AGGLOMERATION ECONOMIES, KNOWLEDGE SPILLOVERS, TECHNOLOGICAL DIVERSITY AND SPATIAL CLUSTERING OF INNOVATIONS

Stefano Breschi

Introduction

This paper focuses on the localisation and clustering in space of innovative activities. Starting from the 80s, a flourishing literature has been growing on various topics related to the spatial dimension of technical change, from studies of innovative milieus, to industrial and technology districts, scientific parks and, more generally, local systems of innovation and production. Still more recently, several works conducted in the spirit of the so-called economic geography have begun to explore the geographical aspects of knowledge externalities and the localised relationships between private and University R&D and the localisation of innovative firms, stressing the fundamental role that the proximity among agents plays in mediating the processes of knowledge creation, transmission and appropriation. A robust result emerging from this literature is that innovations, far from being scattered and randomly distributed, tend to cluster geographically in some regions, provinces and towns. Moreover, some geographical areas are persistently better than others in producing innovations and their share of innovations far exceeds their share of manufacturing activities, thus providing evidence of some sort of localised increasing returns.

Drawing upon this literature, this paper addresses from an empirical perspective the analysis of the spatial patterns of innovation. To date, the empirical research on this subject has mainly focused upon the United States. The reason for that has to do with the lack of regional data both on innovative and economic activities for Europe. While the situation has changed a little bit in the last few years for what concerns regional data on innovative activities, thanks to the initiative of academic research centres and to the decision of Eurostat to start collecting European Patent Office data at the regional level, the same cannot be said for what regards other economic and infrastructure data. Most of these data are still collected at a spatial (NUTS 2) *and* sectoral (NACE 2-digit) level of aggregation, which is rather unsatisfactory from the perspective of the empirical researchers. On the one hand, knowledge spillovers and external economies are likely to take place more strongly at spatial scales closer to the definition of NUTS 3 than NUTS 2. On the other hand, industrial sectors within NACE 2-digit branches strongly differ both in terms of technological regimes (opportunity, appropriability and knowledge base) and structural features (economies of scale, barriers to entry and skills of workforce). Due to these data limitations, this paper will restrict the analysis to the case of Italy, for which consistent data at the NUTS 3 level (provinces) and NACE 3-digit could be found at relatively low costs. Preliminary investigations concerning other European countries have indeed evidenced quite serious difficulties in finding and collecting comparable data.

Agglomeration economies, knowledge spillovers and technological diversity

Several theoretical explanations have been proposed in the literature to account for the uneven capability to innovate across regions and for the highly skewed spatial distribution of innovative activities. In what follows, three major headings are identified that group together most of the issues examined in recent times. For each of them, a brief discussion is carried out stressing the problems posed for the empirical measurement and the methodology of spatial statistics.

Agglomeration economies

Pred (1966) is probably one of the first economists to explicitly advance the idea that the generation of innovations should concentrate in large towns and industrially strong regions to a greater extent than either population or manufacturing production thanks to various kinds of agglomeration economies. The idea rests upon the advantages accruing both on the demand and the supply side to firms located geographically close to each other. On the demand side, the benefits to firms are related to the reduction of search and transaction costs as well as to higher incentives to innovate due to a strong and rapidly growing local demand. On the supply side, the external economies may refer to the availability of a localised pool of specialised workers, to the exploitation of a fixed social capital, such as communication and transportation infrastructures, and to a greater circulation of ideas and information. However, as Swann (1996) has emphasised, agglomeration can also bring negative effects by producing congestion costs, both on the demand and the supply side.

The empirical measurement of agglomeration economies poses, however, several problems. A first problem relates to the fact that most of the benefits attached to clustering arises from forces of an invisible nature and that "leave no paper trail by which they may be measured and tracked" (Krugman, 1991). Until now, studies of agglomeration of innovative activities have in fact made use of employment and production data as proxies of agglomeration economies. However, even the use of proxy variables raises a second problem of measurement related to the notion of *spatial autocorrelation*. Indeed, most economic data are collected using administrative regions as spatial units of observation. It is nonetheless clear that such geographical partition does not correspond to actual markets and that any mismatch between the spatial unit of observation and the spatial extent of economic phenomena under consideration will result in spatial measurement errors and spatial autocorrelation between these errors in adjoining locations (Anselin, 1998). The problem of spatial autocorrelation may thus have serious consequences for econometric analyses based on contiguous cross-sectional data.

Knowledge spillovers

A peculiar type of benefit accruing from clustering has to do with knowledge spillovers, namely those ideas and findings borrowed by firm (industry) *i* from firm (industry) *j* (Griliches, 1995). In recent times, a growing number of authors have addressed the question of measuring to what extent knowledge spillovers are geographically localised (i.e. firms and industries located close in space to other innovative firms and industries are more likely to benefit from such spillovers than firms and industries located farther) and the impact of localised knowledge spillovers on the innovative performance of regions. The expectation is that knowledge spillovers are indeed localised because spatial proximity matters in facilitating the transmission and the acquisition of (tacit and complex) pieces of knowledge.

These empirical studies differ somewhat in terms of research design, although they all focus on the United States. Jaffe, Trajtenberg and Henderson (1993) compare spatial patterns of citations, that are assumed to be a good proxy for knowledge spillovers, for a sample of Universities and corporate patents with a control sample of randomly drawn patents, finding significant evidence that citation patterns (i.e. knowledge spillovers) are strongly localised at both the state and SMSA levels.[1] Jaffe (1989) uses patent counts, while Audretsch and Feldman (1996), Feldman (1994), and Acs, Audretsch and Feldman (1994) utilise a 1982 dataset of innovation counts compiled by the U.S. Small Business Administration. These studies all use federal states as spatial unit of observation and they measure localised knowledge spillovers through an index of *geographic coincidence* of industry R&D and University research for each state. At this level, they find evidence of a strong and positive relationship between innovative activity and both industry and University research. However, the evidence is much weaker and mixed for what concerns the role of spatial proximity in affecting the strength of knowledge spillovers. In the words of Jaffe (1989), "there is only weak evidence that spillovers are facilitated by the geographic coincidence of university and research labs within the state. (…) the effect comes more clearly within technical areas than it does in the total across areas " (p. 968). Anselin, Varga and Acs (1997) argue that the lack of uniform results regarding the importance of spatial interaction at the *local* level could be due to the fact that the unit of analysis (the state) only partially captures this interaction, but also to the formal specification of local spatial interaction in the form of a geographic coincidence index. They propose to measure knowledge spillovers through the a set of *spatially lagged* variables designed to capture the effect of University and private R&D in counties surrounding a SMSA within a given distance band from the centre of the SMSA. Their results provide convincing evidence that localised knowledge spillovers from University and private R&D significantly affect innovative performance of SMSAs.

In sum, this short review of the empirical literature on localised geographic spillovers points out two crucial questions for empirical research. On the one hand, the choice of the spatial unit of observation is absolutely fundamental to detect localised knowledge spillovers. On the other hand, the smaller the spatial unit of observation the more essential becomes to correctly specify the spatial extent of local interaction. The choice of spatial weights used to build spatially lagged effects of research and innovative activities have then to be based upon some prior information on the geographic extent of knowledge spillovers (e.g. from micro studies).

Technological diversity

A debate that has regained attention in the recent times is whether the regional specialisation within a narrow set of economic activities is more conducive to knowledge spillovers and innovation or if diversity and variety, by widening the pool of complementary competencies, better encourage innovative activities. Generally speaking, results of empirical analyses provide strong support for the thesis that the co-location of diverse and related industries tends to promote innovation and growth (Feldman and Audretsch, 1995; Glaeser et al., 1992). However, more empirical research needs to be done, particularly for what regards the identification of *related* industries and technologies. This paper will adopt a measure of *technological proximity* based upon the analysis of co-classification codes contained in patent documents.

The three issues just discussed provide the base upon which this paper is built. In particular, this paper will attempt to test two broad hypotheses:

- Agglomeration (dis)economies working *within* regions (i.e. *without* taking into account the spatial interaction with contiguous regions) account for the uneven capability to innovate across regions. In other terms, regions enjoying higher levels of agglomeration economies and knowledge spillovers and lower levels of congestion costs also tend to produce a higher number of innovations.
- Technological diversity and related diversification of regions tend to encourage innovation by promoting higher levels of knowledge spillovers as well as multiple and complementary sources of new ideas.

Before testing these two hypotheses, however, Section 3 provides a brief description of the data used, while in Section 4 an exploratory analysis of spatial data is carried out using methodology and techniques borrowed from spatial statistics.

Sources of data

This paper makes use of several sources of data. A brief description of each source is provided below.

Patent data

This paper will use the EPO-CESPRI database. The data set contains all *patent applications* to EPO (European Patent Office) from 1978 onward, by firms and institutions of all countries seeking protection for their innovations in any of the 18 countries adhering to the Munich Convention, which established EPO. For each patent applicant the data set also identifies its spatial localisation as given by the address contained in the patent document.

Patents have been classified according to a technology-oriented classification that distinguishes 5 technology areas and 30 technology sub-fields based on the International Patent Classification (IPC). This classification has been elaborated jointly by FhG-ISI, the French Patent Office (INPI) and the Observatoire de Sciences and des Techniques (OST). Its most updated version is reported in Appendix 1. A concordance table between IPC codes and NACE 3-digit codes has been also built and reported in Appendix 1.

Patent applications have been processed at the *regional* level taking NUTS 3 as spatial unit of observation. This spatial partition corresponds to 95 administrative provinces for the case of Italy. The number of patents applied for by provinces in the time period from 1987 to 1994, by technological class, has been used here as a measure of innovative strength. To this purpose, it should be noted that in addition to measuring innovation (with all the strengths and weaknesses and the methodological problems associated to this; see Griliches, 1991), patent applications are a very good indicator of firms' and regions' technological competencies. The fact that firms located in a certain region have applied for patents in a given technological field means that such firms are at, or close to, the technological frontier and have advanced technological competencies in that field. A word of caution needs to be spent about the use of applicant's address to locate patents in space. This approach has the drawback of overestimating the actual degree of spatial concentration of innovative activities, given the propensity of headquarters to patent innovations developed by establishments located in other regions. However serious, this problem is likely to cause concern only in those technological fields (notably, chemical) in which multi-plant firms play a very important role.

Patent data have been also used to measure the *knowledge proximity* between different technological fields. Taking *all* patent applications to the EPO, Breschi et al. (1998) have built a matrix of knowledge proximity across the 30 technological classes, by considering the frequency of co-occurrences of technological classes in *primary* and *secondary* classification codes. Such matrix has then been used to identify clusters of *related technologies*; namely technologies close from a cognitive perspective.

Other innovation data

Another commonly used innovation indicator is R&D expenditures. Unfortunately, this statistics only reports aggregate data and it is not available at the sectoral level. Moreover, most national statistical offices and Eurostat itself only report data at the NUTS 2 level (regions). In this paper, R&D expenditures at the NUTS 2 level will be used under the implicit assumption that the benefits flowing from R&D spread equally across all NUTS 3 provinces belonging to the same region. In addition to that, use will also be made of data on investments in innovative capital goods at the NUTS 2 level. Both data are drawn from the Innovation Survey carried out by ISTAT in 1992 within the Community Innovation Survey (CIS) project.[2]

Employment data

The 1991 Italian Industrial Census has been used to measure the industrial strength of provinces. In particular, this paper will make use of data on the number of firms, establishments and employees at the NUTS 3 and NACE 3-digit levels. Data on population and area of provinces for the year 1991 were drawn from ISTAT (1993).

Infrastructure and milieus data

Data on the number of business telephone subscribers in 1989 at the NUTS 3 level will be used as a proxy of agglomeration economies arising from increased circulation of ideas and information (SIP, 1990). An attempt to measure congestion costs (i.e. costs arising from clustering of firms) has been made by using data on road traffic provided by the Italian Auto Industry Association (ANFIA, 1994). In particular, an index of congestion costs at the NUTS 3 level was built as the number of circulating vehicles per kilometre of non-urban road. Lorries were assumed to weight double.

Exploratory spatial data analysis

This section provides some exploratory data analysis on the spatial distribution of innovative and manufacturing activities in Italy at the NUTS 3 level.

The first indicator examined here is an aggregate index of spatial concentration, namely the Herfindahl equivalent number. This is defined as:

(1) $$HEN = 1 / \sum_{i=1}^{n} S_{ij}^{2}$$

where S_{ij} is the share of a given variable of province *i* in sector *j* and *n* is the total number of provinces in a country. The *HEN* index has been calculated for each of the 30 technological sub-fields using four different variables: patents, employment, number of establishments, population (Table 1).

Comparing patents and employment one can see that innovative activities display much higher levels of spatial concentration than manufacturing activities across all industrial sectors. This is true also if we compare the spatial concentration of patents and establishments. The difference in the values of HEN index is so large that even a moderate bias in the spatial attribution of patenting activity would not change the result too much.

A closer look at the data, however, also reveals the existence of rather large differences across sectors in the extent of spatial concentration of innovations and manufacturing activities. For what regards innovations, they are highly spatially concentrated in most chemical and electrical-electronic sectors. However, also productive activities in these sectors appear to be relatively more spatially concentrated than in other sectors (with some notable exceptions, like food chemistry and basic materials chemistry). A more mixed picture emerges instead with respect to mechanical engineering and industrial equipment sectors, although a relatively more spatially diffused pattern seems to characterise all these branches, both in terms of patenting and manufacturing activities. Also in this case, noteworthy exceptions are represented by transports and engines industries.

Table 1

Herfindahl equivalent numbers
(Italy, NUTS 3)

Class	*Patents*	*Employment*	*Establishments*	*Population*
1. Electrical engineering	4.7	16.8	22.5	44.5
2. Audiovisual technology	3.1	9.8	33.8	44.5
3. Telecommunications	2.5	11.0	29.6	44.5
4. Information technology	2.7	4.3	11.6	44.5
5. Semiconductors	1.6	11.9	14.5	44.5
6. Optics	5.5	6.3	11.2	44.5
7. Control technology	6.5	8.3	12.5	44.5
8. Medical technology	11.9	23.9	34.7	44.5
9. Organic chemistry	2.5	14.8	19.4	44.5
10. Polymers	2.1	14.8	19.4	44.5
11. Drugs	3.9	5.1	7.8	44.5
12. Biotechnology	3.0	-	-	44.5
13. Materials	5.6	22.6	22.7	44.5
14. Food chemistry	8.2	54.7	68.1	44.5
15. Basic materials chemistry	3.5	22.1	29.5	44.5
16. Chemical engineering	6.9	22.1	26.5	44.5
17. Coating technology	4.0	22.7	22.3	44.5
18. Materials processing	13.3	44.1	55.6	44.5
19. Thermal processes	14.1	22.2	29.5	44.5
20. Environmental technology	7.5	-	-	44.5
21. Machine tools	14.2	14.8	17.0	44.5
22. Engines	7.4	15.5	19.8	44.5
23. Mechanical elements	8.9	19.6	21.2	44.5
24. Handling	15.6	19.7	29.1	44.5
25. Food processing	19.5	25.9	22.7	44.5
26. Transport	4.5	8.4	40.9	44.5
27. Nuclear engineering	4.2	-	-	44.5
28. Space technology	4.2	-	-	44.5
29. Consumer goods	12.8	-	-	44.5
30. Civil engineering	16.4	22.6	22.7	44.5

A final result to point out is that both manufacturing and patenting activities are more spatially concentrated than population thus providing indirect evidence of agglomeration economies and localised increasing returns.

While informative about the degree of spatial concentration of a variable, the type of analysis conducted so far does not provide much information about the way the variable is *spatially structured*. A given value of spatial concentration can indeed correspond to different spatial configurations of data. Several techniques have been developed in the field of spatial statistics in order to detect spatial patterns and some of them will be applied to our data (Haining, 1990).

Before turning to that, however, a few maps are reported that illustrate the spatial distribution of patenting activities for a number of technological fields. In particular, maps have been produced for five technological fields that appear to be particularly representative of broader technological areas. An examination of these maps provides some preliminary insights into the spatial arrangement of innovative activities and the main differences existing across sectors.

The spatial distribution of patents across NUTS 3 Italian provinces in handling technologies (Map 1) is in many respects very similar to that of other mechanical engineering and industrial equipment sectors and could be defined as a "localised diffusion" pattern of innovative capabilities. As a matter of fact, such competencies are largely diffused across contiguous provinces (and across NUTS 2 regional borders), but at the same time they do not spread all over the country remaining within the boundaries of specific bunches of provinces. A rather similar pattern emerges with respect to control technology (Map 2), even though there is less evidence of diffusion of innovative capabilities across contiguous provinces and more concentration in a few key areas is observed. Maps 3 and 4 that refer, respectively, to drugs and telecommunications sectors illustrate quite neatly a spatial pattern common to most chemicals and electrical and electronic sectors. Innovations tend to concentrate in just one core province and they hardly diffuse to neighbouring areas. Finally, transport industry (Map 5) represent a case where two strong innovative provinces (located far apart from each other) are present, without much spatial interaction with contiguous provinces.

As already mentioned, mapped data contain not only information about the values of variables but also information about how these values are arranged in space. However, for the sake of parsimony and statistical testing, it is obviously desirable to have some numerical summary of the observed spatial patterns. In this respect, a notion that is largely employed in spatial statistics is that of *spatial autocorrelation*. In very general terms, spatial autocorrelation exists whenever a variable exhibits a regular pattern over space and its values at a set of locations depend on values of the same variable at other locations. More specifically, a situation of positive autocorrelation arises when locations "close" in space take similar values of a given variable, whereas a situation of negative autocorrelation is present when nearby locations take dissimilar values of a certain variable. Spatial autocorrelation analysis can thus be helpful in identifying regularities in the spatial patterns and, particularly, to identify *clusters* of innovative provinces. Please note that the word *close* above has been put in quotes because the meaning of closeness and the definition of a measure of spatial interaction between pairs of locations is a crucial task in the calculation of autocorrelation statistics. I will return later to this point.

Several indexes have been proposed in the spatial statistics literature to assess the presence of spatial autocorrelation. In this paper, I will present results based on one of the most commonly used statistics of spatial autocorrelation, namely the Moran's I. Formally, this statistic is:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \mu)\cdot[x_j - \mu]}{\sum_{i=1}^{n}(x_i - \mu)^2} \qquad (2)$$

where n is the number of observations, w_{ij} is the element in a spatial weights matrix W corresponding to the observation pair (i,j), x_i and x_j are observations for location i and j (with mean μ) and S_0 is a scaling constant

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$$

namely, the sum of all weights.

For a row-standardised spatial weights matrix, the normalising factor S_0 equals n (since each row sums to one), and the statistics simplifies to a ratio of a spatial cross product to a variance:

$$I^* = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \mu)\cdot[x_j - \mu]}{\sum_{i=1}^{n}(x_i - \mu)^2} \qquad (2')$$

Moran's index is very similar but not equivalent to a correlation coefficient and it is not centred on 0. The theoretical mean of I is $-1/(n-1)$. This means that the expected value is negative and it depends only on the sample size (n). However, the mean will tend to zero as the sample size increases. A value of the Moran's I coefficient larger (smaller) than its expected value indicates therefore positive (negative) spatial autocorrelation. The problem of testing the null hypothesis (H_0) of no spatial autocorrelation against the alternative (H_1) that the data are spatially autocorrelated , however, is typically based upon a standardised z-value. This is computed by subtracting the theoretical mean and dividing by the theoretical standard deviation: $z_I = [I - E(I)]/SD(I)$, where $E(I)$ is the theoretical mean and $SD(I)$ is the theoretical standard deviation. The theoretical variance of Moran's I depends on the stochastic assumptions that are made. One of the most common approaches is to assume that the variable in question follows a normal distribution (assumption N). In other terms, under the normality assumption for H_0 the observed map, consisting of n observations, is the result of n independent drawings from a normal population. A second often used approach is to assume that the observed map of values is one possible arrangement of the set of n values. The reference distribution for the statistics is then obtained by considering the $n!$ possible permutations of the n values. In other terms, under this hypothesis (often referred to as randomisation assumption R), each value observed could equally likely have occurred at all locations. Cliff and Ord (1981) provides a technical discussion and detailed expressions for the moments of Moran's I statistic under the two assumptions. Moreover, they also show that the z-values under the two assumptions follow asymptotically a standard normal distribution so that their significance can be judged by means of a standard normal table.

Before reporting results, a few words must be spent about the form of the spatial weights matrix W. The selection of a spatial weighting matrix is indeed the most important step in calculating a spatial autocorrelation statistic. The generic element w_{ij} of matrix W - where (i,j) corresponds to a given location pair- has to be conceptualised as a measure of the potential spatial interaction between observations at different locations. Elements that are zero indicate a lack of spatial interaction between the two locations (by convention the diagonal elements of the weights matrix are set to zero), while positive elements measure the strength of the interaction between the two locations. There are, of course, many different ways in which this matrix can be specified. The simplest weighting matrix is defined in terms of *simple contiguity*: the generic element w_{ij} is assigned a value of 1 if i and j are neighbours (i.e. they share a border) and 0 otherwise. More complex specifications include distance contiguity (i.e. having centroids within a critical distance band), or in function of inverse distance or squared inverse distance. Throughout this paper, however, a simple definition of the spatial weight matrix based on contiguity will be adopted. Even though this specification provides an admittedly crude representation of the extent of spatial interaction between pairs of provinces, it nonetheless permits to reach some preliminary results upon which more sophisticated definitions can be implemented.

Moran's I* will be calculated here for three orders of spatial contiguity (Moran correlogram). A matrix $W(1)$ will contain elements $w_{ij} = 1$ if i and j are first-order neighbours (i.e. share a boundary) and 0 otherwise. A second-order contiguity matrix $W(2)$ is then defined in a recursive fashion as containing elements $w_{ij} = 1$ if unit i is first-order contiguous to another unit that itself is first-order contiguous to unit j, and zero otherwise. A third-order contiguity matrix $W(3)$ is then defined in a similar manner. All matrixes are row-standardised. The use of second- and third-order contiguity matrices provides a more detailed description and allows a more accurate test of spatial pattern properties compared with the simple use of a first-order contiguity matrix.

Moran's I* has been calculated for patent and employment data. Concerning patents, the absolute number of patents has been used as well as the standardised value by sectoral employment and by total population. For what concerns employment, absolute values by sector have been instead used to compute Moran's statistic. Results are reported in Tables 2 to 5.

Looking at the patenting activity, a first important result emerging from the data refers to the large differences across sectors in the extent and statistical significance of spatial autocorrelation. Taking the absolute number of patents (Table 2), Moran's I* based on first-order contiguity is positive and statistically significant only in 8 sectors. Very interestingly, all these sectors belong to the mechanical engineering and industrial equipment industries, with the only exception of the consumer goods class. What the data suggest is therefore that in most mechanical industries innovative activities tend to cluster together in bunches of contiguous provinces. For some of these sectors (e.g. thermal processes and food processing) also second- and third-order spatial autocorrelation is positive and highly significant, thus providing evidence that innovative capabilities spread across a wide set of contiguous provinces. With respect to other sectors, no clear pattern seems to emerge. However, it is rather interesting to note that third-order spatial autocorrelation is positive and statistically significant in audio-visual, IT and control technology sectors, thus indicating that the spatial organisation of innovative activities in these sectors looks like a set of non contiguous "islands" located not too far from each other. Finally, a rather puzzling result emerges when the total NUTS 3 population weights patents. The value of Moran's I* increases in all sectors, even though it still remains statistically not significant in most chemical and electronic industries.

Table 2

Moran correlogram
(Patents, n=95)

Class	*Order of contiguity*								
	W(1)			W(2)			W(3)		
		N	*R*		*N*	*R*		*N*	*R*
1. Electrical engineering	0.029			0.015			0.038		
2. Audiovisual technol.	-0.014			-0.014			**0.048**		*
3. Telecommunications	-0.020			-0.017			0.041		
4. Information technol.	-0.039			-0.022			**0.108**	*	*
5. Semiconductors	-0.015			-0.012			-0.018		
6. Optics	-0.020			-0.022			0.065		
7. Control technology	0.006			0.002			**0.101**	*	*
8. Medical technology	0.002			0.051			0.034		
9. Organic chemistry	0.004			0.004			-0.027		
10. Polymers	0.003			-0.011			-0.018		
11. Drugs	0.007			0.006			-0.028		
12. Biotechnology	-0.018			0.004			-0.032		
13. Materials	0.001			0.002			-0.017		
14. Food chemistry	0.003			**0.064**		*	0.036		
15. Basic materials che.	0.005			0.003			0.006		
16. Chemical engineer.	**0.040**		*	0.021			**0.043**		*
17. Coating technology	-0.014			-0.012			0.023		
18. Materials processing	**0.201**	*	*	0.060			0.035		
19. Thermal processes	**0.141**	*	*	**0.090**	*	*	**0.086**	*	*
20. Environmental tech.	0.000			-0.007			0.067		*
21. Machine tools	**0.094**		*	0.024			0.059		
22. Engines	0.005			-0.010			0.058		
23. Mechanical elements	0.037			0.038			**0.124**	*	*
24. Handling	**0.133**	*	*	0.065			0.092		*
25. Food processing	**0.223**	*	*	**0.122**	*	*	**0.138**	*	*
26. Transport	0.001			-0.011			**0.100**	*	*
27. Nuclear engineering	-0.035			-0.016			-0.014		
28. Space technology	-0.025			0.057			-0.052		
29. Consumer goods	**0.171**	*	*	0.000			0.029		
30. Civil engineering	**0.140**	*	*	0.048			**0.089**	*	

Note: statistical significance of Moran's I* under the normal assumption (N) and the randomisation assumption (R).
A star (*) indicates statistical significance at the 10% level.

Table 3

Moran correlogram
(Patents per employee, n=95)

Class	*Order of contiguity*								
	W(1)			W(2)			W(3)		
		N	R		N	R		N	R
1. Electrical engineering	**0.181**	*	*	0.028			**0.106**	*	*
2. Audiovisual technol.	0.058			-0.011			-0.018		
3. Telecommunications	-0.035			-0.014			-0.027		
4. Information technol.	-0.028			-0.035			**0.083**	*	*
5. Semiconductors	0.011			0.002			-0.043		
6. Optics	-0.067			-0.001			0.005		
7. Control technology	0.061			-0.041			0.020		
8. Medical technology	0.019			**0.204**	*	*	0.006		
9. Organic chemistry	-0.047			-0.031			-0.029		
10. Polymers	-0.012			-0.013			-0.014		
11. Drugs	-0.012			-0.031			0.025		
12. Biotechnology									
13. Materials	-0.049			-0.033			-0.020		
14. Food chemistry	0.012			0.008			**0.143**	*	*
15. Basic materials che.	-0.065			0.002			-0.018		
16. Chemical engineer.	0.028			0.020			0.056		
17. Coating technology	**0.334**	*	*	-0.018			-0.039		
18. Materials processing	**0.189**	*	*	0.069			-0.020		
19. Thermal processes	0.088			0.057			0.022		
20. Environmental tech.									
21. Machine tools	-0.044			-0.016			0.024		
22. Engines	-0.019			0.006			-0.017		
23. Mechanical elements	0.023			-0.048			0.015		
24. Handling	0.075			0.060			0.027		
25. Food processing	**0.132**	*	*	-0.043			-0.059		
26. Transport	-0.031			-0.039			0.020		
27. Nuclear engineering									
28. Space technology									
29. Consumer goods									
30. Civil engineering	-0.034			-0.021			-0.028		

Note: statistical significance of Moran's I* under the normal assumption (N) and the randomisation assumption (R). A star (*) indicates statistical significance at the 10% level.

Table 4
Moran correlogram
(Patents per capita, n=95)

Class	*Order of contiguity*								
	W(1)			W(2)			W(3)		
		N	R		N	R		N	R
1. Electrical engineering	**0.276**	*	*	**0.103**	*	*	0.055		
2. Audiovisual technol.	0.015			0.006			0.023		
3. Telecommunications	0.007			0.010			0.007		
4. Information technol.	0.064			0.046			**0.128**	*	
5. Semiconductors	0.029			0.022			0.028		
6. Optics	**0.236**	*	*	**0.072**		*	0.006		
7. Control technology	**0.189**	*	*	**0.166**	*	*	**0.109**	*	*
8. Medical technology	0.042			**0.124**	*	*	**0.066**	*	
9. Organic chemistry	**0.107**	*	*	0.012			0.057		
10. Polymers	0.075			0.011			0.051		
11. Drugs	0.010			0.036			0.039		
12. Biotechnology	0.020			0.020			0.022		
13. Materials	0.047			0.032			0.013		
14. Food chemistry	0.017			0.014			**0.049**		*
15. Basic materials che.	0.070			0.004			0.026		
16. Chemical engineer.	**0.216**	*	*	**0.206**	*	*	**0.209**	*	*
17. Coating technology	**0.170**	*	*	0.041			0.033		
18. Materials processing	**0.090**	*	*	**0.063**		*	0.026		
19. Thermal processes	**0.175**	*	*	**0.131**	*	*	0.016		
20. Environmental tech.	**0.185**	*	*	**0.109**	*	*	0.028		
21. Machine tools	0.099		*	**0.089**	*	*	0.024		
22. Engines	0.036			0.051			0.040		
23. Mechanical elements	**0.217**	*	*	**0.206**	*	*	**0.183**	*	*
24. Handling	**0.164**	*	*	**0.080**		*	**0.080**		*
25. Food processing	**0.292**	*	*	**0.125**	*	*	**0.166**	*	*
26. Transport	**0.142**	*	*	0.033			**0.157**	*	*
27. Nuclear engineering	0.027			0.046			0.052		
28. Space technology	0.105			0.010			0.039		
29. Consumer goods	**0.277**	*	*	0.018			0.003		
30. Civil engineering	**0.319**	*	*	**0.193**	*	*	**0.245**	*	*

Note: statistical significance of Moran's I* under the normal assumption (N) and the randomisation assumption (R).
A star (*) indicates statistical significance at the 10% level.

When we turn to employment, a striking result that emerges is that the value of the first-order Moran's I* is generally higher compared with patenting activity: 15 sectors out of 30 display a positive and statistically significant coefficient. Moreover, in most mechanical industries even second- and third-order spatial autocorrelation is positive and statistically significant. What these results can indicate is that manufacturing production tend to present a much more diffused spatial pattern compared to innovative activities, by clustering around groups of contiguous and interconnected areas to a greater extent than innovations, suggesting that centripetal forces in innovative activities can have a stronger impact than in manufacturing. Note, however, that in most electronic and several chemical sectors there is no spatial autocorrelation also with respect to employment, therefore indicating that in these sectors both innovation and production tend to be realised within the boundaries of few isolated provinces.

Further information on the spatial patterns of innovative activities can be derived from inspecting the so-called *Moran scatterplot maps*. As explained above, Moran's I gives a formal indication of the degree of linear association between a vector of observed values X (of dimension n by 1) and a weighted average of the neighbouring values (spatial lags) WX, where W stands for the spatial weights matrix. All possible pairs (x_i, Wx_i) can then be computed and plotted in a bivariate scatterplot. Four types of spatial association can be distinguished in the set of pairs (x_i, Wx_i): two forms of "positive" spatial association- i.e. association between similar values, large x_i surrounded by large Wx_i, or small x_i surrounded by small Wx_i; and two forms of "negative" spatial association- i.e. large x_i surrounded by small Wx_i, or small x_i surrounded by large Wx_i. The position of each observation in the scatterplot can then be converted into distinct colours or shades for visualisation into a Moran scatterplot map. An inspection of such maps can provide several information regarding the way data are arranged in space. In the first place, it may suggest the existence of different regimes of spatial association in different subsets of data, e.g. positive association in one part of the landscape and negative association in another. In the second place, Moran scatterplots are also useful to identify "outliers" in the global patterns of association indicated by the I statistic, namely observations that are extreme with respect to the central tendency reflected by Moran's I coefficient (Anselin, 1988).

Table 5

Moran correlogram
(Employment, n=95)

Class	W(1)	*N*	*R*	W(2)	*N*	*R*	W(3)	*N*	*R*
1. Electrical engineering	**0.128**	*	*	0.020			**0.078**		*
2. Audiovisual technol.	-0.015			-0.014			0.005		
3. Telecommunications	0.012			-0.022			0.021		
4. Information technol.	-0.011			-0.026			**0.128**	*	*
5. Semiconductors	0.066			-0.028			0.057		
6. Optics	**0.112**	*	*	0.024			-0.016		
7. Control technology	**0.131**	*	*	-0.005			0.025		
8. Medical technology	0.068			0.027			**0.067**		*
9. Organic chemistry	**0.115**	*	*	0.000			0.004		
10. Polymers	**0.115**	*	*	0.000			0.004		
11. Drugs	0.043		*	-0.022			-0.014		
12. Biotechnology									
13. Materials	**0.268**	*	*	**0.136**	*	*	**0.151**	*	*
14. Food chemistry	0.058			0.078		*	0.068		
15. Basic materials che.	0.062			0.017			-0.040		
16. Chemical engineer.	**0.237**	*	*	**0.106**	*	*	**0.133**	*	*
17. Coating technology	**0.265**	*	*	**0.144**	*	*	**0.170**	*	*
18. Materials processing	0.079			**0.106**	*	*	**0.113**	*	*
19. Thermal processes	**0.192**	*	*	0.071		*	**0.108**	*	*
20. Environmental tech.									
21. Machine tools	**0.245**	*	*	**0.104**	*		**0.154**	*	*
22. Engines	**0.108**	*	*	**0.131**	*		**0.126**	*	*
23. Mechanical elements	**0.187**	*	*	**0.146**	*	*	**0.156**	*	*
24. Handling	**0.171**	*	*	0.055		*	**0.099**	*	*
25. Food processing	**0.297**	*		**0.153**	*	*	**0.180**	*	
26. Transport	-0.068			-0.002			0.036		
27. Nuclear engineering									
28. Space technology									
29. Consumer goods									
30. Civil engineering	**0.268**	*		**0.136**	*		**0.151**	*	*

Note: statistical significance of Moran's I* under the normal assumption (N) and the randomisation assumption (R). A star (*) indicates statistical significance at the 10% level.

As before, Moran scatterplot maps have been plotted only for a number of representative sectors. Results are reported in Maps 6 to 11.

Map 6 related to the handling sector is a fairly faithful representation of spatial patterns in most mechanical engineering and industrial equipment branches. What one observes from the map is the existence of four systems of *clustered innovative provinces* (white areas). Provinces in these systems are highly innovative and are also surrounded by other equally innovative provinces. It may be then argued that in these sectors knowledge spillovers exert their benefits over a fairly wide set of spatial areas. There are two additional points to note. First, localised clusters cut across regional (NUTS 2) boundaries. For example, one of the local innovation systems comprises the provinces of Mantova (MN), Modena (MO), Bologna (BO), Ravenna (RA), Firenze (FI) and Pisa (PI) that belong to three different NUTS 2 regions. Therefore, any analysis of spatial patterns of innovations based upon such geographical partition seriously risk to underestimate clustering effects. Second, it is worth noting that all clusters of innovative provinces are located in the north of the country and that provinces located in the south constitute a homogenous cluster of non-innovative areas, i.e. non-innovative provinces surrounded by equally non-innovative provinces. It is precisely this spatial pattern that accounts for the positive and high levels of spatial autocorrelation registered in these sectors (see above). A very similar pattern can be also found in the materials processing sector (see Map 9).

Maps 10 and 11 that refer, respectively, to drugs and telecommunications industries, tell us a completely different story. For what concerns drugs industry, what one observes is the existence of three local systems of innovation: the first system comprises the provinces of Vicenza (VI), Padova (PD) and Venezia (VE); the second system includes the provinces of Firenze (FI), Bologna (BO), Siena (SI), Modena (MO) and Pisa (PI): while the third system is centered around the provinces of Milano (MI) and Varese (VA). Finally, a pole of innovation is also present in the province of Rome (RM). What distinguishes the case of drugs from the case of mechanical industries discussed above is that innovative provinces are surrounded, in general, by less (or non) innovative ones. The spatial extent of local innovation systems is thus much narrower. This is even more so in the case of telecommunications industry. Innovative activities are highly spatially concentrated in three sets of provinces: Milano (MI), Torino (TO), and Rome (RM) with l'Aquila (AQ). Surrounding provinces are in general scarcely innovative or non-innovative at all. Therefore, one can conclude that in these industrial sectors knowledge spillovers tend to remain within the borders of the larger metropolitan areas.

Finally, Maps 7 and 8 report Moran scatterplots for two instruments sectors, namely control and medical technology. With respect to medical technology, there is evidence of a strong local system of innovation that comprises the provinces of Bologna (BO), Modena (MO) and Firenze

(FI). Note also that this local system is surrounded by scarcely or non-innovative provinces. For what regards control technology, a strong local system of innovation comprises the provinces of Milano (MI), Pavia (PV) and Varese (VA). However, other less integrated clusters of provinces can be found in other parts of the country.

The main conclusions reached by this exploratory data analysis can be summarised as follows:

- Innovative and manufacturing activities are more spatially concentrated than population. At the same time, innovations are considerably more spatially concentrated than production.
- There are large differences across sectors in the degree of concentration of innovation and production. Concentration is relatively lower in most mechanical engineering and industrial equipment sectors and it is relatively higher in most electrical-electronic and chemicals-drugs sectors.
- There are large differences across sectors also in the spatial patterns of innovative and manufacturing activities. While there is evidence of positive and statistically significant spatial autocorrelation in several mechanical sectors, innovative and productive activities in most chemicals and electrical-electronic sectors are concentrated in few technological "poles" that are surrounded by non-innovative provinces. In general, production displays higher levels of spatial autocorrelation than innovations.
- Looking only at innovations, an inspection of Moran scatterplot maps has shown that in mechanical industries and, to a less extent, instruments sectors there are several local systems of clustered innovative provinces. All these systems are, however, located in the northern part of the country. On the other hand, in most chemical and electronic industries innovative activities tend to concentrate in few metropolitan areas that are surrounded by non-innovative provinces.

Modelling framework

After having described the main spatial properties of innovation and manufacturing data, this section turns the attention to the problem of testing the hypotheses put forward in section 2. To this purpose, a model is estimated where the dependent variable is the number of patents attributed to a specific industry in a particular NUTS 3 province in the period 1987-94 (P).

To reflect the strength of agglomeration economies, two explanatory variables were used here (both measured in logs): the number of employees in the same industry within the province (OWNEMP) and the number of employees in other manufacturing sectors within the province (OTHEMP). Both variables are clearly proxies for all benefits accruing from the clustering of firms within the same or related industries and they are expected to have a positive impact on the number of innovations produced within any given province and industry. An additional variable has been also included in the attempt to capture the strength of information and communication flows among the firms located within a province: the number of 1989 business telephone subscribers (also expressed in logs) (PHSUBS).

In order to capture the impact of local knowledge spillovers two variables have been included in the model: the total R&D expenditures (RDEXP) and the total investments in innovative capital goods (KGEXP). Both measures (expressed in logs) were drawn from the Italian Innovation Survey and, unfortunately, are only available at the NUTS 2 level. As a consequence, it has to be assumed here that the impact of both types of investment affects equally all provinces located within the same NUTS 2 region. Concerning the effect upon the number of expected patents, it is a robust result of the recent empirical literature that higher levels of R&D expenditures should encourage higher levels of innovative activity. Less clear is the likely impact of investments in new capital goods on the innovative performance of a given province. To the extent that such an investment measures the commitment to innovation by firms located in a province, one should expect that high levels of investments in new capital goods be also matched by high numbers of innovations produced. However, to the extent that this sort of investment indicates the relative importance of *embodied* technical change and of less formalised and autonomous innovative activities, one could also expect an inverse relationship between the amount of expenditures in new capital goods and the total number of innovations generated within a given province.

The impact on innovation of diversity in the knowledge base within a given NUTS 3 province has been assessed here through different measures. In the first place, for each industry and for each NUTS 3 province, the total number of patents attributed to *related industries*

(RELTEC) in the period 1978-86 has been used to measure how strong is the province within industries sharing a common knowledge base (see § 3). In the second place, the total number of patents attributed to *other industries* (OTHTEC) in the period 1978-86 has been also employed to assess the overall innovative strength of the province. Finally, the Herfindahl index (HERF) has been included to capture the extent to which any given province is specialised into a small set of technologies or, conversely, is characterised by a wide and differentiated array of competencies.

Finally, two variables have been also considered in order to control for other effects that could affect the innovative performance of provinces. On the hand, the extent to which congestion costs impact negatively on the innovative performance of provinces has been evaluated through a variable that measures the number of circulating cars and lorries per kilometre of non-urban roads within each NUTS 3 province (CONG). On the other hand, for any given NUTS 3 province an index number (Italy=100) summarising its relative position in terms of per-capita value added at factor costs (VADDED) in the year 1991 has been also included.

Thus, the model to be estimated becomes the following:

(3) $$P_{ij} = f(RELTEC_{ij}, OTHTEC_{ij}, HERF_{i.}, OWNEMP_{ij}, OTHEMP_{ij}, RDEXP_{i.}, KGEXP_{i.}, CONG_{i.}, VADDED_{i.})$$

where P_{ij} is the total number of patents of province i (i=1 … 93) in industry j (j=1 …25).[3]

Given the discrete nature of the dependent variable and the high number of zeros (i.e. non innovative provinces), the most appropriate approach to estimate (3) seems to be a Poisson regression model (Hausman, Hall and Griliches, 1984).

In a nutshell, given a vector of counts Y, the Poisson regression model specifies that each y_i is drawn from a Poisson distribution with parameter λ_i, which is related to the regressors x_i. The basic equation of the model is:

(4) $$\Pr(y_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}$$

The most common formulation for λ_i is the log-linear model:

(5) $$\ln(\lambda_i) = x_i\beta$$

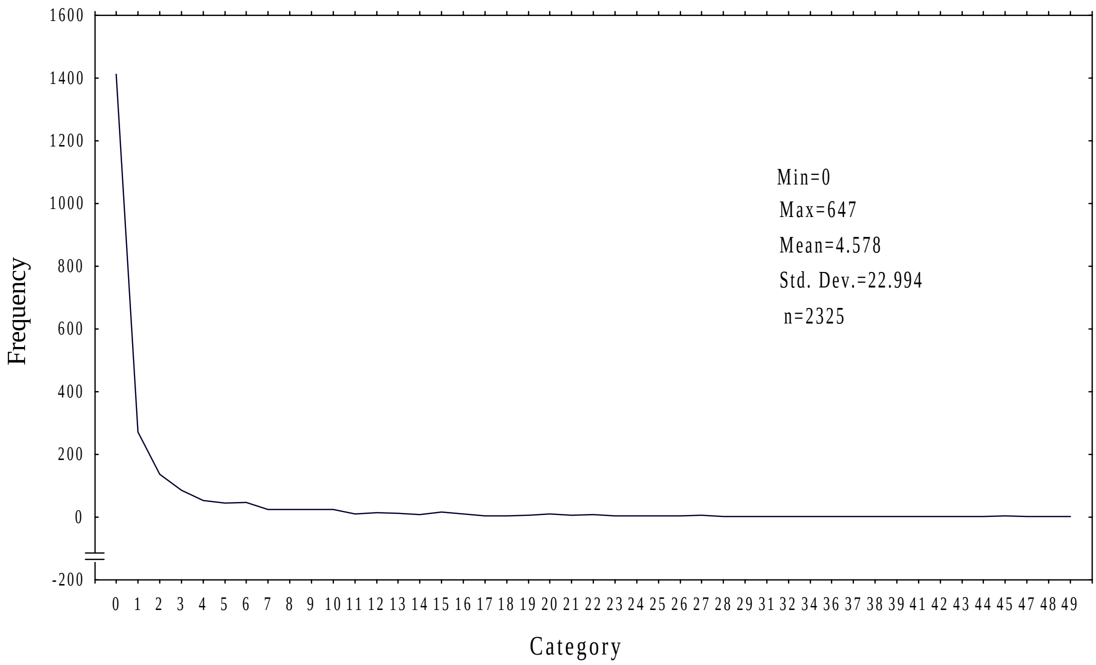
where x_i is a vector of regressors and β a vector of coefficients. From this, it follows that the expected number of events (i.e. patents) per period is given by:

$$E(y_i \mid x_i) = V(y_i \mid x_i) = \lambda_i = \exp(x_i\beta)\,. \tag{6}$$

In this paper, however, an alternative specification of the Poisson model will be used in order to allow for two interrelated features of the dependent variable: the very large number of zeros and the qualitative difference of the zero outcomes from the positive ones. Graph 1 reports the actual distribution of patent counts (for graphical purposes the distribution has been truncated at 49 patent counts). As one can see, zeros are 1412 that represent 61% of all patent counts. The distribution of patent counts is extremely skewed, with a mean of 4.6 and standard deviation of 22.9. In addition to that, it must be also pointed out that six provinces out of 93 have zero patents in all the 25 technological classes.

Graph 1 - Actual distribution of patent counts

(Italy NUTS 3, 25 sectors, 1987-94)



The number of patents can be thus reinterpreted as the observed outcome of a latent variable for “innovative capabilities”. The observed number of patents is zero as long as innovative capabilities of a given province do not exceed a certain critical threshold, while it takes any positive and integer value once the threshold has been crossed.

In settings where there are “too many” (or “too few”) zeros in the data and the zero outcome of the data generating process is qualitatively different from the positive ones, the hurdle

Poisson has been suggested as an alternative (Greene, 1997). In broad terms, the hurdle Poisson model allows for systematic differences in the statistical process governing observations with zero counts and observations with one or more counts. This is achieved by combining a dichotomous model governing the binary outcome of the count being zero or positive and a Poisson model for strictly positive outcomes (Winkelmann and Zimmermann, 1995). In what follows, the extension of the hurdle model proposed by Lambert (1992) and Greene (1994) will be used to estimate equation (3). This model, labelled Zip (Zero-Inflated Poisson) model, allows for excess zeros in count models under the assumption that the population is characterised by two regimes, one where members always have zero counts, and one where members have zero or positive counts. The likelihood of being in either regime is estimated using a logit specification, while the counts in the second regime are estimated using a Poisson specification. Thus, the Zip model specifies that (Lambert, 1992, pg. 3):

$$(7) \qquad y_i \sim 0 \qquad \text{with probability } p_i$$

$$y_i \sim \text{Poisson} \qquad \text{with probability } (1-p_i)$$

so that

$$y_i = 0 \qquad \text{with probability } p_i + (1-p_i)\exp(-\lambda_i)$$

$$y_i = k \qquad \text{with probability } (1-p_i)\exp(-\lambda_i)\lambda_i^k / k!$$

In turn, the parameters p_i and λ_i are assumed to depend on some covariates according to

$$(8) \qquad \ln(\lambda_i) = x_i\beta$$

$$\text{logit}(p_i) = \ln(p_i / 1 - p_i) = z_i\gamma$$

where x_i and z_i are vectors of regressors. Moreover, the regressors that affect the mean of the Poisson process may or may not be the same as the regressors that affect the probability of the zero regime (Lambert, 1992).

Estimates of equation (3) based upon the Zip regression model are reported in Table 6.[4] After various attempts, four regressors have been used for the logit model: OWNEMP, OTHEMP, PHSUBS, CONG and VADDED. For the Poisson model, five additional variables have been introduced: RELTEC, OTHTEC, HERF, KGEXP, RDEXP. Regional fixed effects have been also considered. Italian NUTS 3 provinces have been grouped into five broad areas: *North-West* (Piemonte, Valle d'Aosta and Liguria), *Lombardy*, *North-East* (Veneto, Trentino, Friuli V.G.), *Third-Italy* (Emilia Romagna, Toscana and Marche), *Centre* (Lazio, Umbria, Abruzzo and Molise) and *South* (Puglia, Campania, Calabria, Sicilia and Sardegna). In addition to that, for the Poisson model, dummy variables for the four largest metropolitan areas (Milan, Turin, Rome and Bologna) have been also introduced. These four provinces account for about 20% of total population, 23% of total manufacturing employment and 52% of total patenting activity. For a complex mix of historical, political and economic reasons, that our explanatory variables are not able to capture fully, these four metropolitan areas have become the location of intense innovative processes. Finally, in both models sectoral fixed effects have been explicitly considered (even though value and statistical significance of coefficients are not reported in the tables).

Looking at the first column of Table 6, the likelihood of being in the positive regime (i.e. to have a number of patents ≥ 0) is positively affected by the strength of agglomeration economies. In particular, the stronger the industrial base of a province the higher the probability of being in the positive regime. However, the positive and statistically significant coefficient of VADDED also points out the presence of important productivity effects. Provinces enjoying higher levels of value added per capita are also less likely to be in the zero regime. It is also interesting to note that the dummy variables for regions (using South as baseline) are not statistically significant, with the exception of *D_Centre*.

Looking instead at the second column of Table 6, the expected number of patented innovations increases with the strength of agglomeration economies within a province (the coefficients of OWNEMP and OTHEMP are positive and statistically significant) and with the level of knowledge spillovers (RDEXP). On the other hand, higher congestion costs (CONG) reduce the expected number of innovations.

Table 6
Hurdle Poisson regression estimates

	Logit P(0/1)	**Poisson**
RELTEC		-0.002
		(-14.841)
OTHTEC		-0.002
		(-20.852)
HERF		0.511
		(2.256)
OWNEMP	-0.137	0.364
	(-2.219)	(23.206)
OTHEMP	-0.133	0.156
	(-2.234)	(10.213)
PHSUBS	-1.351	0.460
	(-7.099)	(7.860)
RDEXP		0.317
		(6.783)
KGEXP		-0.404
		(-7.893)
CONG	-0.003	-0.006
	(-0.383)	(-2.851)
VADDED	-0.024	-0.009
	(-3.010)	(-3.854)
D_NorthWest	-0.579	0.379
	(-1.383)	(2.808)
D_Lombardy	-0.354	0.681
	(-0.741)	(4.586)
D_NorthEast	-0.573	1.213
	(-1.482)	(10.705)
D_ThirdItaly	-0.283	0.548
	(-0.811)	(4.521)
D_Centre	-1.264	0.242
	(-3.728)	(2.002)
D_Turin		3.319
		(26.202)
D_Milan		9.520
		(24.764)
D_Bologna		1.617
		(19.993)
D_Rome		1.794
		(9.026)
Constant	16.778	-5.011
	(10.330)	(-9.476)
N	2325	
Log L	-5147.79	
Predicted frequency of zeros	1410.8	

Notes: t-values among parentheses. Sectoral fixed effects are included in the regression. The predicted frequency of zeros is obtained by summing over the individual predicted probabilities for that outcome.

Somewhat less clear and more difficult to interpret are the results concerning the effect of technological diversity and of investments in innovative capital goods. Both types of variables negatively impact on the expected number of innovations produced within a province. Regarding the role of investments in capital goods (KGEXP), as it was argued above, they are likely to operate as substitutes for more formalised and autonomous innovative activities. From this perspective, the result observed indicates that firms located in regions where innovation is mainly embodied in new capital goods, are also less likely to be autonomously innovative and therefore less able to generate patented innovations.

For what concerns, instead, the effect of variables related to technological diversity, one possible interpretation of the results obtained runs as follows. The largest number of patents *and* the lowest Herfindahl indexes (i.e. the widest base of technological competencies) are recorded in the four metropolitan areas mentioned above (Milan, Turin, Rome and Bologna). After controlling for these four areas, therefore, a higher number of patents are expected to take place in provinces with relatively narrower technological base and higher Herfindahl indices. In order to assess the correctness of this interpretation, a regression model has been estimated excluding the four metropolitan areas (see Table 7). The results clearly indicate that a wider technological base is likely to encourage innovative activity in any technological sector.

Finally, it is worth noting that regional fixed effects are positive and statistically significant. In particular, provinces located in Lombardy, North-East and Third-Italy are expected to generate a significantly higher number of patented innovations compared to the provinces located in the South (and in the Centre). In other terms, while spatial location does not affect the likelihood of being in either regimes, it does impact on innovative performance of provinces being in the positive (Poisson) regime (see Table 6). However, the coefficients of regional dummy variables become statistically not significant and they even change sign once we exclude metropolitan areas from the analysis (see Table 7). This result therefore indicates that fixed regional effects mainly occur through large metropolitan agglomerations and a possible line of future research is that of modelling spatial interactions and spillovers among contiguous provinces, particularly between large urban agglomerations and surrounding areas.

Table 7

Hurdle Poisson regression estimates
(excluding metropolitan areas)

	Logit P(0/1)	**Poisson**
RELTEC		0.003
		(7.754)
OTHTEC		0.003
		(11.554)
HERF		-0.668
		(-2.852)
OWNEMP	-0.225	0.211
	(-3.601)	(10.354)
OTHEMP	-0.170	0.024
	(-2.838)	(1.269)
PHSUBS	-1.162	0.132
	(-6.001)	(2.167)
RDEXP		0.145
		(2.970)
KGEXP		-0.183
		(-3.417)
CONG	-0.005	-0.004
	(-0.667)	(-1.572)
VADDED	-0.020	-0.001
	(-2.389)	(-0.379)
D_NorthWest	-0.743	-0.169
	(-1.723)	(-1.214)
D_Lombardy	-0.596	-0.177
	(-1.209)	(-1.151)
D_NorthEast	-0.779	0.122
	(-1.955)	(1.005)
D_ThirdItaly	-0.552	-0.110
	(-1.565)	(-0.869)
D_Centre	-1.212	-0.293
	(-3.622)	(-2.428)
Constant	15.635	-0.701
	(9.363)	(-1.293)
N		2225
Log L		-4052.32
Predicted frequency of zeros		1406.4

Notes: t-values among parentheses. Sectoral fixed effects are included in the regression. The predicted frequency of zeros is obtained by summing over the individual predicted probabilities for that outcome.

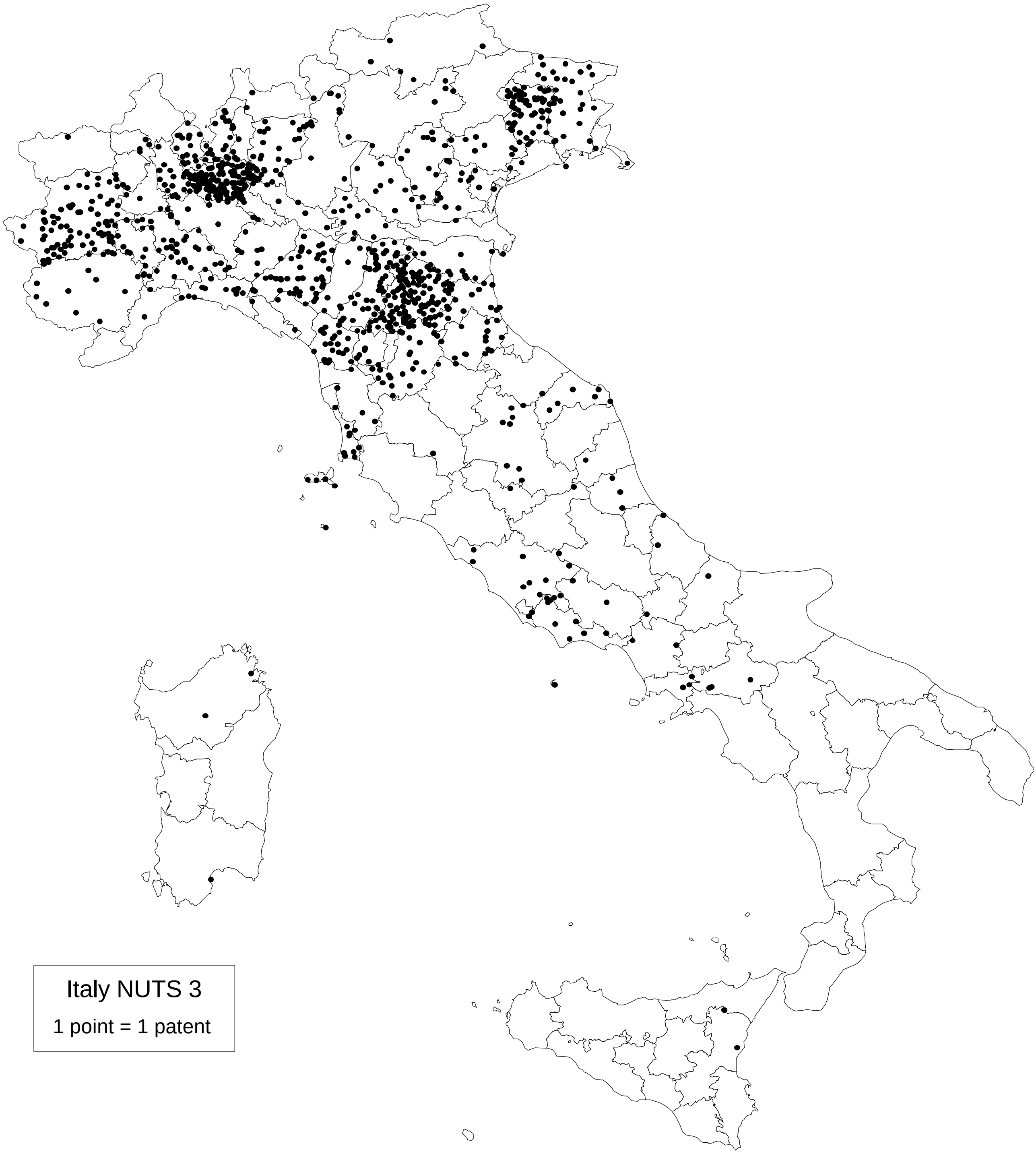
Conclusions

This paper has examined from an empirical perspective the spatial patterns of innovative activities, with reference to the Italian case. The analysis carried out shows that both innovative and manufacturing activities are spatially concentrated, even though innovations are considerably more geographically clustered than production. There are also large differences across sectors in the degree of concentration of innovation and production. Concentration is relatively lower in most mechanical engineering and industrial equipment sectors and it is relatively higher in most electrical-electronic and chemical-drugs sectors.

Beside that, spatial data analysis has shown that industrial sectors also largely differ with respect to the way innovative and manufacturing activities are arranged in space. In particular, while there is evidence of significantly positive spatial autocorrelation in several mechanical sectors, in most chemical and electronic sectors innovative and productive activities are concentrated in few technological "poles" that are surrounded by non-innovative provinces. Moreover, production displays higher levels of spatial autocorrelation than innovations. Spatial analysis has also evidenced that in mechanical industries and, to a much less extent, instruments sectors, there are several local systems of clustered innovative provinces. All these systems are, however, located in the northern part of the country. On the other hand, in most chemical and electronic industries innovative activities tend to concentrate in few metropolitan areas surrounded by non-innovative provinces.

Results of regression estimates provide support to the hypothesis that the strength of agglomeration economies and localised knowledge spillovers has a positive impact on the innovative performance of provinces. Innovative performance is also positively associated with a wide and diversified base of local technological competencies, whereas congestion costs have a negative impact on the number of patented innovations. Finally, higher investments in innovative capital goods result in a lower number of innovations, by operating as substitutes for more formalised and autonomous innovative activities.

Map 1
Spatial distribution of patents: Handling technologies (Italy NUTS 3 – 1987-94)



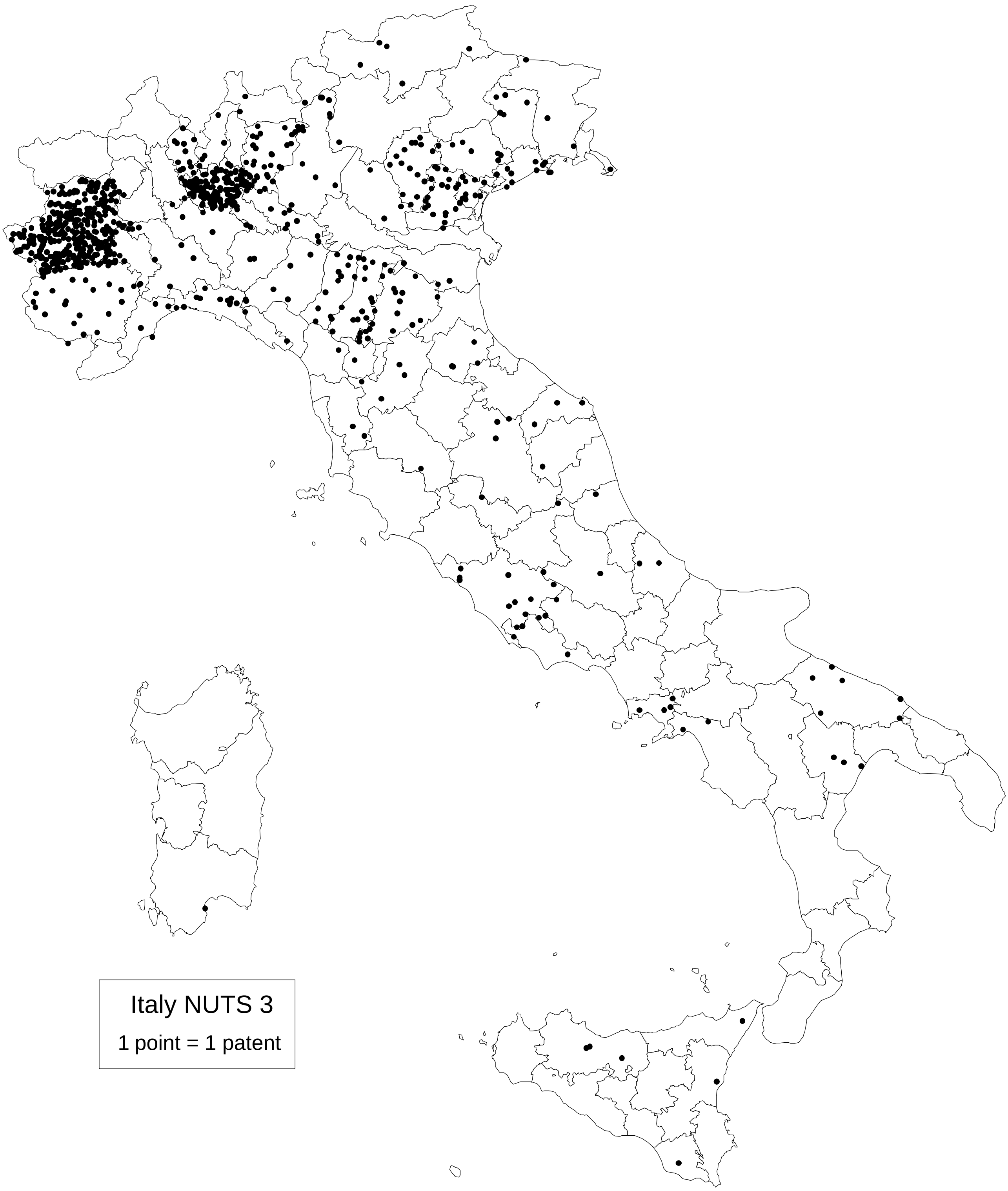
Map 2
Spatial distribution of patents: Control technology (Italy NUTS 3 – 1987-94)

Map 3
Spatial distribution of patents: Drugs (Italy NUTS 3 – 1987-94)

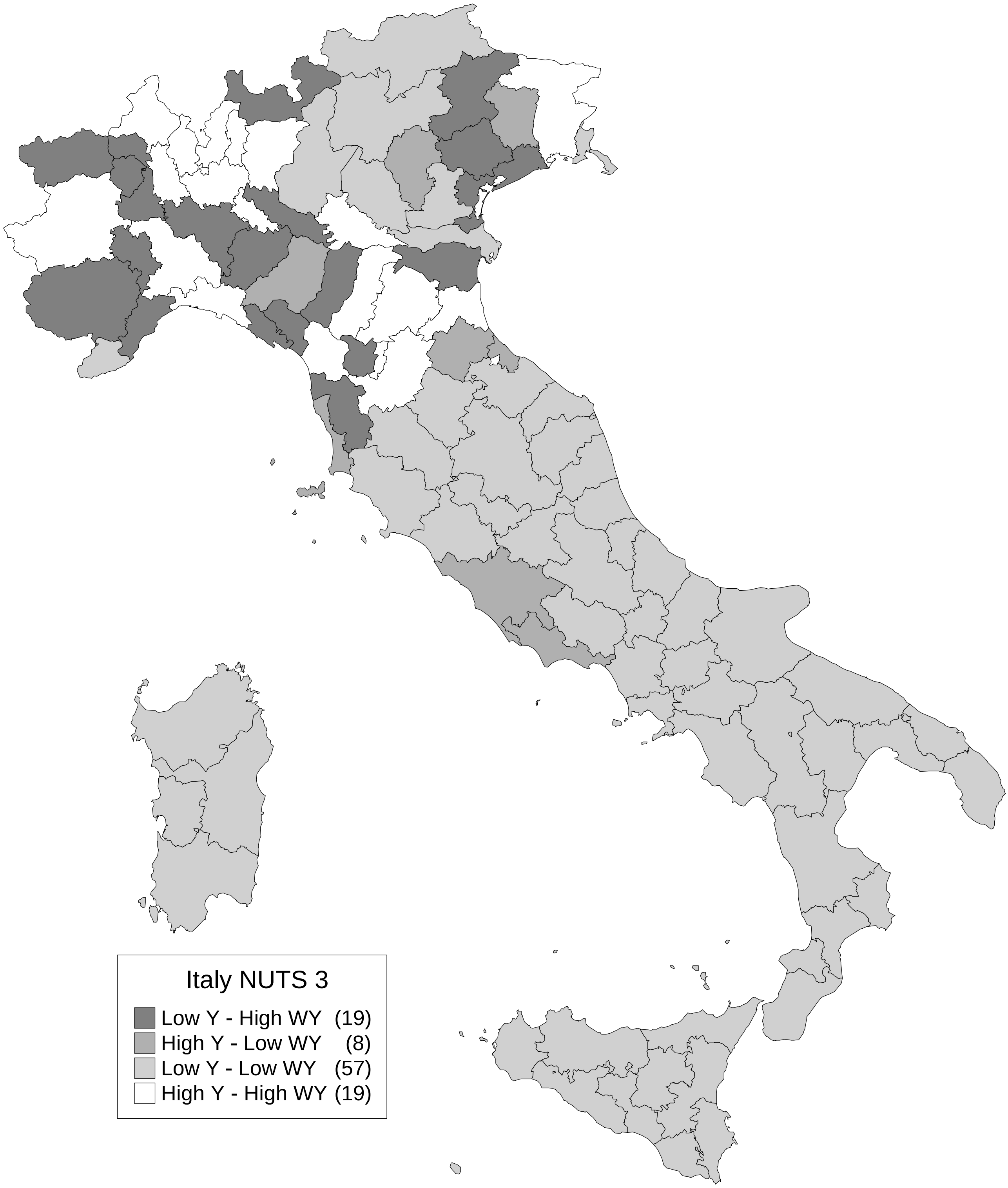
Map 4

Spatial distribution of patents: Telecommunications (Italy NUTS 3 – 1987-94)

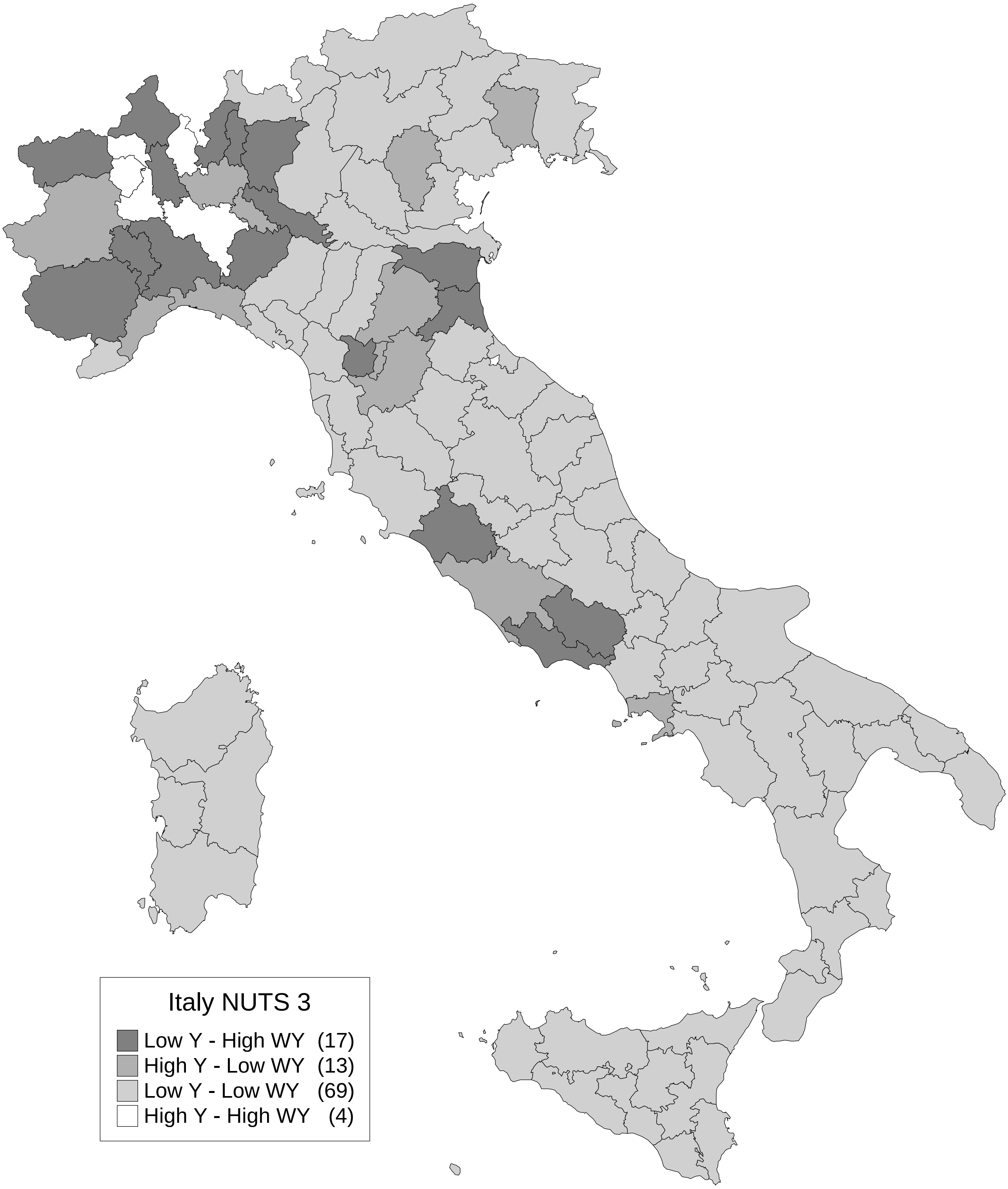
Map 5
Spatial distribution of patents: Transports (Italy NUTS 3 – 1987-94)



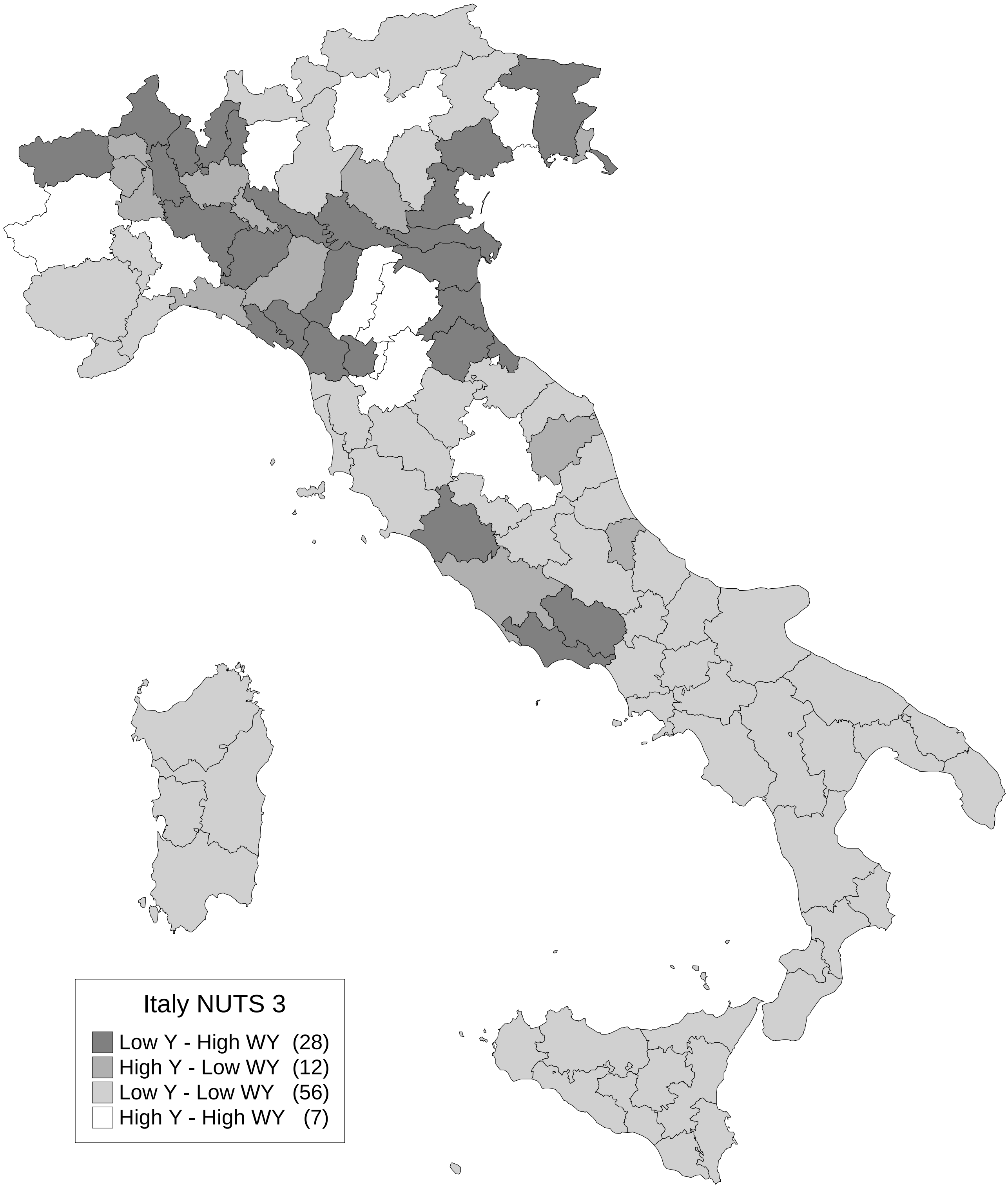
Map 6
Moran scatterplot: Handling technologies (Italy NUTS 3 – 1987-94)



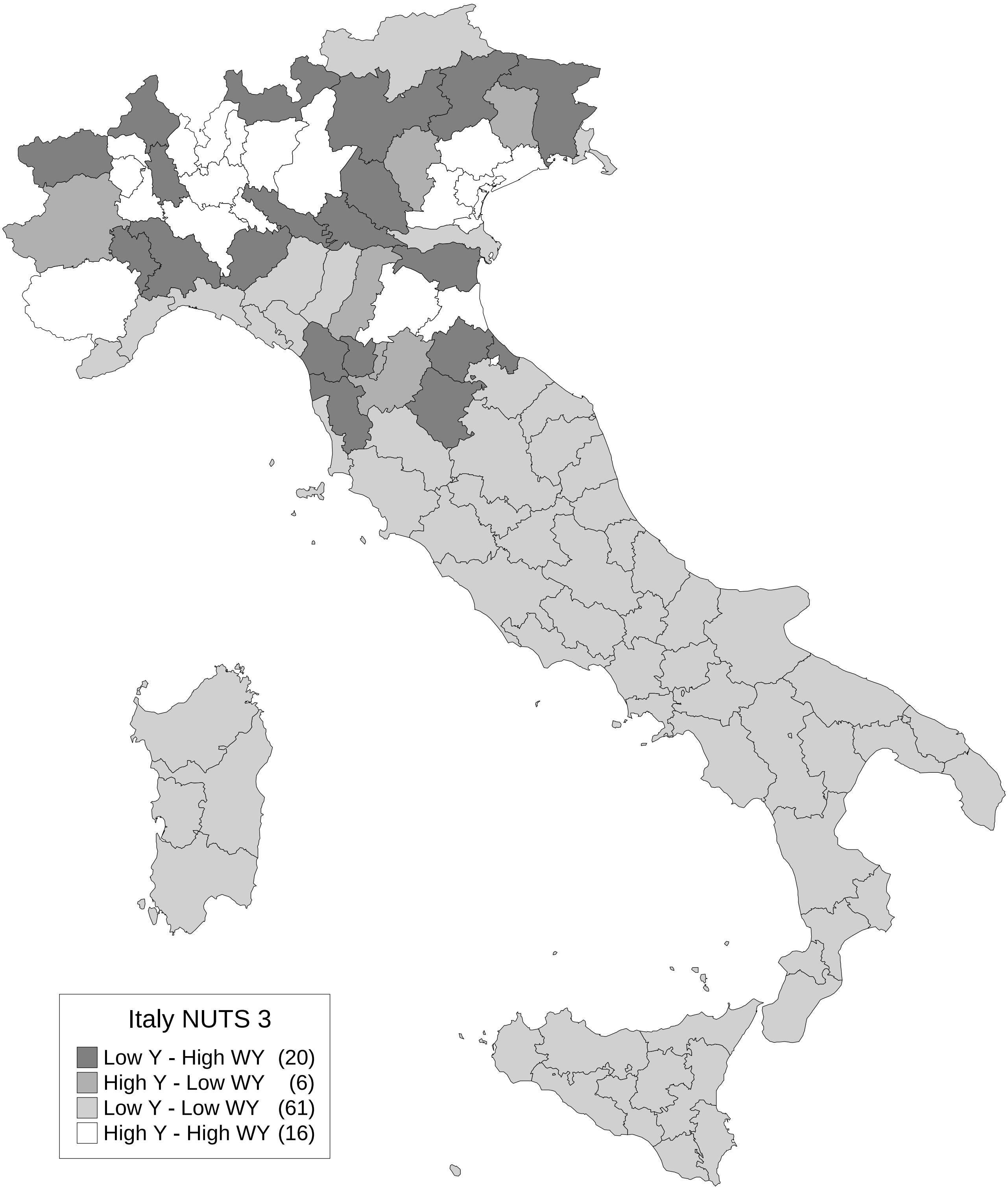
Map 7
Moran scatterplot: Control technology (Italy NUTS 3 – 1987-94)



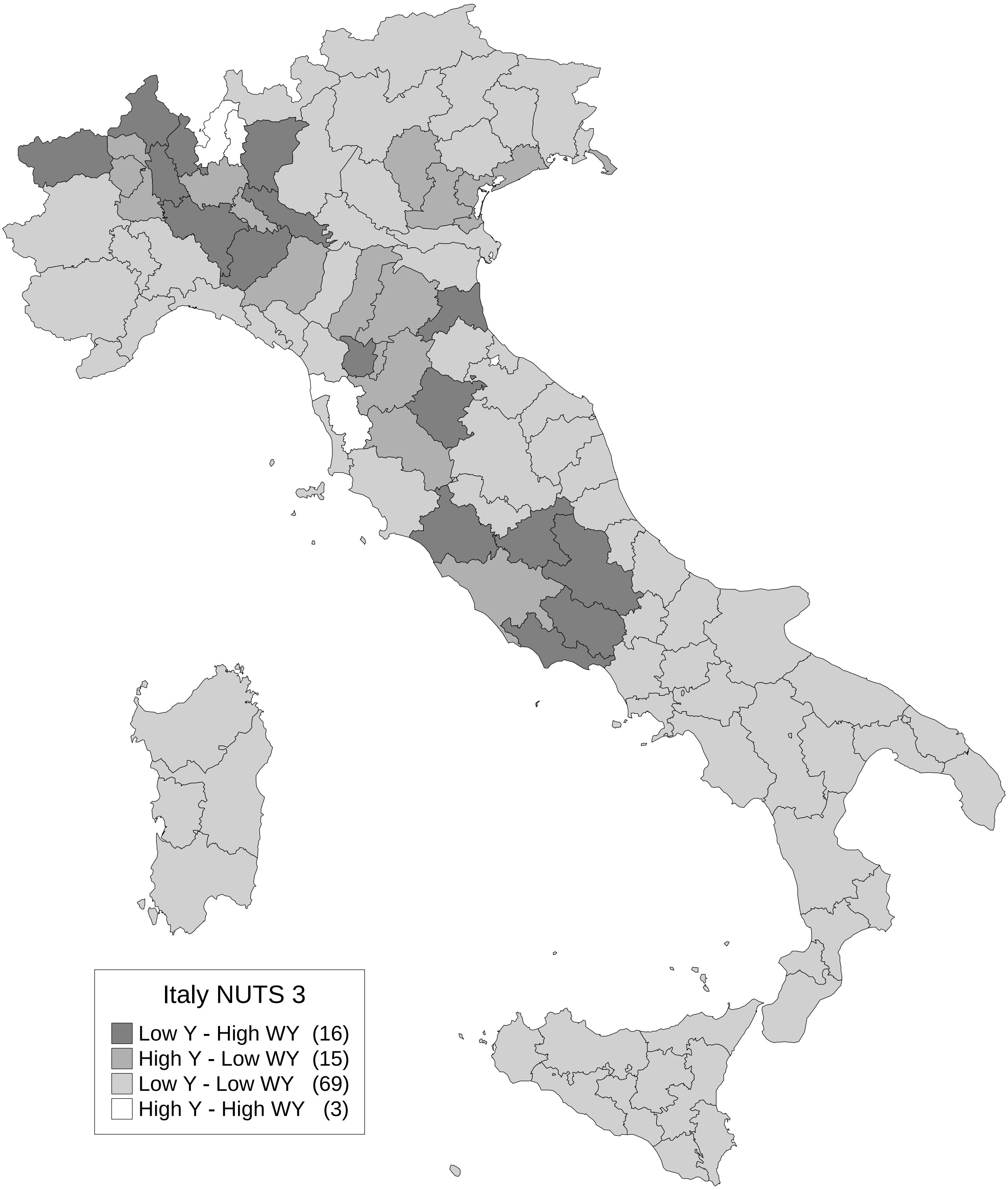
Map 8
Moran scatterplot: Medical technology (Italy NUTS 3 – 1987-94)



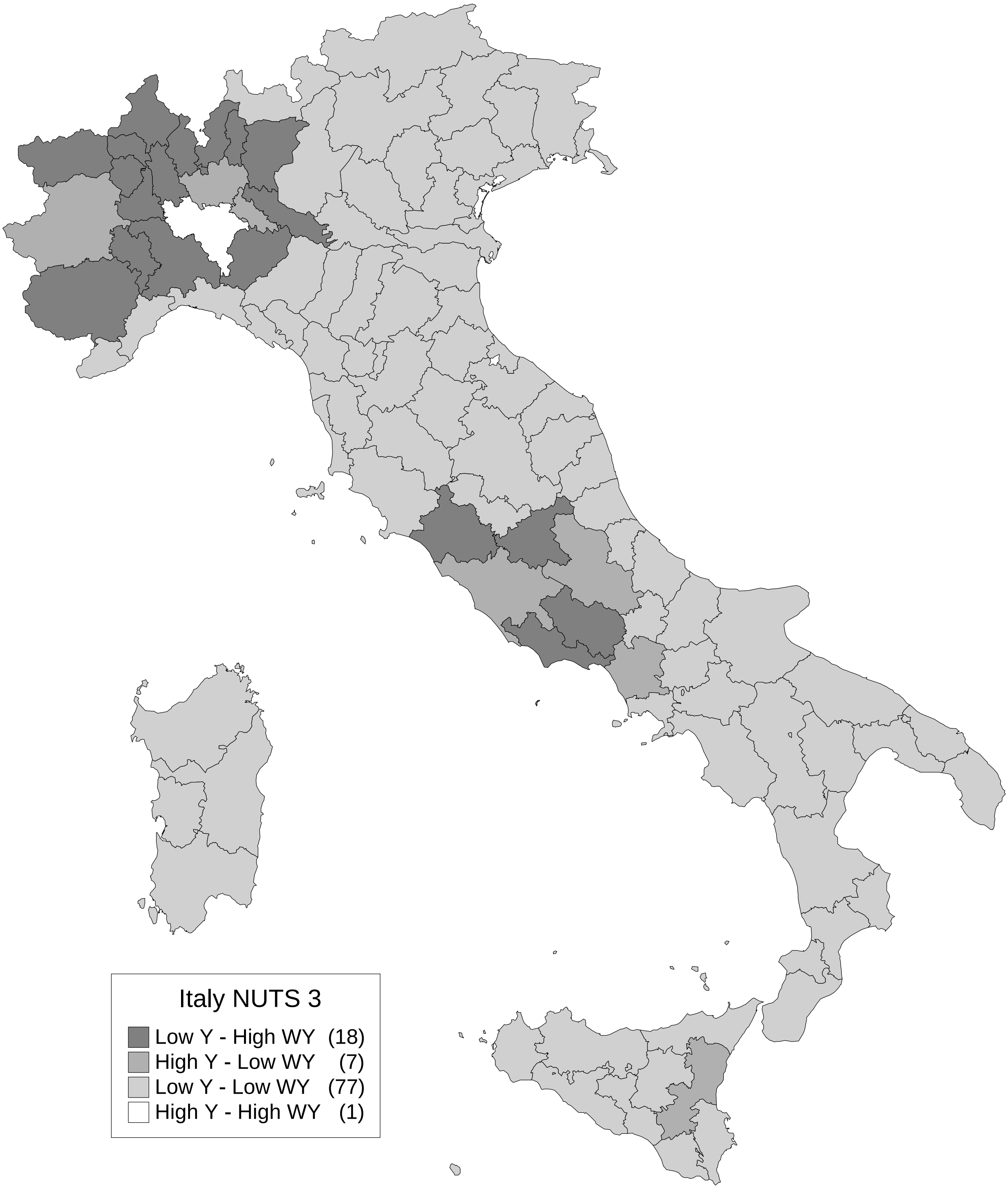
Map 9
Moran scatterplot: Materials processing (Italy NUTS 3 – 1987-94)



Map 10
Moran scatterplot: Drugs (Italy NUTS 3 – 1987-94)



Map 11
Moran scatterplot: Telecommunications (Italy NUTS 3 – 1987-94)



Appendix 1 – Concordance table between IPC technology codes and NACE classification

Industry-Technology	IPC code	NACE code
I. Electrical engineering		
1. Electrical machinery and apparatus, electrical energy	F21; G05F; H01B,C,F,G,H,J,K,M, R,T; H02; H05B,C,F,K	31
2. Audio-visual technology	G09F,G; G11B; H03F,G,J; H04N-003,-005, 009,-013,-015, -017,R,S	32.2 + 32.3
3. Telecommunications	G08C; H01P,Q; H03B,C,D,H,K,L,M; H04B,H,J,K,L,M, N-001, -007,-011,Q	32
4. Information technology	G06; G11C; G10L	30
5. Semiconductors	H01L	32.1
II. Instruments		
6. Optics	G02; G03B,C,D,F,G,H; H01S	33.4
7. Analysis, measurement, control technology	G01B,C,D,F,G,H,J,K,L,M,N, P,R,S,V, W; G04; G05B,D; G07; G08B,G; G09B,C,D; G12	33.2 + 33.5
8. Medical technology	A61B,C,D,F,G,H,J,L,M,N	33.1
III. Chemistry, pharmaceuticals		
9. Organic fine chemistry	C07C,D,F,H,J,K	24.1
10. Macromolecular chemistry, polymers	C08B,F,G,H,K,L; C09D,J;C13L	24.1
11. Pharmaceuticals, cosmetics	A61K	24.4
12. Biotechnology	C07G; C12M,N,P,Q,R,S	-
13. Materials, metallurgy	C01; C03C; C04; C21; C22; B22	26 + 27
14. Agriculture, food chemistry	A01H; A21D; A23B,C,D,F,G,J,K,L; C12C,F,G,H,J; C13D,F,J,K	15
15. Chemical and petrol industry, basic materials chemistry	A01N; C05; C07B; C08C; C09B,C,F, G,H,K; C10B,C,F, G,H,J,K,L,M; C11B,C,D	23.1 + 23.2 + .24.1 + 24.2
IV. Process engineering, special equipment		
16. Chemical engineering	B01B,D (without -046 to -053), F,J,L;B02C; B03; B04; B05B; B06; B07; B08; F25J; F26	28.3 + 29.2 + 29.5
17. Surface technology, coating	B05C,D; B32; C23; C25; C30	28.5 + 29.5
18. Materials processing, textiles, paper	A41H; A43D; A46D; B28; B29; B31; C03B; C08J; C14; D01; D02; D03; D04B,C,G,H; D05; D06B,C,G,H,J,L,M,P,Q; D21	29.5
19. Thermal processes and apparatus	F22; F23B,C,D,H,K,L,M,N,Q; F24; F25B,C; F27; F28	28.2 + 28.3 + 29.2
20. Environmental technology	A62D; B01D-046 to -053; B09; C02; F01N; F23G,J	-
V. Mechanical engineering, machinery		
21. Machine tools	B21; B23; B24; B26D,F; B27; B30	29.4
22. Engines, pumpes, turbines	F01B,C,D,K,L,M,P; F02; F03; F04; F23R	29.1
23. Mechanical elements	F15; F16; F17; G05G	28.5 + 29.1
24. Handling, printing	B25J; B41; B65B,C,D,F,G,H; B66; B67	29.2
25. Agricultural and food processing, machinery and apparatus	A01B,C,D,F,G,J,K,L,M; A21B,C; A22; A23N,P; B02B; C12L; C13C,G,H	29.3 + 29.5
26. Transport	B60; B61; B62; B63B,C,H,J; B64B,C,D,F	34.1 + 35 (exc. 35.3)
27. Nuclear engineering	G01T; G21; H05G,H	-
28. Space technology, weapons	B63G; B64G; C06; F41; F42	-
29. Consumer goods and equipment	A24; A41B,C,D,F,G; A42; A43B, C; A44; A45; A46B; A47; A62B,C; A63; B25B,C,D,F,G,H; B26B; B [illegible] B43; B44; B68; D04D; D06F,N; D07; F25D; G10B,C,D,F,G,H,K	-
30.Civil engineering, building, mining	E01;E02;E03;E04; E05;E06;E21	29.5

Appendix 2 - Italian NUTS 3 provinces

References

ANFIA, 1994, *L'automobile in cifre*, Codex, Torino.

Anselin L., Varga A. and Acs Z., 1997, *Entrepreneurship, geographic spillovers and University research: A spatial econometric approach*, ESRC WP 59, University of Cambridge.

Anselin, L., 1998, "Spatial dependence in linear regression models with an introduction to spatial econometrics", in A. Ullah and D. Giles (eds.) *Handbook of Applied Ecomomic Statistics*, Marcel Dekker, New york.

Audretsch D.B. e Feldman M., 1996, "Knowledge spillovers and the geography of innovation and production", *American Economic Review*, 83, pp. 630-640.

Breschi S., 1998, "Spatial patterns of innovation: evidence from patent data", in Malerba F. and Gambardella A. (eds.) *The Organisation of Innovative Activity in Europe*, Cambridge University Press, Cambridge (forthcoming).

Breschi, S., Lissoni, F. and Malerba, F., 1998, "Knowledge proximity and firms' technological diversification", ISE TSER paper.

Cliff, A.D. and Ord, J.K., 1981, *Spatial Processes. Models and Applications*, Pion, London.

Feldman M., 1994, *The Geography of Innovation*, Kluwer Academic, Boston.

Greene W.H., 1997, *Econometric Analysis*, Prentice Hall International, London.

Griliches Z., 1991, "Patent statistics as economic indicators: a survey", *Journal of Economic Literature* 28, 1661-1707.

Haining, R., 1990, *Spatial data analysis in the social and environmental sciences*, Cambridge University Press, Cambridge.

Hausman, J., Hall B.H. and Griliches Z., 1984, "Econometric models for count data with an application to the patents-R&D relationship", *Econometrica*, 52, 909-938.

ISTAT, 1998, *Statistiche sulla ricerca scientifica e tecnologica*, Roma.

ISTAT, 1993, *Popolazione e movimento anagrafico dei comuni*, Roma.

Jaffe A.B., 1989, "Real Effects of Academic Research", *American Economic Review*, vol. 79, n. 5, pp. 957-70.

Jaffe A.B., Trajtenberg M., Henderson R., 1992, *Geographic localization of knowledge spillovers as evidenced by patent citations*, Nber, W.P. 3993, Cambridge Mass..

Krugman P., 1991, *Geography and Trade*, Mit Press and Leuven University Press, pp. 142.

Lambert, D., 1992, "Zero-Inflated Poisson regression with an application to defects in manufacturing", *Technometrics*, 34 (1), 1-14.

Mansfield E., 1995, "Academic research underlying industrial innovations: sources, characteristics and financing", *Review of Economics and Statistics*, 77, pp. 55-65.

Pred A., 1966, *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800-1914. Interpretative and Theoretical Essays*, Mit Press, Cambridge Mass., 1966.

SIP, 1990, *Struttura socio-demografica e telefonica dei comuni italiani*, Roma.

Storper M., 1992, "The limits to globalisation: Technology districts and international trade", *Economic Geography* 68, 1, 60-93.

Swann, G.M.P., 1996, “Technology evolution and the rise and fall of industrial clusters”, *Revue Internationale de Systemique*, 10 (3), pp. 285-302.

Swann, G.M.P. e M. Prevezer, 1996, “A comparison of the dynamics of industrial clustering in computing and biotechnology ”, *Research Policy*, 25, pp. 1139-1157.

Winkelmann R. and Zimmermann K.F., 1995, “Recent developments in count data modelling: Theory and application”, *Journal of Economic Surveys*, 9(1), 1-24.

WIPO (1994) *International Patent Classification, 6th edition (1994)*, volume 9 Guide, survey of classes and summary of main groups, Geneva.

Notes

[1] SMSA stands for Standard Statistical Metropolitan Area.

[2] The ISTAT Innovation survey has been based upon a questionnaire sent to more than 22,000 firms. It must be pointed out that the survey could be extremely useful in order to measure other characteristics of clusters conducive to (or hampering) innovative activities. However, published data are only reported at the NUTS 2 level.

[3] Two provinces, Aosta and Trieste, have been dropped from the analysis due to their small area compared to the other provinces and their peculiar geographical location. Concerning sectors, 5 industries have been dropped (biotechnology, environmental technology, nuclear engineering, space technology, and consumer goods) since it could not be found a satisfactory concordance between IPC codes and the NACE classification.

[4] STATA 5.0 has been used to estimate the model.