Specialized Agglomerations with Lattice Data: Model and Detection

Christian HAEDO^a and Michel MOUCHART^b

^a *Centro Interuniversitario de Investigaciones sobre Desarrollo Económico, Territorio e Instituciones (CIDETI), Alma Mater Studiorum-Università di Bologna in the Argentine Republic*

^b *Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA) and Center for Operations Research and Econometrics (CORE), UCLouvain, Belgium*

December 13, 2014

Abstract

This paper develops new statistical and computational methods for the automatic detection of spatial clusters displaying an over- or under- relative specialization spatial pattern. A probability model is used to provide a basis for a space partition into clusters representing homogenous portions of space as far as the probability of locating a primary unit is concerned. A cluster made of contiguous regions is called an agglomeration. A greedy algorithm detects specialized agglomerations through a model selection criteria. A random permutation test evaluates whether the contiguity property is significant. Finally this algorithm is run on Argentinean data. Evaluating the proposed methodology concludes the paper.

Keywords: relative specialization, specialized agglomeration, lattice data, spatial clustering, spatial cluster detection, permutation bootstrap.

Corresponding Author: Michel MOUCHART, ISBA, voie du Roman Pays 20 - L1.04.01, B-1348 Louvain-la-Neuve(B); e-mail: michel.mouchart@uclouvain.be; phones: office +3210474318, mobile +32472335927

1 Introduction

1.1 Spatial clustering framework

Spatial cluster detection is an important issue in many fields, ranging for instance from astrophysics to forest ecology, including epidemiology and many others. There is accordingly quite a vast and rapidly expanding literature. In particular, Openshaw *et al.* (1988), Anselin (1995), Ester *et al.* (1996 and 1998), Chaudhuri and Marron (1999 and 2000), Wakefield *et al.* (2000), Glaz *et al.* (2001 and 2009), Lawson and Denison (2002), Diggle (2003), Kulldorff *et al.* (2003), Tango (2010), Saini and Kaur (2014) illustrate the relevance of this topic in all these fields as well as the great variety of approaches for detecting spatial clusters. This paper is basically motivated by spatial economics.

A significant group of contributions uses the spatial scan statistic first proposed by Naus (1965) and further developed by Cressie (1977), Openshaw *et al.* (1988), Besag and Newell (1991), Kulldorff and Nagarwalla (1995), Kulldorff (1997), Glaz *et al.* (2001 and 2009), Tango and Takahashi (2005), Kulldorff *et al.* (2006), Neill *et al.* (2006), Loh (2011), among many others; these contributions have been used to detect disease cluster in epidemiology. Our approach is essentially an elaboration of the basic ideas of the scan method proposed by Besag and Newell (1991) in which, given a complete set of regions, individual regions are extracted and progressively aggregated to contiguous regions in order to find the most significant cluster, evaluating all possible cluster schemes that can be formed from these regions. We work in a strictly discrete space, with lattice data on a fixed set of polygons. A model selection procedure, based on the BIC criterion, is used in the algorithm for building optimal clusters. In most works, significance testing is concerned with a null hypothesis of spatial uniformity whereas in this paper we are concerned with the significance of the contiguity condition, hence the reliance on permutation bootstrapping.

1.2 Economic background of industrial agglomeration

Economic activity is spatially concentrated. Spatial concentration generates agglomeration economies, notably upstream-downstream linkages, which help firms become more productive. These positive effects involve a critical mass of workers and infrastructure, and dense networks of suppliers and collaborators. The central role of agglomeration economies on the spatial structure of the economy has inspired a large literature focused on trying to understand their causes or origins and the dynamic of these economies, as well as most adequate industrial policies for their consolidation and promotion wherever the industrial agglomerations are weak or nonexistent. It has been recognized that the industrial policy objectives can be better fulfilled if they are more sensitive to places and sectors in design and delivery (Donato 2002 and Nathan and Overman 2013).

The explanations about the existence and determinants of the industrial agglomerations started with the pioneer works of Marshall (1890), Scitovsky (1954), Arrow (1962), Becattini (1979), and

Romer (1986), regarding the localization economies of firms located near other firms of the same industry; the literature often refers to it as forces that lead to the clustering of firms of similar activities and to regions that tend to specialize). Furthermore, Jacobs (1969), Henderson (1985), Lucas (1988) and Glaeser *et al.* (1992) are concerned with the aspect of the urban economies that refer to the advantages arising from the spatial concentration of firms of diverse activities (urbanization economies tend to lead to diverse regions). The advantages arise from the industrial agglomeration environments. The focus of this paper is on the identification of specialized agglomeration for each activity separately while controlling whether a condition of contiguity, among the regions conforming a putative agglomeration, gives evidence of a role played by localization economies. The Application, Section 3, reveals that the results of that control may be different from one activity to another one.

Agglomeration economies may appear in different geographical scales and may involve different disaggregation levels, and consequently, a certain space scale is not necessarily equivalent to another (Arbia 1989 and 2001, Krugman 1991, Arbia and Espa 1996, Anas *et al.* 1998, Duranton and Overman 2005, Arbia *et al.* 2008, among others). In this paper, we analyze data referring to a discrete space (lattice or areal data regarding administrative divisions), and the boundaries cannot be ignored, given that economic conditions can change abruptly due to changes in the tax system, in transportation costs, or to the impact of public policies at regional and sectorial levels.

The specialized agglomerations may be analyzed either in absolute (with respect to a given country) or in relative terms. In both cases, the forces that lead to the concentration tend to specialize the regions. In this paper, the specialized agglomerations are defined from the analysis of concentration in relative terms, in order to operate a distinction with respect to “just” agglomerations based on the analysis in absolute terms.

1.3 Identification of specialized agglomerations

Our objective is to make a contribution to the detection of specialized agglomerations with lattice data understood as a cluster of contiguous regions, and focuses the attention on the relative industrial concentration of each activity separately, measured through the discrepancy between the activity specific distribution of the regions (*i.e.* the distribution of regions conditional on the activity), and the overall distribution of the regions (*i.e.* the marginal distribution of the regions); detail on this concept in *e.g.* Haedo and Mouchart (2012). This approach differs from Mori and Smith (2011) who used the area of regions as benchmark distribution of “complete spatial randomness”, and should be rather interpreted as absolute concentration.

An automatic (algorithmic) procedure for detecting agglomerations of relative industrial concentration is proposed, starting from a considerable extension of Donato and Haedo (2002) and the third chapter of Haedo (2009). We are not looking for the detection of one particular agglomeration, we rather look for a cluster scheme, *i.e.* a finite partition of regions such that each cluster is an

agglomeration (thus a cluster of contiguous regions, possibly a single region). Here, the concept of specialized agglomeration refers to the homogeneity, within the agglomeration, of the regions relatively to industrial concentration and to the heterogeneity between the agglomerations; thus for a particular activity, some agglomerations are over-specialized while others are under-specialized.

This paper makes a dual contribution to the field of agglomeration detection. Firstly, we rely on a structural modeling approach; on this approach see *e.g.* Mouchart, Russo and Wunsch (2009). In particular, we endeavour to structure the model around distributions characterized by parameters meaningful from a spatial economics point of view. Based on a three-step structure of individual multiple choices, the formation of agglomerations is accordingly modeled through a multinomial distribution, where each agglomeration is characterized by a probability of individual location that is different among agglomerations and is depending on the activity. Secondly, we develop a greedy algorithm for the construction of a cluster scheme that is coherent with the statistical model just developed and that relies on a natural concept of specialization.

This algorithm is initialized by postulating a null hypothesis of no specialized agglomeration, *i.e.* an initial cluster scheme of a unique cluster made of all regions, and proceeds by separating progressively from that unique cluster regions potentially able to produce most significant partitions, through a model selection criterion based on the highest values of BIC on the null hypothesis H_0 of no specialized agglomeration; note that our null hypothesis is to be compared with Mori and Smith (2011) where their null hypothesis H_0 should be rather interpreted as no “just” agglomeration. A next step evaluates whether the contiguity property is significant, by a bootstrapping method based on random permutations. The stopping rule is based on the observation that the sequence of the optimum BIC as a function of the number of steps is roughly unimodal; we therefore take as a stopping rule the change of the sign of the slope of a least squares linear trend on the past steps. Finally, the algorithm is run on Argentinean data.

The paper is concluded by evaluating the achievements of the proposed methodology.

2 Selection of a best specialized agglomerations scheme

In this section, we propose a structural model of agglomeration formation and an algorithm used to identify a best cluster scheme of specialized agglomerations. We start by describing the data base underlying the algorithm along with some notational device, firstly at the level of the regions, next at the level of the clusters of regions.

2.1 Structure of the data and notation

For a given country, let us consider a finite set of administrative regions $i \in \mathcal{I} = \{1, \dots, I\}$, and a finite set of activities $j \in \mathcal{J} = \{1, \dots, J\}$. The administrative nature of the regions refers to two

aspects. Firstly, the number of regions is finite. Secondly, the boundaries of the regions are designed exogenously, *i.e.* independently of the problem under analysis; in particular the areas of the different regions are typically quite heterogeneous. Furthermore, this administrative nature is crucial for the availability of the data.

For each pair $(i, j) \in \mathcal{I} \times \mathcal{J}$, we observe the number N_{ij} of primary units; these could be typically number of employees or number of firms. These data refer to lattice (or, areal) data as the N_{ij} are characteristics of the area defined by region i . Putting together these data, we obtain a two-way $I \times J$ contingency table $\mathbf{N} = [N_{ij}]$ that also produces row, column and table totals denoted as follows:

$$N_{i\cdot} = \sum_{j=1}^J N_{ij}; \quad N_{\cdot j} = \sum_{i=1}^I N_{ij}; \quad N_{\cdot\cdot} = \sum_{i=1}^I \sum_{j=1}^J N_{ij} = \sum_{j=1}^J N_{\cdot j} = \sum_{i=1}^I N_{i\cdot}. \quad (1)$$

This paper is concerned with the relative industrial concentration, *i.e.* the concentration of one activity in some regions. Following Haedo and Mouchart (2012), the basic tools for these analyses are derived from the profiles provided by the contingency table $\mathbf{N} = [N_{ij}]$; more explicitly¹: region i may be characterized by the profile (or conditional distribution) of the i -th row $p_{\cdot|i} = (p_{1|i}, \dots, p_{j|i}, \dots, p_{J|i})$ to be compared with the global row profile (or marginal distribution) $p_{\cdot\cdot} = (p_{\cdot 1}, \dots, p_{\cdot j}, \dots, p_{\cdot J})$, where $p_{j|i} = \frac{N_{ij}}{N_{i\cdot}}$ and $p_{\cdot j} = \frac{N_{\cdot j}}{N_{\cdot\cdot}}$; similarly, activity j may be characterized by the profile (or conditional distribution) of the j -th column $p_{i|\cdot} = (p_{1|j}, \dots, p_{i|j}, \dots, p_{I|j})$ to be compared with the global column profile (or marginal distribution) $p_{\cdot\cdot} = (p_{1\cdot}, \dots, p_{i\cdot}, \dots, p_{I\cdot})$, where $p_{i|j} = \frac{N_{ij}}{N_{\cdot j}}$ and $p_{i\cdot} = \frac{N_{i\cdot}}{N_{\cdot\cdot}}$.

It may be convenient to refer to the areas rather than to the (arbitrary) labels of the regions. Thus we also denote the country, considered as an area, as Ω and the disjoint regions as Ω_i . Evidently, the regions Ω_i provide a partition of the country Ω : $\Omega_i \neq \emptyset$, $\Omega_i \cap \Omega_{i'} = \emptyset$ ($i \neq i'$) and $\bigcup_{i=1}^I \Omega_i = \Omega$. Therefore, $p_{i\cdot} = p(\Omega_i)$.

Let us operate a partition of the I regions into M agglomerations, or “grouped regions” to be called “g-regions” for the ease of exposition. This regrouping may be written in terms of the labels:

$$\mathcal{I} = \{1, 2, \dots, I\} = \bigcup_{m=1}^M \mathcal{I}_m \quad \mathcal{I}_m \cap \mathcal{I}_{m'} = \emptyset \quad (m \neq m') \quad \#(\mathcal{I}_m) = I_m \quad \sum_m I_m = I \quad (2)$$

Let us define accordingly $N_{m\cdot} = \sum_{i \in \mathcal{I}_m} N_{i\cdot}$ and $N_{m,j} = \sum_{i \in \mathcal{I}_m} N_{ij}$. Using g to denote relative frequencies on the space of the g-regions, we successively define $g_{\cdot\cdot} = (g_{1\cdot}, \dots, g_{m\cdot}, \dots, g_{M\cdot})$ and $g_{m|\cdot} = (g_{1|j}, \dots, g_{m|j}, \dots, g_{M|j})$, where $g_{m\cdot} = \sum_{i \in \mathcal{I}_m} p_{i\cdot} = \frac{N_{m\cdot}}{N_{\cdot\cdot}}$ and $g_{m|j} = \sum_{i \in \mathcal{I}_m} p_{i|j} = \frac{N_{m,j}}{N_{\cdot j}}$. Similarly for the conditional probabilities:

$$p_{i\cdot|m} = \frac{p_{i\cdot}}{g_{m\cdot}} \mathbb{I}_{\{i \in \mathcal{I}_m\}} = \frac{N_{i\cdot}}{N_{m\cdot}} \mathbb{I}_{\{i \in \mathcal{I}_m\}} \quad \text{and} \quad p_{i|j,m} = \frac{p_{i|j}}{g_{m|j}} \mathbb{I}_{\{i \in \mathcal{I}_m\}} = \frac{N_{ij}}{N_{m,j}} \mathbb{I}_{\{i \in \mathcal{I}_m\}}. \quad (3)$$

¹When the components of a vector are indexed by i (regions) or by j (activities), we use an arrow above the index that defines the components of the vector.

This regrouping may also be viewed in terms of a cluster scheme of areas, $C = \{\mathcal{I}_{1|C}, \dots, \mathcal{I}_{m|C}, \dots, \mathcal{I}_{M|C}\}$, consisting of disjoint regional clusters:

$$\Omega_{\mathcal{I}_{m|C}} = \bigcup_{i \in \mathcal{I}_{m|C}} \Omega_i \quad \text{with} \quad \bigcup_{m=1}^M \Omega_{\mathcal{I}_{m|C}} = \Omega \quad (4)$$

Thus, when we want to make explicit the role of a particular cluster scheme C , we also write, instead of $g_{m\cdot}$: $g_{m\cdot|C} = g(\Omega_{\mathcal{I}_{m|C}}) = \sum_{i \in \mathcal{I}_{m|C}} p_{i\cdot}$.

Two extreme cases should be noted:

a) $M = I$: no grouping; *i.e.* each region remains isolated, more explicitly:

$$C_{min} = \{\{1\}, \{2\}, \dots, \{I\}\} \quad \mathcal{I}_{m|C_{min}} = \{i\} \quad (5)$$

$$g_{\vec{m}\cdot|C_{min}} = p_{\vec{i}\cdot} \quad g_{\vec{m}|j;C_{min}} = p_{\vec{i}|j} \quad p_{i\cdot|m;C_{min}} = \mathbb{I}_{\{\{i\}=\mathcal{I}_{m|C_{min}}\}} \quad (6)$$

$$p_{i|j,m;C_{min}} = \mathbb{I}_{\{\{i\}=\mathcal{I}_{m|C_{min}}\}} \mathbb{I}_{\{N_{ij}>0\}} \quad (7)$$

b) $M = 1$: maximal grouping, *i.e.* each region enters a unique g-region, more explicitly:

$$C_{max} = \{1, 2, \dots, I\} \quad \mathcal{I}_{m|C_{max}} = \mathcal{I} \quad m \equiv 1 \quad (8)$$

$$g_{\vec{m}\cdot|C_{max}} = g_{\vec{m}|j;C_{max}} = \{1\} \quad p_{i\cdot|m;C_{max}} = p_{i\cdot} \quad p_{i|j,m;C_{max}} = p_{i|j} \quad (9)$$

2.2 The multinomial model

We now develop a model in terms of the $N_{\cdot}$ primary units labeled by u , *i.e.* $u \in \mathcal{U}$ with $\#(\mathcal{U}) = N_{\cdot}$ along with a localization function $\ell : \mathcal{U} \rightarrow \Omega$ where $\ell(u)$ stands for the localization of u in Ω and an activity function $a : \mathcal{U} \rightarrow \mathcal{J}$ where $a(u)$ stands for the activity of the primary unit u . For each pair (i, j) , the primary unit u is associated with a binary variable $x_{ij}^u = \mathbb{I}_{\{\ell(u) \in \Omega_i, a(u)=j\}}$. Clearly, $\sum_{u \in \mathcal{U}} x_{ij}^u = N_{ij}$.

The algorithm developed in Section 2.4 allows a flexible way of dealing with the regions to be taken into account when modeling a particular activity j ; thus, for a given j it may be specified that only the regions with $N_{ij} > 0$, or only the regions with N_{ij} larger than some pre-specified limit (*e.g.* 5) will be the object of modeling. For a particular activity j we have I_j regions to be taken into account and we construct the set $\mathcal{A}_j$ of the primary units in the relevant regions and active in activity j with $A_j = \#(\mathcal{A}_j)$. Next we define an $A_j \times I_j$ matrix $X_j = [x_{ij}^u]$ where $i = 1, \dots, I_j$ in columns (up to an eventual reordering of the regions, putting first the relevant ones) and u , in rows, run over $\mathcal{A}_j$. As X_j is an incidence matrix with elements equal to x_{ij}^u for $u \in \mathcal{A}_j$, the sum of each row is equal to 1.

We now introduce a model aimed at representing how the data have been generated and may be eventually interpreted; the notation now distinguishes unknown parameters, in Greek letters, and functions of data (estimators or statistics) in Latin letters although we also use Greek letters with hat for estimators. We start with an arbitrary cluster scheme $C = \{\mathcal{I}_{1|C}, \dots, \mathcal{I}_{m|C}, \dots, \mathcal{I}_{M|C}\}$.

The stochastic model accordingly involves three categorical random elements, namely: regions (i), activity (j), and cluster (m). When randomly drawing a statistical unit u from the universe $\mathcal{U}$, we therefore need to specify a trivariate distribution $\pi_{i,j,m}$. The process is decomposed as follows:

1. individual u selects an activity j according to a distribution $\pi_{\cdot j}$;
2. conditionally on j , individual u selects a cluster m according to a distribution $\gamma_{m|j;C}$;
3. conditionally on (j, m) , individual u selects a region Ω_i within cluster $\Omega_{\mathcal{I}_{m|C}}$ according to a distribution $\rho_{i|j,m;C}$.

In short, we consider as structural the following decomposition:

$$\pi_{i,j,m|C} = \pi_{\cdot j} \gamma_{m|j;C} \rho_{i|j,m;C} \quad i \in \mathcal{I}_{m|C} \quad (10)$$

From (4), we have

$$\gamma_{m|j;C} = \sum_{i \in \mathcal{I}_{m|C}} \pi_{i|j}, \quad \gamma_{m|C} = \sum_{i \in \mathcal{I}_{m|C}} \pi_{i\cdot} \quad (11)$$

Remark. In the limit case of a one-term partition, *i.e.* $M = 1$ and $C = \{\mathcal{I}\}$, we have: $\gamma_{m|j;C} = \gamma_{m|C} = 1$, $\rho_{i|j,m;C} = \pi_{i|j}$ and (10) boils down to $\pi_{i,j} = \pi_{\cdot j} \pi_{i|j}$. ■

In next subsection we design a procedure for identifying specialized agglomeration, by means of a cluster scheme C different for each activity j . Therefore, we do not discuss the specification of $\pi_{\cdot j}$ and conduct the whole analysis conditionally on j . The concept of specialized agglomeration is to be built progressively. As a first step, we use, as an hypothesis maintained throughout the paper:

$$H_m : \quad \rho_{i|j,m;C} = \rho_{i|m;C} \quad \text{with} \quad \sum_{i \in \mathcal{I}_{m|C}} \rho_{i|m;C} = 1 \quad (12)$$

This hypothesis of conditional independence, namely $i \perp j \mid m; C$, means that once an individual has selected an activity j , he selects a cluster m likely to be suitable for his activity j and when, conditionally on his choice (j, m) , he selects a region, within the cluster m , he considers that *within* $\Omega_{\mathcal{I}_{m|C}}$ the regions exert an attraction independent of his sector of activity. Thus cluster m is considered as specialized as long as it is selected conditionally on j and that the area within $\Omega_{\mathcal{I}_{m|C}}$ is homogeneous as far as attractivity relative to activity j is concerned. In the next subsection, we make the concept of specialized agglomeration more explicit. Moreover, since

$$\rho_{i|m;C} = \frac{\pi_{i\cdot}}{\gamma_{m|C}}, \quad (13)$$

the maintained hypothesis also assumes that the attractivity of region i , within cluster m , only depends of its general (or marginal) size $\pi_{i..}$, relatively to the cluster size, $\gamma_{m|C}$. This hypothesis is built within a multinomial setting and is different from the Mori and Smith (2011) condition of “Complete spatial randomness”, which assumes that (using our notation) $\rho_{i|m;C}$ is equal to the ratio between the area of Ω_i and the area of $\Omega_{\mathcal{I}_m}$. Thus in (13) all the activities are taken into account through $\pi_{i..} = \sum_j \pi_{ij.}$; in other words the role of $\pi_{i..}$ in $\rho_{i|m;C}$ is to provide a proxy for the set of characteristics of region i being favorable to the development of the activity in general. From now on, we explicitly write that the number of clusters depends on the cluster scheme under consideration; hence we shall write $M(C)$ instead of M .

Under our maintained hypothesis we have:

$$\pi_{i,m|j;C} = \gamma_{m|j;C} \rho_{i|m;C} \quad \forall i \in \mathcal{I}_{m|C} \quad (14)$$

and, for a given j , the probability of the data matrix X_j may be written as

$$p(X_j | C) = \prod_{u \in \mathcal{U}} \prod_{1 \leq m \leq M(C)} \prod_{i \in \mathcal{I}_{m|C}} \pi_{i,m|j;C}^{x_{ij}^u} \quad (15)$$

$$\begin{aligned} &= \prod_{1 \leq m \leq M(C)} \prod_{i \in \mathcal{I}_{m|C}} [\gamma_{m|j;C} \rho_{i|m;C}]^{N_{ij}} \\ &= \left[\prod_{1 \leq m \leq M(C)} \gamma_{m|j;C}^{N_{m,j|C}} \right] \cdot [b(X_j)] \end{aligned} \quad (16)$$

where

$$N_{m,j|C} = \sum_{i \in \mathcal{I}_{m|C}} N_{ij} \quad (17)$$

$$b(X_j) = \prod_{1 \leq m \leq M(C)} \prod_{i \in \mathcal{I}_{m|C}} \rho_{i|m;C}^{N_{ij}} \quad (18)$$

As the parameter of the factor $b(X_j)$ does not depend on j , we may factorize the likelihood function as

$$L(X_j | C) = L_1(\gamma_{\vec{m}|j;C} | X_j) L_2([\rho_{i|m;C}] | X_j) \quad (19)$$

This factorization not only separates the parameters $\gamma_{\vec{m}|j;C}$ and $\rho_{i|m;C}$ but also isolates the term specific to activity j .

2.3 The concept of specialized cluster

The previous section proposes a modeling for the formation of an arbitrary cluster scheme of regions. In this section, we identify specialized clusters for a specified activity j and propose a measure of the intensity of the specialization of a country for the specified activity j . We first make the concept of cluster specialization operational.

The starting point of our analysis is the well established Location Quotient (Florence 1939), also known as the (estimated) *Hoover-Balassa coefficient*, for the cell (i, j) that may be written in several forms, as:

$$LQ_{ij} = \frac{\hat{\pi}_{ij}}{\hat{\pi}_i \cdot \hat{\pi}_{\cdot j}} = \frac{\hat{\pi}_{j|i}}{\hat{\pi}_{\cdot j}} = \frac{\hat{\pi}_{i|j}}{\hat{\pi}_i} = \frac{N_{ij} N_{\cdot\cdot}}{N_{i\cdot} N_{\cdot j}} \quad (20)$$

This location quotient reveals the following feature of activity j in region i :

$$\begin{aligned} LQ_{ij} &= 1 && \text{non-specialization} \\ &> 1 && \text{over-specialization} \\ &< 1 && \text{under-specialization} \end{aligned} \quad (21)$$

where the “non-specialization” corresponds *locally* - *i.e.* at the level of the cell (i, j) - to the row-column independence. Indeed, the last two equalities in (20) stress the point that specialization is an issue concerning the global structure at the country level: the absence of specialization of a cell (i, j) means that, *relative to the marginal distributions*, activity j is not over-(nor under-) represented in region i (*i.e.* there is locally no relative regional specialization) *and* that region i is not over-(nor under-) represented for the activity j (*i.e.* there is locally no relative industrial concentration). On the duality between regional specialization and industrial concentration, see Haedo and Mouchart (2012, Table 1). Thus, “location” means that LQ is localized in a cell (i, j) .

At the level of regions, the concept of industrial concentration for a given activity j compares the conditional distributions of the regions $\pi_{i|j}$ for activity j with a benchmark distribution of the regions, say $b_i = (b_1, \dots, b_i, \dots, b_I)$, and is measured by means of some discrepancy $d(\pi_{i|j} \mid b_i)$.

Remark. As indicated, with more detail in Haedo and Mouchart (2012), the absolute industrial concentration takes a uniform distribution, *i.e.* $b_i = \frac{1}{I} \forall i$, as a benchmark and therefore considers that a not concentrated activity would be spread uniformly over the regions. The relative industrial concentration takes the marginal distribution $\pi_{\cdot}$ as a benchmark, *i.e.* $b_i = \pi_{i\cdot}$. In that case, the marginal distribution is taken as an indicator of the global attractivity of a region and a non-concentrated activity would be spread over the regions according to the same marginal distribution $\pi_{\cdot}$. Other benchmark distributions might be and have been considered, in particular: $b_i = \frac{A_i}{\sum_i A_i}$ where A_i is the area of region i (see *e.g.* Mori and Smith 2011) or $b_i = \frac{P_i}{\sum_i P_i}$ where P_i is the population of region i (see *e.g.* Lawson *et al.* 1999, Lawson 2006, among others). Note that in epidemiology, the population P_i may represent the population at risk for a given disease, in which case $b_i = \frac{P_i}{\sum_i P_i}$ may be a reasonable benchmark for defining an epidemic concentration. ■

Now we identify specialized clusters relatively to a specified activity j . Here a cluster $\mathcal{I}_m$ is over-specialized (resp. under-specialized) with respect to activity j when the $\pi_{i|j}$ ’s for $i \in \mathcal{I}_m$ are significantly greater (resp. smaller) than the country-wide average $\pi_{i\cdot}$, taken as a benchmark (remember that $\pi_{i\cdot}$ is an average of the $\pi_{i|j}$ ’s, *i.e.* $\pi_{i\cdot} = \sum_j \pi_{i|j} \pi_{\cdot j}$) and in view of (20) this

is equivalent to the location quotients being significantly larger, or smaller, than 1. Under the maintained hypothesis (12), it may be seen from (15)-(16) that the identification of specialized clusters and the construction of a cluster scheme C of specialized clusters is to be based on the properties of $\gamma_{m|j;C}$.

A (fully) non-specialized cluster $\mathcal{I}_m$ is a cluster where $LQ_{ij} = 1$ for $\forall i \in \mathcal{I}_m$ (equivalently, $\hat{\pi}_{i|j} = \hat{\pi}_{i\cdot}$ or $\hat{\pi}_{j|i} = \hat{\pi}_{j\cdot}$). This hypothesis, extended to each cluster m , *i.e.* $LQ_{ij} = 1 \ \forall i \in \mathcal{I}_m$, implies, due to (11),

$$H_0 : \quad \gamma_{m|j;C}^{(0)} = \gamma_{m|C}^{(0)} \quad m = 1, \dots, M. \quad (22)$$

This hypothesis means that for the activity j and for the cluster scheme C , there is no industrial concentration on the whole country. The maximum likelihood estimate under H_0 is therefore:

$$\hat{\gamma}_{m|j;C}^{(0)} = \hat{\gamma}_{m|C}^{(0)} = \frac{N_{m\cdot|C}}{N_{\cdot\cdot}}. \quad (23)$$

The absence of industrial concentration for activity j , underlying (23), is relative to a particular cluster scheme C . The maximized log likelihood under H_0 is

$$\ln \hat{L}^{(0)}(X_j | C) = \sum_{1 \leq m \leq M(C)} N_{m,j|C} \ln \left(\frac{N_{m\cdot|C}}{N_{\cdot\cdot}} \right) + \ln a(X_j) \quad (24)$$

where $a(X_j)$ corresponds to the term $L_2([\rho_{i|m;C}] | X_j)$ in (19) and gathers the terms unaffected by the null hypothesis. As an alternative hypothesis H_1 , the parameter $\gamma_{m|j;C}$ is left unconstrained and may be estimated as

$$\hat{\gamma}_{m|j;C}^{(1)} = \frac{N_{m,j|C}}{N_{\cdot j}}. \quad (25)$$

Thus when the alternative hypothesis is assumed for each cluster $\mathcal{I}_m$ of a cluster scheme C , the maximized log likelihood under H_1 is

$$\ln \hat{L}^{(1)}(X_j | C) = \sum_{1 \leq m \leq M(C)} N_{m,j|C} \ln \left(\frac{N_{m,j|C}}{N_{\cdot j}} \right) + \ln a(X_j) \quad (26)$$

The log-likelihood-ratio statistic to test H_0 against H_1 , under H_m is

$$\begin{aligned} T(X_j | C) &= -2 \left[\ln \hat{L}^{(0)}(X_j | C) - \ln \hat{L}^{(1)}(X_j | C) \right] \\ &= -2 \left[\sum_{1 \leq m \leq M(C)} N_{m,j|C} \ln \left(\frac{N_{m\cdot|C}}{N_{\cdot\cdot}} \right) - \sum_{1 \leq m \leq M(C)} N_{m,j|C} \ln \left(\frac{N_{m,j|C}}{N_{\cdot j}} \right) \right] \\ &= 2 \left[\sum_{1 \leq m \leq M(C)} N_{m,j|C} \ln \left(\frac{N_{m,j|C}/N_{\cdot j}}{N_{m\cdot|C}/N_{\cdot\cdot}} \right) \right] \end{aligned} \quad (27)$$

$$= 2 \left[N_{\cdot j} \sum_{1 \leq m \leq M(C)} p_{m\cdot|j} \ln \left(\frac{p_{m\cdot|j}}{g_{m\cdot}} \right) \right] \quad (28)$$

$$= 2 N_{\cdot j} d(p_{\bar{m}|j} | g_{\bar{m}\cdot}), \quad (29)$$

where $d(\cdot | \cdot)$ is the (non-symmetric) Kullback-Leibler divergence between two distributions. Therefore, the test statistic (30) may be viewed as a measure of relative industrial concentration of activity j among the clusters (see Haedo and Mouchart 2012, for additional discussion).

The argument of the logarithm in (28) is the location quotient obtained after clustering the regions; more explicitly, it is the cross product ratio in the 2×2 -table

$$\begin{bmatrix} N_{m,j|C} & N_{m\cdot|C} \\ N_{\cdot j} & N_{\cdot\cdot} \end{bmatrix}.$$

Notwithstanding the existence of agglomeration economies, firms' location decision is modeled as independent of these agglomeration economies as the model is essentially a static one that does not present the dynamics of agglomeration formation. The weight $N_{m,j|C}$ in (28) of the logarithm of LQ serves to solve a small areas problem in line with the works of Moineddin *et al.* (2003), O'Donoghue and Gleave (2004) and Guimarães *et al.* (2003 and 2009). From (20), it is seen that the argument of the logarithm in (29) is also equal to $p_{\cdot j|m}/p_{\cdot j}$, making clear the fact that the location quotient for cluster m and activity j is independent of the area of cluster m and of the weight g_m of the cluster m ; this is at variance from Mori and Smith (2011), as mentioned in the Remark above in this section.

In (22), H_0 represents $M(C) - 1$ restrictions for a fixed j under the condition that the sum in m is equal to 1. Therefore under H_0 , the test statistics $T(X_j | C)$ is asymptotically distributed chi-square with $M(C) - 1$ degrees of freedom. The null hypothesis is rejected when the value of the likelihood-ratio test is sufficiently large, or when the corresponding p -value is sufficiently small.

2.4 Detecting a specialized agglomerations scheme

Up to now, we have developed a “space free” analysis as long as the labels of the region is arbitrary and convey no information on the localization of the regions. Now we introduce an idea of distance-based pattern by means of the concept of agglomerations that are clusters made of neighboring regions. The simplest case is obtained when neighboring regions are interpreted as contiguous regions. In this case, only regrouping contiguous regions is of interest; therefore each $\Omega_{\mathcal{I}_{m|C}}$ should be a connected set of regions. The contiguity matrix (or weights matrix) W formally expresses the proximity links existing between all pair of regions. The elements of the $I \times I$ -matrix W are obtained as the values of the following function:

$$w : \mathcal{I} \times \mathcal{I} \longrightarrow \{0, 1\} \quad \text{where} \quad w(i_1, i_2) = \mathbb{I}_{\{i_1 \text{ and } i_2 \text{ are contiguous}\}} \quad (30)$$

Note that W is symmetric ($W = W'$) with 1s on the main diagonal and the sum of the rows (or, of the columns) minus 1 is equal to the number of contiguous regions of each region in the set $\mathcal{I}$. Therefore, the set of regions contiguous to a cluster $\mathcal{I}_{m|C}$ may be written as:

$$v(\mathcal{I}_{m|C}) = \{i_1 \in \mathcal{I} \setminus \mathcal{I}_{m|C} \mid \exists i_2 \in \mathcal{I}_{m|C} : w(i_1, i_2) = 1\} \quad (31)$$

Remark on the W matrix. The concept of contiguity underlying (31) deserves to be made more precise. Contiguity may mean at least one point common in the boundaries of the two contiguous regions, in which case W is a first order queen weights matrix, or contiguity may mean a partly common frontier with more than one point, in which case W is a first order rook weights matrix; for more information on weights matrices see O’Sullivan and Unwin (2010). ■

Our final objective is to construct a cluster scheme C made of specialized agglomerations that maximizes the heterogeneity among agglomerations up to a penalization on the number of parameters. In this context, a cluster scheme is “best” as long as the agglomerations are the most different possible for their specialization with respect to activity j . Here, the heterogeneity among clusters is measured by the divergence $d(p_{\vec{m}|j} \mid g_{\vec{m}.})$, as given in equation (30).

A particular cluster scheme generates a particular model; indeed, from (10), a cluster scheme C may be viewed as a model of agglomeration formation. Therefore, selecting a cluster scheme out of a set of possible schemes may be treated as a problem of model selection. We develop an algorithmic procedure, based on a greedy approach and, at each step, on a model selection criterion, balancing the goodness of fit and the number of parameters.

There are a lot of model selection procedures; particularly relevant for our purposes are the following: Diggle 2000, Burnham and Anderson 2002 and 2004, Kulldorff, Tango and Park 2003, Yang 2005, Gómez-Rubio, Ferrándiz-Ferragud and López-Quílez 2005, Claeskens and Hjort 2008, among others). The critical alpha criterion has the advantage of a unique asymptotic distribution, namely uniform on the unit interval, irrespectively of the number of degrees of freedom; however a drawback of the criterion is to require a comparison among very small numbers. The Akaike information criterion (AIC), (Akaike 1974), and the Bayesian information criterion (BIC) have also been widely discussed.

Maximizing AIC or BIC is equivalent to maximize the divergence $d(p_{\vec{m}|j} \mid g_{\vec{m}.})$, up to different penalizations for the number of parameters. The Bayesian information criterion (BIC), introduced by Schwartz (1978), has proved to yield a consistent estimator of the true model whenever it is included among the candidate models. The best model is that with the smallest

$$BIC^{(1)}(X_j \mid C) = (M(C) - 1) \ln(N_{..}) - 2 \ln(\hat{L}^{(1)}(X_j \mid C)). \quad (32)$$

Reformulating this measure in terms of BIC -differences, we obtain:

$$BIC(X_j \mid C) = BIC^{(0)}(X_j \mid C) - BIC^{(1)}(X_j \mid C) \quad (33)$$

$$= -2 \ln[\hat{L}^{(0)}(X_j \mid C) - \hat{L}^{(1)}(X_j \mid C)] - (M(C) - 1) \ln(N_{..}) \quad (34)$$

$$= T(X_j \mid C) - (M(C) - 1) \ln(N_{..}) \quad (35)$$

and the best cluster schemes C is that with larger value of $BIC(X_j \mid C)$.

A sketch of the algorithm.

The number of possible cluster schemes can be enormous for an even modest number of basic regions. Thus, it is necessary to consider limited search procedures that yield reasonable approximations to best cluster schemes. Our approach is essentially an elaboration of the basic ideas of the scan method proposed by Besag and Newell (1991) in which, given a complete set of regions, individual regions are extracted and progressively aggregated to contiguous regions in order to find the most significant cluster, evaluating all possible cluster schemes that can be formed from these regions. This algorithm may also be viewed in the family of hierarchical divisive clustering algorithms (for an interesting synthesis of this topic, one may consult http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/hierarchical.html, see also <http://nlp.stanford.edu/IR-book/html/htmledition/divisive-clustering-1.html>). Experience suggests that the divisive structure of this algorithm is likely to be suitable for taking due account of the constraint of contiguity.

For each activity j , we develop a greedy forward algorithm that uses the BIC as a selection criterion. We start with a baseline configuration $\mathcal{I}_{[j]}$ made of all regions with $N_{ij} > 0$; we might also specify N_{ij} larger than some specified minimum. For the sake of exposition, we shall assume that $\mathcal{I}_{[j]} = \mathcal{I} \ \forall j$.

Remark. In some cases, the selection of individuals is made on those localized in regions where $LQ_{ij} > 1 + \varepsilon$ or $LQ_{ij} < 1 - \varepsilon$, for some $\varepsilon \geq 0$ (for fixed j), with the objective of analyzing separately over- or under- specialized agglomerations (see *e.g.* Neill, Moore and Cooper 2006). ■

The algorithm generates, as a first (myopic) version, a sequence of configuration $C_{[k]}^*$ as follows:

Step 0. As an initial step, $C_{[0]}^*$ is the one-term partition:

$$C_{[0]}^* = \{\mathcal{I}\} = C_{max}. \quad (36)$$

In this case, $BIC(X_j | C_{[0]}^*) = T(X_j | C_{[0]}^*) = 0$, because $M(C) = 1$.

Step 1 chooses the region which forms a separated one region cluster and that maximizes the value of $BIC(X_j | C)$. There are I possible regions to choose from. For each cluster scheme under consideration $M(C) = 2$ as the outcome of this first step is a two clusters configuration, namely one cluster formed by the chosen region, say $\{i_{[1]}^*\}$, and the other cluster consisting of all the remaining regions, $\mathcal{I} \setminus \{i_{[1]}^*\}$. This second cluster is a “residual” cluster in the sense that it is not composed of homogenous regions only, as far as specialization is concerned, but rather is to be used as a reservoir from which a new region will be extracted in the following step. For this cluster the condition of contiguity is not required. At the first step, the optimal configuration has accordingly the following form:

$$C_{[1]}^* = \{\{i_{[1]}^*\}, \mathcal{I} \setminus \{i_{[1]}^*\}\} \quad (37)$$

This step is completed by evaluating $BIC(X_j \mid C_{[1]}^*)$. If, however, $\pi_{i|j} = \pi_i$, $\forall i$ for a given j , then $C_{[1]}^* = C_{[0]}^*$ and the algorithm stops.

Step 2 looks for a second region, say $i_{[2]}^*$. The algorithm starts by examining each region i_2 in $\mathcal{I} \setminus \{i_{[1]}^*\}$ and evaluates the BIC -criterion for the corresponding cluster scheme $C_{[1],i_2}$ formed from $C_{[1]}^*$ and i_2 as follows. If $i_2 \in v(\{i_{[1]}^*\})$ - see (32) - , two cluster schemes are possible: either configuration a , namely $C_{[1],i_2,a} = \{\{i_{[1]}^*, i_2\}, \mathcal{I} \setminus \{i_{[1]}^*, i_2\}\}$ (with $M(C) = 2$) or configuration b , namely $C_{[1],i_2,b} = \{\{i_{[1]}^*\}, \{i_2\}, \mathcal{I} \setminus \{i_{[1]}^*, i_2\}\}$ (with $M(C) = 3$). The algorithm retains $C_{[1],i_2}$ as a configuration with the highest BIC. If $i_2 \notin v(\{i_{[1]}^*\})$, only $C_{[1],i_2} = \{\{i_{[1]}^*\}, \{i_2\}, \mathcal{I} \setminus \{i_{[1]}^*, i_2\}\}$ is possible and $M(C) = 3$. The optimum $i_{[2]}^*$ and the corresponding optimum cluster scheme are now defined as:

$$i_{[2]}^* = \arg \max_{i_2 \in \mathcal{I} \setminus \{i_{[1]}^*\}} BIC(X_j \mid C_{[1],i_2}) \quad (38)$$

$$C_{[2]}^* = C_{[1],i_{[2]}^*} \quad (39)$$

Note that we may write $C_{[2]}^*$ as follows:

$$C_{[2]}^* = \{\mathcal{I}_{1|[2]}^*, \dots, \mathcal{I}_{M(C_{[2]}^*)|[2]}^*\} \quad (40)$$

where $M(C_{[2]}^*)$, equal either to 2 or to 3, is the number of elements of the best cluster scheme at the end of step 2, and $\mathcal{I}_{M(C_{[2]}^*)|[2]}^* = \mathcal{I} \setminus \{i_{[1]}^*, i_{[2]}^*\}$ is a residual cluster at the end of step 2. If the maximizer does not improve the BIC -criterion of the previous step, the algorithm stops.

Next steps iterate step 2. More specifically, at the end of step k , the cluster scheme has the form

$$C_{[k]}^* = \{\mathcal{I}_{1|[k]}^*, \mathcal{I}_{2|[k]}^*, \dots, \mathcal{I}_{M(C_{[k]}^*)|[k]}^*\} \quad (41)$$

where $M(C_{[k]}^*)$ is the number of agglomerations after k steps and $\mathcal{I}_{M(C_{[k]}^*)|[k]}^*$ is a residual cluster, thus $M(C_{[k]}^*) \leq k$. Note that some agglomerations might be contiguous without being merged into a unique one because of too different values of the location quotients: this is decided through the BIC -criterion that balances the effect on the divergence against the penalization for the number of parameters, see equations (30) and (36).

Stopping rule. The algorithm stops when no choice of region from the residual cluster of the preceding step provides an increase in the criterion. If we write k^* for the last step before stopping the algorithm, we simplify the notation by writing C^* instead of $C_{[k^*]}^*$.

More on the stopping rule.

The proposed algorithm is a heuristic one: it economizes on a complete enumeration of all possible partitions, but does not ensure that the absolute optimal selection has been achieved. It is also myopic because the algorithm stops as soon as the value of the criterion deteriorates for the

first time. The performance of such an algorithm is to be evaluated from two criteria. Firstly, the proximity of the obtained solution with respect to the true optimal solution should be appreciated in terms of the value of the objective function (*i.e.* is $BIC(C^*)$ close to the BIC of the true optimum?) and in terms of the configuration of the final partition (*i.e.* is the partition C^* close to the true optimum partition?). Secondly, once the number of regions becomes sizable, the computing time becomes an important issue, particularly in view of the simulations to be performed through the permutation bootstrap, to be discussed in the next subsection. These two criteria crucially depend on adopting a more or less sophisticated stopping rule. Neill (2006) proposed quite an interesting discussion on the elaboration of a stopping rule, in a however quite different analytical context.

As a preliminary phase in the elaboration of the stopping rule we first let the algorithm run on simulated data till the exhaustion of the residual cluster (*i.e.* till $\mathcal{I}_{M(C^*)||[k]}^*$ has a simple region) with a moderate number of regions (roughly, between 20 and 100). We observed that (i) the value of the optimum BIC as a function of the number of steps is roughly unimodal but with local peaks of minor importance (ii) once the optimum BIC substantially decreases, the number of agglomerations $M(C_{[k]}^*)$ increases dramatically giving, in BIC , more weight to the penalization of the number of parameters (producing even negative values of the final BIC 's) and producing agglomerations with too small numbers of regions for being significant. Thus stopping the algorithm once the value of BIC decreases for the first time is probably too myopic a rule.

In order to protect against a situation where a first local maximum be followed by higher values of BIC and to avoid a complete exhaustion of the residual cluster, we introduce a more sophisticated stopping rule as follows. We define “ h ”, as a parameter of backward horizon *i.e.* , the number of past criterion values that enters the stopping rule; later we discuss the specification of h . Write, as before, k for the label of the step of the algorithm $k = 1, 2, \dots$, and y_k for the value of the criterion at the end of step k . Once $k \geq h$, consider a linear trend of the last h realizations, *i.e.* evaluate a least squares linear trend of y on the $h + 1$ past steps. The slope of this trend is proportional to the covariance $\frac{1}{h+1} \sum_{i=k-h}^k (y_i - \bar{y})(i - \bar{i})$. The algorithm stops once the slope becomes null or negative and the optimum values is obtained by examining the $h + 1$ last values of y_k , namely

$$k^* = \arg \max_h \{y_k, y_{k-1}, \dots, y_{k-h}\} \quad (42)$$

$$y^* = y_{k^*} = \max \{y_k, y_{k-1}, \dots, y_{k-h}\} \quad (43)$$

The specification of the horizon h may require some trials. Our experience suggests that $h = 60$ is a reasonable choice, ensuring a fair protection against a local maximum followed by higher values of the criterion or against a situation where the objective function would have a very flat behavior for a sizeable number of steps, followed by a steeper ascent behavior.

A related issues is the penalization imposed by $M(C)$, the number of agglomerations in the cluster scheme. AIC weights $(M(C) - 1)$ by a factor of -2 while BIC uses a factor of $-\ln(N_{..})$.

The motivation of BIC is to improve the asymptotic convergence toward a “true” model. This motivation is however irrelevant in the present case. Indeed, what would be the meaning of a “true” agglomeration or of a “false positive”? As a matter of fact we are looking for a “reasonable” cluster scheme with a number of agglomerations that is neither too high nor too low and with agglomerations homogeneous as far as relative concentration is concerned. Thus we have in mind no structural model explaining the frontiers of agglomeration. If this modeling effort were the purpose of the research then the present algorithm might be viewed as an exploratory step for orienting the search for a reasonable model. Experiences show that basing the stopping rule on the p -value of $T(X_j | C)$ produces cluster schemes with too many agglomerations, many of them too small: there is a definite need for penalizing $M(C)$. The choice between AIC and BIC may depend both on I , the number of regions, and on N ., the number of primary units. We recommended some explorations of this issue before embarking to the bootstrapping phase.

2.5 Testing the role of contiguity for specialized agglomerations

Consider a cluster scheme that has been observed, or determined, for a given activity j for which there is some degree of industrial concentration, as measured by an index of relative industrial concentration of the form $d(p_{i|j} | p_i)$. The question is to try to understand why the industrial concentration of activity j tends to cluster into some agglomerations. As a first step it is natural to ask whether the contiguity among regions is a significant factor of clustering into over- or under-specialized agglomerations. Indeed, as a difference from geo-localized data, lattice data provides no information on the intra-regional localizations. But a major role of contiguity, among regions, in the formation of specialized agglomerations provides evidence that, for a specific activity, localization economies, inside an agglomeration, have more impact. Formally, we want to test the null hypothesis that the vector of the location quotients, for a fixed activity j , is invariant for the group of permutations of its coordinates.

The permutation bootstrap is a standard methodology for facing such a question. Indeed, re-distributions of LQ 's among all regions without replacement has been used when assessing spatial dependence between neighbouring regions, see *e.g.* Manly (1991), Zoellner and Schmidtman (1999), Good (2000) and Lawson (2006). Each redistribution is simulated independently of the contiguities among regions, thus independently of the matrix W , and if for each simulation we run the algorithm and compute the optimal BIC , then we may appreciate where the optimal $BIC(X_j | C^*)$ is localized relatively with the distribution of the simulated BIC . In particular, one may decide that when $BIC(X_j | C^*)$ is far in the tail of the distribution of the simulated BIC , it is a signal that contiguity is a significant factor of clustering. The significance of $BIC(X_j | C^*)$ is accordingly evaluated through the bootstrap distribution. More specifically, let us write $BIC^b(X_j | C_b^*)$ for the BIC of the optimal cluster scheme obtained as a result of the b -th simulation and $\hat{F}_{BIC}^B$ for

the empirical distribution function of $BIC^b(X_j | C_b^*)$ obtained after B simulations. The empirical p -value of $BIC(X_j | C^*)$ is therefore:

$$\begin{aligned} p\text{-value}[BIC(X_j | C^*)] &= 1 - \hat{F}_{BIC}^B(BIC(X_j | C^*)) \\ &= \frac{1}{B} \sum_{1 \leq b \leq B} \mathbb{I}_{\{(BIC^b(X_j | C_b^*) > BIC(X_j | C^*)\}} \end{aligned} \quad (44)$$

Thus the bootstrap p -value is, in general, simply the proportion of the bootstrap test statistics $BIC^b(X_j | C_b^*)$ that are more extreme than the observed test statistic $BIC(X_j | C^*)$: rejecting the null hypothesis whenever $p\text{-value}[BIC(X_j | C^*)] < \alpha$ is equivalent to rejecting it whenever $BIC(X_j | C^*)$ exceeds the $1-\alpha$ quantile of $\hat{F}_{BIC}^B$.

Each simulated resampled distribution is independent from each other and from the contiguity matrix W ; it is therefore possible to run the algorithm for each one in parallel or sequentially. In this work, the algorithm was programmed to run the simulations sequentially but considering that the computing time becomes an important issue, in the future the parallel or sequential options will be considered an additional parameter of the algorithm to those mentioned in Appendix.

3 Application to a pair of activities in Argentina

Let us now experiment this algorithm on Argentine data. The data base is made of an $I \times J$ -contingency table $\mathbf{N} = [N_{ij}]$ where $I = 511$ is the number of regions (partidos/departamentos), $J = 23$ the number of activities and N_{ij} the number of firms of less than 200 employees; these data come from the Firms Directory of *Fundación Observatorio PyME* (small and medium firms) of Argentina for the year 2010, at the level of the first 2-digit (divisions) of the International Standard Industrial Classification of manufacturing activities (ISIC- Rev.3.1, items 15 to 37). We want to detect and compare the partitions into specialized agglomerations for two activities, namely Manufacture of wearing apparel, dressing and dyeing of fur ($j = 18$) and Manufacture of machinery and equipment n.e.c. ($j = 29$).

Figure 1 gathers some basic information on the industrial structure of the small and medium firms in Argentina. Above in the figure, two maps give the spatial distribution of the primary units for activities 18, distributed across 250 regions, and 29, distributed across 253 regions. Below a map gives the same information for all the manufacturing activities. For these three maps, the intensity of darkness in each region is proportional to the number of primary units according to a discretized scale provided at the South-East of each map. Finally a table provides the same information in a numerical format with the regions (partidos/departamentos) aggregated into the 24 provinces. This table permits an evaluation of the relative weight of the two activities of interest with respect to the global economy. Thus in terms of small and medium firms activity 18 represents 8.3% and activity 29 represents 4.5%.

Activity $j=18$

Activity $j=29$

Total

Total for provinces

Province	Primary units			Total
	Act18	Act29	Others	
Buenos Aires	2836	1722	35164	39722
Capital Federal	5346	1688	29396	36430
Catamarca	19	4	285	308
Chaco	30	13	959	1002
Chubut	22	14	781	817
Córdoba	426	517	8680	9623
Corrientes	16	11	635	662
Entre Ríos	47	54	2344	2445
Formosa	12	2	244	258
Jujuy	13	12	533	558
La Pampa	14	23	667	704
La Rioja	8	2	219	229
Mendoza	133	227	4081	4441
Misiones	25	46	1483	1554
Neuquén	13	13	686	712
Río Negro	19	20	915	954
Salta	26	18	876	920
San Juan	40	26	1127	1193
San Luis	30	24	762	816
Santa Cruz	6	4	340	350
Santa Fe	595	802	10263	11660
Santiago del Estero	9	4	558	571
Tierra del Fuego	8	6	258	272
Tucumán	46	25	1185	1256
Total	9739	5277	102441	117457

Figure 1: Argentinean manufacture activities

Activity $j=18$

Figure 2: Specialized agglomerations for two argentinean manufacture activities

The results from the analysis of the two activities are provided in Figure 2, under a specification of the parameters of the algorithm as given in Appendix. On the left-hand side is given a map of Argentina with the over-specialized agglomerations in dark and on the right-hand side some figures characterizing the numerical aspects of the algorithm. East-side of the maps of the country there is a map of the region around the city of Buenos Aires (the so-called GBA: Great Buenos Aires) at a lower scale than the maps of the country. Similarly, on the west side, lower scale maps are given for the regions around three cities, namely (from north to south) Tucumán, San Juan and Mendoza.

Furthermore, Figure 2 also orders, through a color scale, the agglomerations according to their contributions to the test statistics and Tables 1 and 2 provide further numerical evidences on this issue. These tables make explicit that, in the test statistic (27), the factor $N_{m,j|C}$ multiplying the log of the location quotient induces a re-ordering of the contributions of the agglomerations, as compared with the order of the location quotient alone; this is indeed one aspect of this algorithm for taking the issue of the small areas into account.

The best specialized agglomeration scheme shows that in GBA only 1 agglomeration with 5 regions is over-specialized in activity 29 but 3 agglomerations (two with 1 region and one with 8 regions) are over-specialized in activity 18. For the country in general the numbers of over-specialized agglomerations in activity 18 and in activity 29 are roughly the same, namely 6 and 7, but their spatial distribution is quite different. Indeed, activity 18 is mostly concentrated in the center of the country in 24 regions of the GBA and in the rest of the province of Buenos Aires whereas activity 29 is more spread over the country in 33 regions, located in the GBA, in the rest of the province of Buenos Aires and in the provinces of Santa Fe, Córdoba and Mendoza. Moreover, the lines under the Tables 1 and 2 show that the over-specialized agglomerations of activity 18 gathers 67% of firms of the whole activity 18 at a national level whereas activity 29 gathers 61% of firms of the whole activity 29. These results show that in both activities the firms have a high tendency to co-localize in specialized agglomerations.

Table 1: *Specialized Agglomeration: Optimal Cluster Scheme of Activity 18 (Here $C_{18}^* = C$)*

$\mathcal{I}_{m C} (M(C) = 6)$	$N_{m,18 C}$	$N_{m \cdot C}$	$LQ_{\mathcal{I}_{m C}}^{18}$	$N_{m,18 C} \ln \left(LQ_{\mathcal{I}_{m C}}^{18} \right)$
$\mathcal{I}_{1 C}$	6101	44371	1.66	3085.89
$\mathcal{I}_{2 C}$	253	1440	2.12	189.98
$\mathcal{I}_{3 C}$	46	80	6.93	89.08
$\mathcal{I}_{4 C}$	27	177	1.84	16.46
$\mathcal{I}_{5 C}$	118	1269	1.12	13.53
$\mathcal{I}_{6 C}$	23	234	1.19	3.91

$$\sum_{1 \leq m \leq 6} N_{m,18|C} = 6568 \text{ and } N_{\cdot 18} = 9739 \text{ (percentage 67\%)}$$

It is of interest to compare Figures 1 and 2. Indeed Figure 1 gives a picture of the agglomerations

Table 2: *Specialized Agglomeration: Optimal Cluster Scheme of Activity 29 (Here $C_{29}^* = C$)*

$\mathcal{I}_{m C} (M(C) = 7)$	$N_{m,29 C}$	$N_{m \cdot C}$	$LQ_{\mathcal{I}_{m C}}^{29}$	$N_{m,29 C} \ln(LQ_{\mathcal{I}_{m C}}^{29})$
$\mathcal{I}_{1 C}$	589	5182	2.53	546.71
$\mathcal{I}_{2 C}$	2103	43653	1.07	146.81
$\mathcal{I}_{3 C}$	283	4942	1.27	68.67
$\mathcal{I}_{4 C}$	134	1854	1.61	63.71
$\mathcal{I}_{5 C}$	67	999	1.49	26.84
$\mathcal{I}_{6 C}$	24	184	2.90	25.58
$\mathcal{I}_{7 C}$	8	44	4.05	11.18

$$\sum_{1 \leq m \leq 7} N_{m,29|C} = 3208 \text{ and } N_{\cdot,29} = 5277 \text{ (percentage 61\%)}$$

in terms of the *absolute* concentration of the activities whereas Figure 2 provides a similar picture in terms of the *relative* concentration of the activities.

This greedy algorithm is aimed at detecting specialized agglomerations the actual meaning of which depends, among other issues, on some characteristics of the numerical behavior of the algorithm. On the right-hand side of each map of Figure 2 are shown four figures illustrating, from North to South, the following aspects:

1. The evolution of the objective function, namely the BIC, at each step of the algorithm with a mark for its maximum value (vertical line). In both cases the curve is roughly unimodal although for activity 29 there are some local maxima. The maximum value of the objective function was found at step 48 (value 3212) and at step 35 (value 565) for activities 18 and 29 respectively.
2. The control for local *vs.* global maxima is operated by means of a window of $h = 60$ steps: from the 61-th step on, a least-squares slope is evaluated at each step for the last 60 steps and the algorithm is stopped once the slope change the sign (and therefore crosses the zero value). Once the algorithm is stopped, the maximum is located within the last 60 values of the objective function. The second figure provides the trajectory of the slope, starting from the 61-th step. The change of sign of the slope was obtained in steps 79 and 71 for activities 18 and 29 respectively.
3. As mentioned in Section 2.5, an important issue for interpreting the detected agglomerations is to evaluate how far the contiguity among regions is a significant factor of clustering. This issue is faced by means of a bootstrap test of invariance for the group of permutations of the coordinates. The third figure provides an histogram of the results of $B = 1000$ bootstrap replications with a mark (vertical line) corresponding to the maximum obtained at the optimizing phase. These simulations show that the contiguity is a significant factor of the clustering for

the activity 29 and not significant for activity 18. This suggests that the intra-regional localization economies of agglomeration have a major presence in activity 18 and consequently gives less weight to the contiguity whereas the inter-regional spread has more impact in activity 29.

4. Finally, the fourth figure presents the same data as the third one on a cumulative form under the scale of a standard normal distribution. These last two figures reveal that the distributions of the bootstrap replications is clearly unimodal, bell-shaped and roughly normal without noticeable pathologies in the tails.

4 Concluding remarks

This paper is focused on a particular problem of spatial economics, namely the detection of specialized agglomerations emerging from lattice data. A significant feature of this new algorithm is its ability to take due account of the small areas problem, by weighting the log of the location quotient in such a way that the pathology of this one is faced. The literature cited in the Introduction makes clear that the proposed structural model and the algorithmic detection have a potentially wide field of applications in spatial statistics.

The role of contiguity in the formation of specialized agglomerations is multifaceted and has been sketched in Section 2.5. The application to Argentinean data suggests that it may be different for different activities and, more specifically, allows one to detect activities with a major role of contiguity, activity 29 in this case, suggesting a higher impact of inter-regional localizations economies and those with a minor role, activity 18 in this case.

The bootstrap test of invariance for coordinates permutation therefore delivers an information complementary to the information provided by the optimization result and permits to detect different types of specialized agglomerations. This is also an issue on the very concept of agglomeration. From a narrow spatial approach, one could consider that the contiguity is so an essential ingredient that a region may be incorporated in an agglomeration even if its relative specialization is different from that of the agglomeration and this incorporation is left to be decided by the optimizing process.

But if the contiguity is of a different nature, one might require that each region within a given agglomeration should display a same sign for the *log* of its Location Quotient. This is indeed a flexibility of the algorithm one parameter of which, namely “sign”, specifies whether the regions entering an agglomeration should, or should not, have a same sign of *log LQ*. Another flexibility of the algorithm allows the choice between different types of the distance underlying the condition of contiguity. (see Appendix).

Acknowledgements

Michel Mouchart gratefully acknowledges financial support from IAP research network grant *nrP6/03* of the Belgian government (Belgian Science Policy). Both authors gratefully acknowledge the financial support of CIDETI, that promoted intercontinental cooperation. The help of Cristian S. Rocha, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (Arg.) has been instrumental in shaping the algorithm developed in this paper and is deeply acknowledged. A special thank is due to Vicente Donato for the impetus he gave to the development of the topic of this paper. The authors are also grateful to Dominique Peeters (UCL) for several clarifying discussions during the elaboration of this paper and to Paul Wilson (Clemson University) who helped to substantially improve the presentation of an earlier version of this paper. The authors are also indebted to two anonymous referees whose comments have been instrumental in helping the authors to improve the quality of their exposition.

References

- AKAIKE, H. (1974), A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716-723.
- ANAS, A., ARNOTT, R. AND SMALL, K. (1998), Urban spatial structure. *Journal of Economic Literature* **36**: 1426-1464.
- ANSELIN, L. (1995), Local indicator of spatial association - LISA. *Geographical Analysis* **27**: 93-115.
- ARBIA, G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Dordrecht: Kluwer.
- ARBIA, G. (2001), The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems* **3**: 271-281.
- ARBIA, G. AND ESPA, G. (1996), *Statistica Economica Territoriale*. Padova: CEDAM.
- ARBIA, G., ESPA, G. AND QUAH, D. (2008), A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. *Empirical Economics* **34**: 81-103.
- ARROW, K. (1962), The economic implications of learning by doing. *Review of Economic Studies* **29**: 155-73.
- BECATTINI, G. (1979), Dal “settore” industriale al “distretto” industriale. Alcune considerazioni sull’unità di indagine dell’economia industriale. *Rivista di Economia e Politica Industriale* **1**: 7-21.
- BESAG, J. AND NEWELL, J. (1991), The detection of clusters in rare diseases. *Journal of the Royal Statistical Society* **154**: 143-155.
- BURNHAM, K. AND ANDERSON, D. (2002), *Model Selection and Multi-Model Inference*. New York: Springer-Verlag Inc.
- BURNHAM, K. AND ANDERSON, D. (2004), Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods and Research* **33**: 261-304.
- CHAUDHURI, P. AND MARRON, J. (1999), SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94**: 807-823.
- CHAUDHURI, P. AND MARRON, J. (2000), Scale space view of curve estimation. *The Annals of Statistics* **28**: 408-428.

- CLAESKENS, G. AND HJORT, N. (2008), *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.
- CRESSIE, N. (1977), On some properties of the scan statistic on the circle and the line. *Journal of Applied Probability* **14**: 272-283.
- DIGGLE, P. (2000), Overview of statistical methods for disease mapping and its relationship to cluster detection. In P. Elliott, J.C. Wakefield, N.G. Best and D.G. Briggs (eds.), *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press.
- DIGGLE, P. (2003), *Statistical Analysis of Spatial Point Patterns*. London: Edward Arnold.
- DONATO, V. (2002), Políticas públicas y localización industrial en Argentina. CIDETI Working Paper 2002/01, Alma Mater Studiorum-Università di Bologna in the Argentine Republic.
- DONATO, V. AND HAEDO, C. (2002), *La Nueva Geografía Industrial Argentina*. Buenos Aires: Bononiae Libris and Unión Industrial Argentina.
- DURANTON, G., AND OVERMAN, H. (2005), Testing for localization using micro-geographic data. *Review of Economic Studies* **72**: 1077-1106.
- ESTER, M., KRIEGEL, H-P, SANDER, J. AND XU, X. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han and U. Fayyad, *Proceedings of the Second International Conference on Knowledge Discovery and Data*. Menlo Park: AAAI Press.
- ESTER, M., KRIEGEL, H-P., SANDER, J. AND XU, X. (1998), Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery* **2**: 169-194.
- FLORENCE, P. (1939), *Report of the location of industry*. Political and Economic Planning, London, UK.
- GLAESER, E., KALLAL, H., SCHEINKMAN, J. AND SHLEIFER, A. (1992), Growth in cities. *Journal of Political Economy* **100**: 1126-1152.
- GLAZ, J., NAUS, J. AND WALLENSTEIN, S. (2001), *Scan Statistics*. New York: Springer-Verlag.
- GLAZ, J., POZDNYAKOV, V. AND WALLENSTEIN, S. (2009), *Scan Statistics: Methods and Applications*. Boston: Birkhauser Publishers.
- GÓMEZ-RUBIO, V., FERRÁNDIZ-FERRAGUD, J. AND LÓPEZ-QUÍLEZ, A. (2005), Detecting clusters of disease with R. *Journal of Geographical Systems* **7**: 189-206.

- GOOD, P. (2000), *Permutation Tests- A Practical Guide to Resampling Methods for Testing Hypotheses*. New York: Springer-Verlag.
- GUIMARÃES, P., FIGUEIREDO, O. AND WOODWARD, D. (2003), A tractable approach to the firm location decision problem. *Review of Economics and Statistics* **84**: 201-204.
- GUIMARÃES, P., FIGUEIREDO, O. AND WOODWARD, D. (2009), Dartboard tests for the location quotient. *Regional Science and Urban Economics* **39**: 360-364.
- HAEDO, C. (2009), Measure of Global Specialization and Spatial Clustering for the Identification of “Specialized” Agglomeration. Ph.D. thesis, Bologna: Dipartimento di Scienze Statistiche “P. Fortunati”, Università di Bologna (IT). http://amsdottorato.cib.unibo.it/1735/1/Christian_Haedo_tesi.pdf
- HAEDO, C. AND MOUCHART, M. (2012), A stochastic independence approach for different measures of concentration and specialization. CORE Discussion Paper 2012/25, UCL Louvain-la-Neuve (B). http://www.uclouvain.be/cps/ucl/doc/core/documents/coredp2012_25web.pdf
- HENDERSON, J.V. (1985), *Economic Theory and the Cities*. Orlando: Academic Press.
- JACOBS, J. (1969), *The Economy of Cities*. London: Jonathan Cape.
- KRUGMAN, P. (1991), Increasing returns and economic geography. *Journal of Political Economy* **99**: 483-499.
- KULLDORFF, M. (1997), A spatial scan statistic. *Communications in Statistics-Theory and Methods* **26**: 1481-1496. 799-810.
- KULLDORFF, M. AND NAGARWALLA, N. (1995), Spatial disease cluster: detection and inference. *Statistics in Medicine* **14**: 799-810.
- KULLDORFF, M., HUANG, L., PICKLE, L. AND DUCZMAL, L. (2006), An elliptic spatial scan statistic. *Statistics in medicine* **25**: 3929-3943.
- KULLDORFF, M., TANGO, T. AND PARK, P. (2003), Power comparisons for disease clustering tests *Computational Statistics and Data Analysis* **42**: 665-684.
- LAWSON, A. (2006), *Statistical Methods in Spatial Epidemiology*. New York: Wiley.
- LAWSON, A., BIGGERI, A., BÖHNING, D., LESAFFRE, E., VIEL, J-F. AND BERTOLLINI, R. (eds.) (1999), *Disease Mapping and Risk Assessment for Public Health*. Chichester: John Wiley & Sons.

- LAWSON, A., AND DENISON, D. (eds.) (2002), *Spatial Cluster Modelling*. London: Champan & Hall\CRC.
- LOH, J. (2011), K-scan for anomaly detection in disease surveillance. *Environmetrics* **22**: 179-191.
- LUCAS, R. (1988), On the mechanics of economic development. *Journal of Monetary Economics* **22**: 3-42.
- MANLY, B. (1991), *Randomization and Monte Carlo methods in biology*. London: Chapman & Hall\CRC.
- MARSHALL, A. (1890), *Principles of Economics*. London: Macmillan.
- MOINEDDIN, R., BEYENE, J. AND BOYLE, E. (2003). On the location quotient confidence interval. *Geographical Analysis* **35**: 249-256.
- MORI, T. AND SMITH, T. (2011), An industrial agglomeration approach to central place and city size regularities. *Journal of Regional Science* **51**: 694-731.
- MOUCHART, M., RUSSO, F. AND WUNSCH, G. (2009), Structural modelling, exogeneity, and causality. In H. Engelhardt, H-P. Kohler and A. Fürnkranz-Prskawetz (eds.), *Causal Analysis in Population Studies: Concepts, Methods, Applications*. Dordrecht: Springer.
- NATHAN, M. AND OVERMAN, H. (2013), Agglomeration, clusters, and industrial policy. *Oxford Review of Economic Policy* **29**: 383-404.
- NAUS, J. (1965), Clustering of random points in two dimensions. *Biometrika* **52**: 263-267.
- NEILL, D. (2006), Detection of Spatial and Spatio-Temporal Clusters. Ph.D. thesis, Pittsburgh: School of Computer Science, Carnegie Mellon University (US).
- NEILL, D., MOORE, A. AND COOPER, G. (2006), A Bayesian spatial scan statistic. *Advances in Neural Information Processing Systems* **18**: 1003-1010.
- O'DONOGHUE, D. AND GLEAVE, B. (2004), A note on methods for measuring industrial agglomeration. *Regional Studies* **38**: 419-427.
- OPENSHAW, S., CRAFT, A.W., CHARLTON, M. AND BIRCH, J.M. (1988), Investigation of leukaemia cluster by use of a geographical analysis machine. *Lancet* **1**: 272-273.
- O'SULLIVAN, D. AND UNWIN, D.J. (2010), *Geographic Information Analysis*. New Jersey: John Wiley & Sons.
- ROMER, P. (1986), Increasing returns and long-run growth. *Journal of Political Economy* **64**: 1002-1037.

- SAINI, G. AND KAUR, K. (2014), Regionalization as spatial data mining problem based on clustering: review. *International Journal of Computer Engineering and Applications* **6**: 163-173.
- SCHWARTZ, G. (1978), Estimating the dimension of a model. *Annals of Statistics* **6**: 461-464
- SCITOVSKY, T. (1954), Two concepts of external economies. *Journal of Political Economy* **62**: 143-151.
- TANGO, T. (2010), *Statistical Methods for Disease Clustering*. New York: Springer.
- TANGO, T. AND TAKAHASHI, K. (2005), A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* **4** (11).
- WAKEFIELD, J., KELSALL, J. AND MORRIS, S. (2000), Clustering, cluster detection and spatial variation in risk. In P. Elliot, J. Wakefield, N. Best and D. Briggs (eds.), *Spatial Epidemiology. Methods and Applications*. Oxford: Oxford University Press.
- YANG, Y. (2005), Can the strengths of AIC and BIC be shared? *Biometrika* **92**: 937-950.
- ZOELLNER, I. AND SCHMIDTMANN, I. (1999), Empirical studies of cluster detection: different cluster tests in application to German cancer maps. In A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre and J-F. Viel (eds.), *Disease Mapping and Risk Assessment for Public Health*. Chichester: John Wiley.

Appendix: The main parameters of the algorithm

The algorithm treats each activity j independently and requires the specification of several parameters. The values used in the application on Argentinean data are identified by: *Arg*. The main parameters are:

- contiguity matrix: simple first order queen or rook matrix; see Remark on the W matrix in Section 2.4) (*Arg*: rook);
- selection of the regions i :
 - either according to N_{ij} : either all regions, or only those with $N_{ij} > 0$, or only those with $N_{ij} >$ some fixed quote; ignoring the regions where $N_{ij} = 0$ actually accelerates the processing of the algorithm (*Arg*: $N_{ij} > 4$);
 - or according to the sign of $\log LQ$: all, only this with $\log LQ > 0$ (over-specialized) or only those with $\log LQ < 0$ (under-specialized), possibly with 0 replaced by $+or - \varepsilon$.

Note: both criterion may be combined or not;

- specification of the total number of primary units to be taken into account in the evaluation of the criterion: either $N_{..}$ (coded as δ TRUE), or the sum of the N_{ij} corresponding to the regions actually selected in the previous step (coded as δ FALSE). The case “FALSE” actually ignores the unselected regions although present in the country (*Arg*: $\delta =$ TRUE);
- parameter “sign”: “sign” = TRUE when the agglomerations contain only regions with a same sign of the $\log LQ$, *i.e.* only over-specialized or only under-specialized regions, otherwise “sign” is FALSE (*Arg*: sign = TRUE; more explicitly: if $i_{[1]}^*$ and i_2 are aggregated into a same cluster, then the log of the location quotients $LQ_{i_{[1]}^*,j}$ and $LQ_{i_2,j}$ have the same sign; *i.e.* both regions $i_{[1]}^*$ and i_2 are either over-specialized or under-specialized).

Note: the parameter “sign” is activated at each step of the algorithm but if, in the second parameter, the regions have been selected on the sign of $\log LQ$, the parameter “sign” is always TRUE;

- number of bootstrap replications (*Arg*: B = 1000);
- criterion: BIC or AIC (*Arg*: BIC);
- two parameters for the stopping rule:
 - firstly a selection between stopping according to a change of sign in the trajectory of the criterion or according to the local slope of the trajectory of the criterion (*Arg*: change of sign);
 - secondly the width of the window within which is evaluated the change of sign or the slope of the criterion (*Arg*: h = 60).