Spatial location and economic development on the Spanish mainland
J.M. Albert(1), J. Mateu(2), V. Orts(1)
(1) Department of Economics, (2) Department of Mathematics,
Universitat Jaume I, Castellón, Spain. EMail: albert@eco.uji.es

Abstract

A bi-parametric model (auto-Poisson) is fitted on the spatial point model where points represent economic magnitudes in different Spanish geographical areas. We use spatial statistical techniques. This way we obtain as a result, the value of two parameters which present a strong disparity. This indicates very unequal development potentials among these areas due to their own location.

1 Introduction

The literature on "market potential" begun by Harris [7]. This literature argues that the desirability of a region as a production site depends on its access to markets, and that the quality of that access may be described by an index of "market potential" which is a weighted sum of the purchasing power of all regions, with the weights depending inversely on distance. Thus if $Y_k$ is the income of region $k$, and $D_{jk}$ is the distance between $j$ and $k$, then the market potential of region $j$ would be determined by an index of the form

$$ M_j = \sum_k Y_k g(D_{jk}) $$

where $g(.)$ is some declining function (Krugman [10]).

Indices of market potential do seem to have quite a lot of power in explaining the location of industries across the United States or Western Europe, and the location of particular activities within urban areas. In the 1950s Harris [7] and others drew striking maps of market potential surfaces for the United States. More recently, market potential studies for the European Commission have led to similar maps that show a clear relationship between "centrality" and income per capita (Krugman [11]). The concept of market potential is approximated through a points cloud. Each point is the location of a
given purchasing power. The distance between the points is the distance between different locations, i.e., we use a spatial point pattern.

A spatial point pattern is a data set \( \{(x_i, y_i) \mid i = 1, \ldots, n\} \) consisting of \( n \) locations in an essentially planar region. Some examples are the locations of cell nuclei in a microscopic tissue section, trees in a forest, or cases of disease in a geographical region. An assumption that is fundamental for the analysis of such a data is that they can usefully be regarded as a partial realisation of a stochastic point process (Cox and Isham [2]). Many systems of individuals can also be described by attaching measurable quantities \( m_i \) to the locations, like, for instance, diameter of a tree. This last case is an example of marked points patterns.

The reason for the interest in a pattern of points is that, with an appropriate choice of scale, even huge objects may be better represented by a point. Given suitable scales, the actual sizes of objects that may be represented this way are unbounded. At one end microscopes are required, and at the other end, telescopes are needed. Spatial statistical techniques will be discussed in Section 2. The data set is analysed in section 3. The fourth section presents the results ending with some final conclusions in section 5.

2 Methodology

A formal definition of complete spatial randomness (CSR) is that the events in a region \( A \) constitute a partial realisation of a homogeneous, planar Poisson process (Diggle [4]). This process incorporates a single parameter, \( \lambda \), the intensity, or mean number of events per unit area. The actual number of events in \( A, n \) say, is an observation from a Poisson distribution with mean \( \lambda |A| \), where \( |A| \) denotes the area of the region \( A \).

If we considerer \( n \) as fixed, we arrive at the following definition of CSR: (1) each of the \( n \) events is equally likely to occur at any point within \( A \); (2) the \( n \) events are located independently of each other. Our interest in CSR is that it represents an idealized standard which, if strictly unattainable in practice, may nevertheless be tenable as a convenient first approximation. Most analysis begin with a test of CSR, and there are several good reasons for this. Firstly, a pattern for which CSR is not rejected scarcely merits any further formal statistical analysis. Secondly, tests are used as a means of exploring a set of data, rather than because rejection of CSR is of intrinsic interest. Thirdly, CSR acts as a dividing hypothesis to distinguish between patterns which are broadly classifiable as "regular" or "aggregated."

2.1 Quadrats methods

Throughout this section we shall start working with an observed pattern and attempt to deduce the nature of the process that gave rise to that pattern. Quadrat sampling involves collecting counts of events in subsets of the study
region. Traditionally, these subsets are rectangular (hence the name of quadrats), although any shape is possible. Quadrats may be placed either randomly or layed out contiguously in the region.

This is very easy to implement scattered quadrats, as you only need to position quadrats randomly in the study region (Fig.1) and count the numbers of events that fall in each quadrat. It is true that scattered quadrats counts provide some limited information about the nature of a point pattern and that is our present concern.

We have already remarked that the counts from equi-sized quadrats scattered over a Poisson pattern will be observations from a Poisson distribution with parameter equal to the product of the intensity of the events per unit area and the area of the quadrats. A useful characteristic of the Poisson distribution is that its parameter is equal both to the mean and the variance of the distribution. If the pattern is more regular than a Poisson one, then the quadrats counts will be more uniform in size and will therefore have a relatively small variance (when compared with the size of the mean). On the other hand, if there are clusters, then some quadrats will have large counts so that the variance of the counts will be relatively large. A natural test of a Poisson distribution is therefore provided by examining the value of the ratio sample variance/sample mean. If we denote the observed counts in n quadrats by \( x_1, x_2, \ldots, x_n \), then these counts have mean \( \bar{x} = \sum x_i / n \) and variance

\[
s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)}
\]

where the summations are over the values of \( i \) from 1 to \( n \). Hoel \[8\] showed that

\[
ID = \frac{(n-1)s^2}{\bar{x}},
\]

called the index of dispersion, has an approximate \( \chi_{n-1}^2 \) distribution under CSR. This approximation is reasonable provided \( n>6 \) and \( \bar{x} > 1 \). Values of \( ID > \chi_{n-1}^2; 1 - \alpha \) are indicative of clustering and regularity is given by values of \( ID < \chi_{n-1}^2; \alpha \). Perry and Mead \[13\] have examined the behaviour of the index of dispersion test and conclude that it is remarkably sensitive at detecting a lack of homogeneity in a point pattern.

A number of other indices have been suggested, principally for situations where it is thought that clusters may be present. All are sensitive to changes in quadrat sizes. Perry and Mead \[13\] concluded that, in case of ID, the behaviour of the test was principally dependent upon the size of the mean of the quadrat counts: the larger the mean the more likely was the ID test to recognize any heterogeneity that was present.

David and Moore \[3\] suggested that the quantity

\[
ICS = (s^2 / \bar{x}) - 1
\]

would provide an approximate index of clumping or contagiousness. The notation ICS was suggested by Douglas \[6\] as mnemonic for the index of
cluster size. For a Poisson pattern ICS has mean 0 and is independent of the quadrat size. An interpretation of a positive value for ICS is as the number of other events intimately associated with a randomly chosen event. A negative value for ICS indicates some regularity in the positioning of the events.

Assuming ICS measures cluster size, then the index of cluster frequency ICF = $\bar{x} / ICS$ (Douglas [6]) should measure the average number of clusters per quadrat. Lloyd's [12] index of mean crowding, $IMC = \bar{x} + ICS$, indicates the average number of events sharing a quadrat with an arbitrary event and is obviously dependent on quadrat size.

The analysis of grids of contiguous quadrats takes advantage of information on quadrat locations. A grid of contiguous quadrats is a spatial lattice (see Fig. 2). The advantage of such a grid is that neighbouring quadrats can be combined so that we may obtain information about quadrats of more than one size.

Diggle [4] proposed the same index of dispersion (ID) as in the scattered case, but now partitioning the study region $A$ into a regular kxk grid of contiguous square sub-regions of equal area, and then calculating the index ID to test CSR. Particularly, significantly small values of ID indicate a tendency towards a regular spatial distribution of events in $A$, whereas significantly large values indicate aggregation.

Obviously, we can use the other indices (ICS, IDF, and ICM) in contiguous quadrats.

2.2 Fitting a spatial model: the auto-Poisson model

In this section we suggest an objective method of estimation based on the auto-Poisson distribution (Besag [1]). Let $S$ be a finite region of the plane. We construct a point process on $S$ in the following manner. First, we overlay $S$ with an infinite grid of (small) square cells, each of area $A$ (Fig. 2) and each identified by integer cartesian co-ordinates $i=(r,s)$ with respect to a convenient origin. Let $A_i$ be both the intersection of $S$ with cell $i$ and the corresponding area. Next, we assign an auto-Poisson distribution of counts of points over all cells for which $A_i > 0$; specifically, we demand that the conditional probability of observing $x_i$ points in cell $i$, given all other counts, is

$$P(x_i) = \frac{-\mu_i(x_i)}{\mu_i(x_i) \mu_i(x_i)} x_i \times \mu_i(x_i), \quad x_i = 0, 1, ..., \lambda_i \rho_i$$

where

$$\mu_i(.) = A_i \lambda \rho_i$$
Figure 1: Scattered quadrats

Figure 2: Contiguous quadrats
Ecosystems and Sustainable Development

\[ t_i = x_{r-1,s} + x_{r+1,s} + x_{r,s-1} + x_{r,s+1} \]

and in a second-order scheme

\[ t_i = t_i(\text{first-order}) + x_{r-1,s-1} + x_{r+1,s+1} + x_{r-1,s+1} + x_{r+1,s-1} \]

For \( \lambda > 0 \) and \( 0 < \rho \leq 1 \), this produces a valid and unique probability distribution over the cells (Besag [1]; \( \rho = 0 \) requires a trivial relaxation of the positivity condition). It follows that \( \rho > 1 \) is indicative of a cluster process although defines a non-integrable probability function. The odds of a realization \( x \), in the minimal sample space, to a realization of zeros, is given by

\[
P(x) / P(0) = \lambda \sum_i x_i \sum_j x_j \prod (A_i x_i / x_i !)
\]

where the double summation is over all pairs of cells within the range defined by a first or second-order scheme. Finally, for a given realization of this scheme, we generate a point pattern \( \pi \) by distributing the points in \( A_i \) uniformly and independently. It follows that the likelihood ratio \( I_A \) of the density of such a process, with respect to the unit Poisson process on \( S \), satisfies,

\[
I_A(\pi) / I_A(\emptyset) = \lambda n(\pi) \sum x_i x_j
\]

where \( \emptyset \) denotes the empty realization and \( n(\pi) \) is the number of points in \( \pi \). By letting \( A \to 0 \), we generate a sequence of processes for which \( I_A(\pi) \to I(\pi) \), almost surely; that is, a sequence which converges to the corresponding Strauss process.

From the above result, we obtain a simple procedure for the estimation of \( \lambda \) and \( \rho \). We place a fine grid over the pattern, enumerate the corresponding cell counts and fit the auto-Poisson scheme. In principle, maximum likelihood estimation of the parameters can be invoked. However, the joint distribution involves an extremely awkward normalizing constant and this precludes the implementation of maximum likelihood. As one alternative, the pseudo-likelihood estimation method may be used. Quite generally, given the conditional probability specification of a spatial system, the method ascribes to the parameters those values that maximize the pseudo-likelihood (PL) function

\[
PL = \prod_i P(x_i | x_j \neq i)
\]

In practice, \( A = \text{constant for all cells} \) and we maximize the log(PL) using the following expression

\[
\log(PL) = -\lambda A \sum_{i=1}^n t_i + (\sum_{i=1}^n x_i) \log(A) + (\sum_{i=1}^n t_i x_i) \log \lambda + (\sum_{i=1}^n t_i x_i) \log \rho - \log(\prod_{i=1}^n x_i !)
\]
where $\lambda$ are indicative of the process intensity and $\rho$ shows (when $>1$) an evidence of clustering.

### 3 Data

The data set are the coordinates of points. Each point represents the location of 20000 inhabitants crowd with 769021 pts (pts of 1991) of income per capita. Each crowd is defined by its coordinates $(x,y)$ in the plane being $x$ longitude (eastings) and $y$ latitude (northings). The data set was transformed adequately to have planar coordinates. We only have taken into account Spanish mainland towns over 20000 inhabitants (in 1991). We suppose an area of $(0.5 \times 0.5)$ measurement units is used for every crowd of inhabitants. Each measurement unit be equal to 1500 metres (approximately). We assign $p_{i2}$ points to each town $i$ as follows

$$p_{i2} = p_{i1} \times R_i;$$

$$R_i = \frac{\text{income per capita in } i}{\text{smallest income per capita}};$$

$$p_{i1} = \frac{h_i}{20000}; \quad (h_i = \text{inhabitants in } i)$$

The number of inhabitants in each Spanish town was taken out from INE [9], and provincial income per capita data and coordinates from BBV [14] and Atlas Nacional de España [5], respectively.

A first analysis of the data shows that out of 15 peninsular autonomous communities, only four presented more than 200 points, which is the necessary minimum for a reasonable utilization of the proposed model. This means that, with our data, only four communities have income enough to treat them individually. They are the following: Madrid, Cataluña, Comunidad Valenciana and Andalucia. At provincial level, there are only two cases which surpass the cited minimum: Madrid and Barcelona.

To analyze more peninsular landscape we have joint territories with different autonomous administrations as follows: País Vasco, Navarra, The Rioja, Cantabria and the province of Burgos have been joint together; Galicia, Asturias and Castilla-León (except the provinces of Burgos and Soria) have also been joint together. The rest of peninsular autonomous communities all together (Aragon, Castilla-La Mancha, Extremadura and Murcia) do not even reach the minimum.

### 4 Results

To analyze the characteristics of market potential inside the different Spanish peninsular autonomous communities (from now on, market potential), for all
cases, we have used a 50x50 grid of contiguous quadrats. Based on this grid, we have obtained the statistics ID, ICS, ICF and IMC. We have also fitted an auto-Poisson model to obtain the intensity ($\lambda$) and cluster ($\rho$) parameters values of the income. The bigger these two parameters values are, the greater will be the market potential for we will have more income and it will be more grouped. The results are shown in table 1. Analysing table 1 we obtain the following results:

<table>
<thead>
<tr>
<th>Table 1: Results obtained for the different studied regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>area</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Madrid</td>
</tr>
<tr>
<td>Barcelona</td>
</tr>
<tr>
<td>Cataluña</td>
</tr>
<tr>
<td>Comunidad Valenciana</td>
</tr>
<tr>
<td>Andalucia</td>
</tr>
<tr>
<td>Galicia (and others)</td>
</tr>
<tr>
<td>Pais Vasco (and others)</td>
</tr>
</tbody>
</table>

a) The provinces of Madrid and Barcelona present at the same time, the larger income intensity and very grouped income. Therefore we are in the presence of high indices of market potential.

b) Concerning the autonomous communities (excepting Madrid, the one which is at the same time province and autonomous community), an important difference exists between the income intensity in Cataluña and the Comunidad Valenciana and the remaining ones. Cataluña has more income intensity than the Comunidad Valenciana, but this community presents a cluster level significantly greater. Doing a multiplicative integration of both parameters, we obtain as a result that market potential inside Cataluña is greater than in Comunidad Valenciana.

All three remaining communities present a strong cluster level, but the income intensity allow us a clear classification (from greater to smaller) in their market potential as follows: Pais Vasco, Andalucia and , finally, Galicia.
5 Conclusions

The spatial point model and the spatial statistical techniques described in this paper, have been useful instruments to determine the market potential inside different geographical zones on Spanish mainland. These techniques have allowed us to obtain, for each studied territorial unit, its purchasing power (given by the income intensity $\equiv \lambda$) and the crowd of such purchasing power (income cluster $\equiv \rho$). Assuming that the greater the market potential inside a region the greater its probability of attracting industries will be, the positive relationship between the market potential and the possibility of economic development will lead us to the following conclusions:

a) The possibilities of economic development of different territorial units on the Spanish mainland presents a great disparity due to their own location.

b) We can order the territorial units according to its possibilities of economic development (from greater to smaller) as follows: area of Madrid, eastern Mediterranean coast, Pais Vasco-Valley of the Ebro, Andalucia (overcoat the Valley of the Guadalquivir), and far away from the previous ones are Galicia-Asturias-Castilla-León. The possibilities of economic development of the remaining territorial units are worse (Extremadura, Castilla-La Mancha and Aragón).

c) The results of this work are affected by the original data (towns with more than 20000 inhabit.). It would be very interesting to make it again but with data of towns that surpass, for example, 5000 inhabitants.

d) Although it is interesting to obtain a synthetic index of the market potential using only $\lambda$ and $\rho$. We guess that the multiplicative integration of both parameters might be a useful option.

References


134 Ecosystems and Sustainable Development


