG2
Latvia	Latvia	LV00
Liechtenstein	Liechtenstein	LI00
Norway	Vestlandet	NO05
Norway	Agder og Rogaland	NO04
Norway	Sør-Østlandet	NO03
Norway	Hedmark og Oppland	NO02
Norway	Oslo og Akershus	NO01

Norway	Trøndelag	NO06
Norway	Nord-Norge	NO07
Spain	Ciudad Autónoma de Ceuta (ES)	ES63
Spain	Ciudad Autónoma de Melilla (ES)	ES64
Switzerland	Espace Mittelland	CH02
Switzerland	Région lémanique	CH01
Switzerland	Nordwestschweiz	CH03
Switzerland	Zentralschweiz	CH06
Switzerland	Zürich	CH04
Switzerland	Ticino	CH07
Switzerland	Ostschweiz	CH05
United Kingdom	Merseyside	UKD7
United Kingdom	Cheshire	UKD3

2. List of removed regions for spatial dependence analysis and spatial regression models

Island regions

Country	**Region**	**NUTS ID**
Cyprus	Cyprus	CY00
Estonia	Estonia	EE00
Finland	Länsi-Suomi	FI19
France	Bretagne	FR52
Greece	Notio Aigaio	EL42
Portugal	Região Autónoma dos Açores (PT)	PT20
Portugal	Região Autónoma da Madeira (PT)	PT30
Spain	Canarias (ES)	ES70
Spain	Extremadura	ES43
Spain	Andalucia	ES61
Sweden	Övre Norrland	SE33