Maria Michela Dickson, Giuseppe Espa, Diego Giuliani, Flavio Santi

Assessing Spatial Patterns of Firms Using Micro-Level Data

(doi: 10.14650/100115)

Scienze Regionali (ISSN 1720-3929)
Fascicolo 2, maggio-agosto 2021

Assessing Spatial Patterns of Firms Using Micro-Level Data

Maria Michela Dickson, Giuseppe Espa, Diego Giuliani, Flavio Santi

Abstract: Firm-level geo-referenced databases are becoming more and more a valuable resource in the hand of regional scientists. This paper aims at providing an overview of the main empirical approaches, based on spatial point pattern statistics, that allow to properly use this kind of data to analyze the distribution of economic agents on the geographical space. Both methodological and applied aspects are covered.
Keywords: micro-geographic data, spatial patterns of firms, spatial statistics.
JEL classification: R12, R30, C49.

1. Introduction

Over the past fifteen years, there has been a marked increase in the availability of geo-referenced firm-level data. The *Italian Statistical Archive of Active Enterprises* (ASIA), the *Amadeus* database of *Bureau Van Dijck* and the *US Census Bureau's Longitudinal Business Database* (LBD) are just some examples of common firm-level databases that are now geo-referenced in almost their entirety. This important fact, together with the recent developments of technologies for the treatment of large amounts of data, has given rise to a promising stream of literature on the methodologies with which to analyse spatial patterns of economic agents (see, among many others, Duranton, Overman, 2005; Marcon, Puech, 2010; Bocci, Rocco, 2016; Arbia *et al.*, 2017; Cainelli, Ganau, 2018).

The use of micro-geographic data, instead of regional aggregates, has two main advantages. The first concerns the fact that the inferential conclusions drawn from the analyses of data aggregated at regional level are affected by

Maria Michela Dickson: Department of Economics and Management, University of Trento, Via Inama 5, 38122 Trento, Italy. E-mail: mariamichela.dickson@unitn.it
Giuseppe Espa: Department of Economics and Management, University of Trento, Via Inama 5, 38122 Trento, Italy. E-mail: giuseppe.espa@unitn.it
Diego Giuliani: Department of Economics and Management, University of Trento, Via Inama 5, 38122 Trento, Italy. E-mail: diego.giuliani@unitn.it, *corresponding author*
Flavio Santi: Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy. E-mail: flavio.santi@univr.it

ISSN 1720-3929

a statistical bias arising from the discretionally chosen partition of space (i.e. the so-called *modifiable areal unit problem* bias: see Arbia, 1989). The second advantage is that the theoretical models explaining spatial patterns of firms are generally grounded on the behaviour of the individual economic agent and hence can be tested empirically on regional aggregates only under the restrictive assumption of a homogeneous firm's behaviour within the region.

This paper purpose is to present some of the main empirical approaches proposed in the regional science literature to assess micro-spatial patterns of firms. It also illustrates their implementation in different paradigmatic real-case scenarios. Since all these approaches have been developed by borrowing, explicitly or implicitly, from the methodological framework of spatial point pattern statistics, the article begins by introducing this statistical background. There follows a review of the methods for the analysis of the spatial distribution of firms in the context of both homogeneous and inhomogeneous territories. Finally, the paper concludes with some remarks and suggestions for future developments in this field.

2. Basic concepts and definitions

Traditional spatial econometrics based on the Cliff-Ord methodological framework (Cliff, Ord, 1972) assumes that the spatial locations of statistical units are exogenously given. Hence it is not appropriate to analyze and model spatial patterns of economic agents using micro-geographic data. Indeed, while it is natural and reasonable to treat the location of a region as given, spatial locations of micro-geographic units cannot be considered as fixed because they are not naturally given but result from the behaviour of economic agents.

The natural statistical framework in which to analyse and model spatial patterns of economic agents using micro-geographic data is, instead, that of *spatial point pattern statistics*. This subbranch of the broader spatial statistics discipline is devoted essentially to identifying and summarizing the structure and characteristics of patterns formed by entities (such as business establishments) that are distributed in space. In the typical cases of practical interest, where the space is two-dimensional and finite, entities can be suitably represented by points on a planar map and hence can be identified by two spatial coordinates (say longitude and latitude). An observed set of spatial coordinates representing a micro-geographic distribution of interest is often called a *spatial point pattern*.

The statistical analysis of an observed spatial point pattern is generally based on the assumption that its spatial coordinates are the realization of an underlying stochastic mechanism called *spatial point process*. While a comprehensive and rigorous introduction to this kind of stochastic process falls outside the scope of this paper (see instead Diggle, 2003 or Møller, Waagepetersen, 2004), here it is sufficient to define a spatial point process simply as a stochastic mechanism that generates a set of points $x_i = (x_{1i}, x_{2i}) : i = 1, 2, \ldots$, where x_{1i} and x_{2i} represent, respectively, the longitudinal and latitudinal coordinates of the ith point (Diggle, 2003). As we shall show, point processes can be used to identify and study interesting characteristics of an observed spatial point pattern by making comparisons between the theoretical properties of an *a priori* specified underlying point process and the corresponding empirical complements estimated on the data.

The most important properties of a spatial point process are the so-called *first-order* and *second-order intensity functions*. Let us refer to the generic location x. If we consider that dx is an infinitesimally small spatial region around x, that $N(dx)$ denotes the number of points located in it, and that $|dx|$ represents its area, then the first-order intensity function can be defined as

$$\lambda(x) = \lim_{|dx| \to 0} \left\{ \frac{E[N(dx)]}{|dx|} \right\} \qquad [1]$$

(Diggle, 2003; Cressie, 1993).

Equation [1] can be interpreted as the expected number of points per unit observed in an infinitesimal region around the generic location x. Consequently, $\lambda(x)dx$ corresponds to the probability of finding a point in the close vicinity of x. If the first-order intensity is constant throughout the space, i.e. when $\lambda(x) = \lambda$, it corresponds to the expected number of points per unitary area and the point process is said to be *stationary* (Diggle, 2003).

Using the same notation as in Equation [1], the second-order intensity function can be defined as

$$\lambda_2(x, y) = \lim_{|dx|, |dy| \to 0} \left\{ \frac{E[N(dx)N(dy)]}{|dx||dy|} \right\} \qquad [2]$$

where x and y indicate two distinct generic locations (Diggle, 2003; Cressie, 1993). Intuitively, $\lambda_2(x, y)dxdy$ can be seen as the probability of observing two points within two infinitesimal regions around, respectively, the locations x and y (Diggle *et al.*, 2007). Equation [2] essentially describes the kind and level of spatial interactions among points. In many practical circumstances it is reasonable to assume that the underlying spatial point process is *isotropic*, which implies that $\lambda_2(x, y)$ depends exclusively on the distance s between locations x and y and not on their specific positions: that is $\lambda_2(x, y) = \lambda_2(s)$ (Diggle, 2003). To interpret the values of the second-order intensity function better, we can rely on a scaled version of Equation [2] that is called *pair correlation function* (Ripley, 1976, 1977), $g(s) = \lambda_2(s)/\lambda(x)\lambda(y)$. Indeed, if the process generates points in x or y independently from one another, then $g(s) = 1$. This value indicates absence of spatial interaction among points. On the other hand, if the process generates points in x or y with a probability that is higher (or lower) than if they were independently generated, then $g(s) > 1$ (or $g(s) < 1$). The former case describes a situation of spatial attraction of points while the latter one identifies spatial repulsion (or inhibition) between points.

If, in addition to being isotropic, the spatial point process is also stationary, say $\lambda(x) = \lambda$, then $g_{st}(s) = \lambda_2(s)/\lambda^2$ and, as a consequence, its second-order properties can be alternatively described by the so-called Ripley's *K*-function (Ripley 1976, 1977), which can be defined as

$$K(s) = 2\pi \int_0^s g_{st}(t)\, t dt \quad [3]$$

$K(s)$ is the cumulative version of $g_{st}(s)$, under stationarity, and provides the expected number of further points located *up to* a distance s from? a generic point (Ripley, 1977). In the economic analyses where the underlying data-generating point process is stationary and isotropic (that is, when the territory can be treated as homogeneous), the *K*-function measures the mean (general) level of spatial interactions among the economic agents (such as firms or households) up to each distance s. The *K*-function has an important advantage over the pair correlation function because it can be more straightforwardly estimated from an observed point pattern and, for many point processes, its theoretical form can be explicitly derived.

The approximately unbiased estimator of the *K*-function for a spatial point pattern with n points, observed in a study region A, is

$$\hat{K}(s) = \frac{|A|}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} w_{ij}^{-1} I(s_{ij} \leq s) \quad [4]$$

where $|A|$ is the total area of A, s_{ij} is the distance between the ith and jth observed points, and $I(s_{ij} \leq s)$ is the indicator function such that $I(s_{ij} \leq s) = 1$ if $s_{ij} \leq s$ and 0 otherwise (Diggle, 2003). In most practical cases is computed using the Euclidean distance criterion, but essentially any other definition of distance (such as travel time or economic distance) can be applied as well. The term w_{ij} represents an adjustment factor that is needed to reduce potential negative biases in the estimates for the points located close to the boundary of A. Indeed, points near the boundary may be close to unobserved points located outside A. Consequently, for these points, it may not be possible to count the actual number of further points located up to a distance s. Specifically, w_{ij} is equal to the proportion of the circumference of the circle centred on the ith point that passes through the jth point and lies within A. If s is much smaller than the extension of A, w_{ij} reduces the effects of boundary bias in the proper way.

3. Spatial location patterns of economic agents in a homogeneous space: the CSR test

For many underlying point processes that can be appropriate in modelling the spatial behaviour of firms, it is possible to write the *K*-function in a closed form. This allows the *K*-function to be used for the identification of the underlying point process by comparing its theoretical form with its empirical counterpart estimated on the observed point pattern of interest.

A useful preliminary analysis of a point pattern consists of verifying whether the data are consistent with the so-called hypothesis of *Complete Spatial Randomness* (CSR), which states, essentially, that the observed points have been generated independently of each other (that is, with no attraction or repulsion among them) and with the same probability of being positioned in any possible location of the study region. Therefore, a point pattern under CSR has no spatial structure, neither in the form of spatial inhomogeneity nor because of spatial dependence. Deviations from the CSR hypothesis thus reveal the presence of interesting spatial location patterns that may warrant further investigation.

Formally, a complete spatial random point pattern can be considered as a realization of the homogeneous Poisson process (Diggle, 2003), which generates point patterns containing independent random points, uniformly distributed in a finite region A according to a constant first-order intensity, $\lambda(s) = \lambda$. Therefore, assessing whether an observed spatial point pattern is consistent with the CSR hypothesis corresponds to verifying whether the observed points may have been generated by a homogeneous Poisson process. As shown by Ripley (1976), under this process, the theoretical form of the K-function is such that

$$K(s) = \pi s^2 \ , \ s > 0 \qquad [5]$$

Equation [5] then represents the null hypothesis of CSR and can be considered as a frame of reference within which to build a formal test. Indeed, significant deviations from this reference provide evidence in favour of the alternative hypothesis of either spatial inhomogeneity or spatial dependence, or both. In particular, $K(s) > \pi s^2$ indicates that, on average, there are more points within a distance s from any point than should be expected in the case of CSR. If the observed point pattern represents the spatial distribution of firms in a homogeneous space, this kind of deviation would imply the presence of positive spatial interactions among economic agents, and hence clustering. In contrast, $K(s) < \pi s^2$ means that, on average, there are fewer points within a distance s from any point than expected under CSR. This circumstance suggests, for example, the presence of negative spatial interactions among economic agents that make them locate at a distance s from each other.

In order to help the interpretation, $K(s)$ can be normalized using linear transformations (Besag, 1977), such as

$$L(s) = \sqrt{K(s)/\pi} - s$$

which makes the CSR hypothesis represented by 0 for all values of s.

Since the sampling distribution of $\hat{K}(s)$ under CSR can be derived theoretically only in limited cases, inference about the CSR hypothesis is generally made by assessing the sampling distribution of $\hat{K}(d)$ by Monte Carlo simulations of the homogeneous Poisson process conditional on the observed data. As illustrated by Besag and Diggle (1977), it is indeed possible to obtain significance envelopes for the CSR hypothesis by, first of all, simulating m point patterns from the homogeneous Poisson process with a fixed number of points equal to the size of the observed data. Then, for any of the m generated point patterns, a different $\hat{K}(s)$ (or $\hat{L}(s)$) can be computed. Finally, the $1/(m + 1) \times 100\%$ significance envelopes are obtained from the highest and lowest values of the $\hat{K}(s)$'s (or $\hat{L}(s)$'s) functions computed from the m simulated point patterns. The graph of the empirical function estimated on the data and the related significance envelopes against s makes it possible to draw conclusions about the CSR hypothesis. In particular, deviations of the empirical curve of $\hat{K}(s)$ (or $\hat{L}(s)$) outside – upward or downward – the significance envelopes at some distances s imply significant departures from complete spatial randomness at those distances.

In order to illustrate this inferential procedure, Figure 1 shows the performance of the K-function-based CSR test in different stylized micro-geographic distributions of firms. Specifically, the graphs in the figure portray the behaviour of $\hat{L}(s)$ against the distance s. In each of them, the solid curve represents the empirical function and the shaded area indicates the non-rejection region provided by the corresponding upper and lower significance envelopes for 999 realizations of a homogeneous Poisson process. It is interesting to note that this kind of test can identify significant deviations from CSR (toward spatial concentration or repulsion) at varying spatial scales at once. An important advantage of this feature is that different, and even opposite, spatial patterns taking place in the same area can be detected simultaneously. For instance, the point pattern represented in Case (v) exhibits spatial aggregation at small distances and repulsion at relatively higher distances in the form of small clusters of firms located far apart from each other. By contrast, Case (iv) is characterized by repulsion at small distances and aggregation at relatively higher distances in the form of a large-scale concentration of firms in the left part of the area within which they actually tend to be located while distancing each other.

Regional science applications of the K-function-based CSR test have concerned the assessment of spatial clustering of economic activities in different geographical and industrial contexts; see, among others, Barff (1987), Arbia (1989), Marcon and Puech (2003), Ó hUallacháin and Leslie (2007).

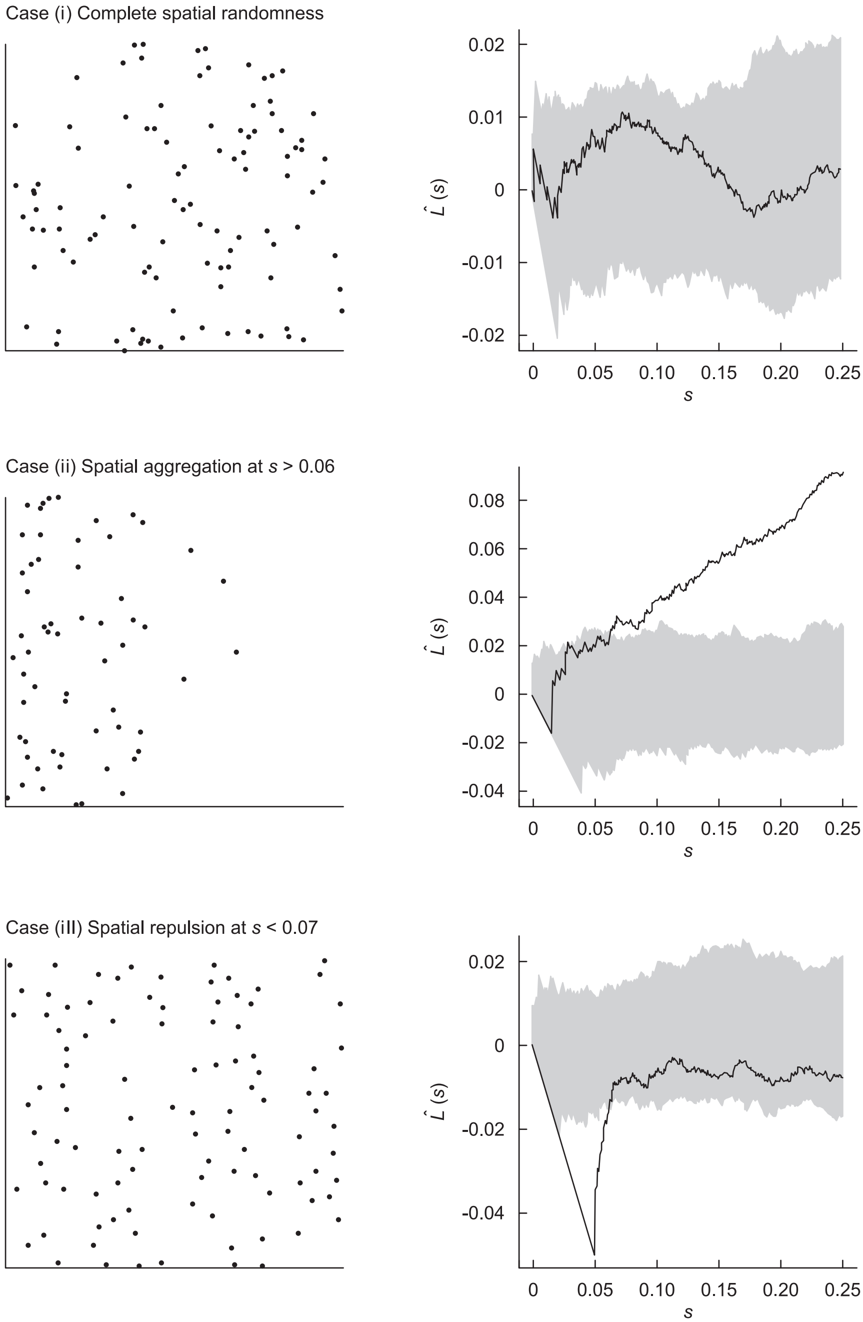

Figure 1: Paradigmatic examples of the application of the *K*-function-based CSR test.
Source: Authors' elaboration.

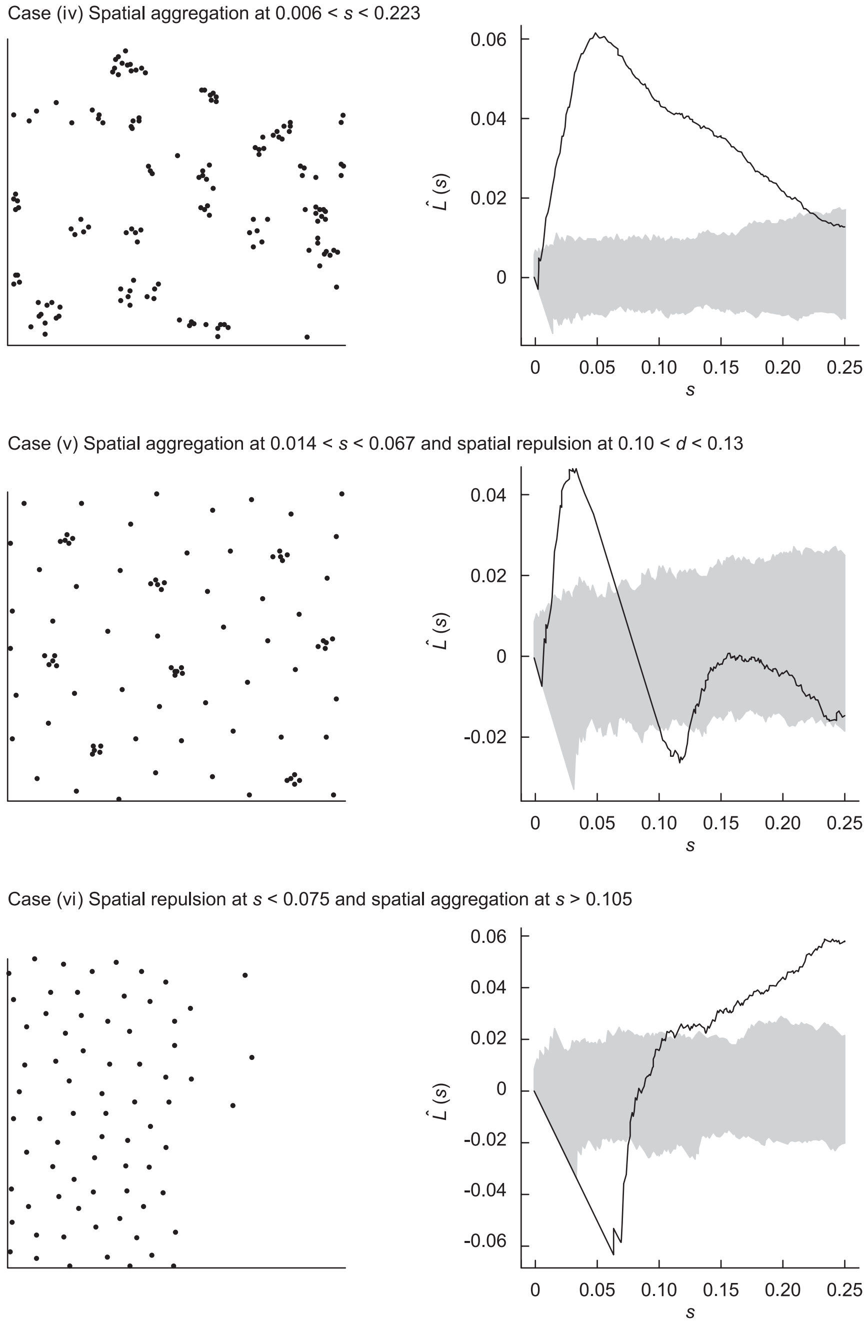

Figure 1: Continue.

4. Spatial location patterns of economic agents in an inhomogeneous space: the *relative* approaches

The CSR test makes it possible to detect the presence of positive or negative spatial interaction among firms only under the assumption that the underlying point process is stationary, which implies that the space in which economic agents operate is essentially homogeneous. In practice, however, the space is seldom homogenous because of the action of exogenous factors that may exclude, limit or expand the possibility to establish firms in certain locations. For example, firms may not locate in some areas due to the occurrence of lawful and geophysical restrictions, or they may locate in certain areas due to the closeness to useful infrastructures, communication routes or because of favourable local taxation. Both the spatial statistics and applied regional science literatures have developed approaches to deal with the relaxation of the assumption of spatial homogeneity. They can be essentially divided into two different kinds: *relative* and *absolute* approaches. The former make it possible to control for the presence of spatial inhomogeneity by comparing the observed point pattern of interest with another observed point pattern that is potentially affected by the same sources of spatial heterogeneity. In a different way, the latter consider the direct estimation of the varying first-order intensity function. The main relative approaches are probably those based on the Diggle and Chetwynd (1991)'s *D*-function, Duranton and Overman (2005)'s *K*-density and Marcon and Puech (2010)'s *M* function. While the most popular absolute method is undoubtedly the one based on the use of the Baddeley *et al.* (2000)'s K_{inhom}-function.

4.1. Diggle and Chetwynd's *D*-function method

Diggle and Chetwynd (1991) developed an approach that makes it possible to relax the assumption of spatial homogeneity by making use of the *K*-function within a case-control framework. This approach can be applied in those circumstances in which it is reasonable and meaningful to split the observed point pattern into a set of *cases* and a set of *controls*. The cases should be the points of interest, such as the firms of a given sector or the firms that exit the market in a certain time period. The controls should constitute a reference set of points located in the same space, such as the firms of all other sectors of the economy or the firms that still operate on the market. According to Diggle and Chetwynd (1991), if cases and controls are affected by the same unobserved exogenous factors of spatial heterogeneity, it is possible to assess actual positive (or negative) spatial dependence among cases if they are more (or less) spatially aggregated than the controls. With the aim of measuring the extra-aggregation (or extra-spreading) of cases with respect to controls, Diggle and Chetwynd (1991) suggested using the following statistic:

$$D(s) = K_{cases}(s) - K_{controls}(s) \quad [6]$$

where $K_{cases}(s)$ and $K_{controls}(s)$ indicate the K-function for the cases and controls respectively. The rationale behind this framework is that the point pattern of controls can properly describe the underlying common varying first-order intensity. In this respect, $D(s) = 0$ suitably represents the null hypothesis of no spatial dependence of cases within an inhomogeneous space. As a consequence, $D(s) > 0$ implies that the cases are relatively more spatially aggregated than the controls, while $D(s) < 0$ means that the cases are relatively less spatially aggregated than the controls. The former circumstance indicates the presence of actual positive spatial interaction, and hence clustering, among cases; the latter one points towards negative spatial interaction, and hence repulsion, among cases.

$D(s)$ can naturally be estimated using $\hat{D}(s) = \hat{K}_{cases}(s) - \hat{K}_{controls}(s)$, with $\hat{K}_{cases}(s)$ and $\hat{K}_{controls}(s)$ obtained through Equation [4], for the observed sets of cases and controls respectively. To determine whether $\hat{D}(s)$ is significantly different from zero and hence test the null hypothesis of no *relative* spatial interaction, Diggle and Chetwynd (1991) considered the so-called hypothesis of *random labelling*, formalized by Cuzick and Edwards (1990). This hypothesis states that the observed points of the pattern are randomly «labelled» as cases or controls and hence it is logically equivalent to the hypothesis of absence of spatial interaction among cases. Therefore, $D(s) = 0$ can be properly tested using a Monte Carlo procedure that simulates m random point patterns, in each of which the observed «labels» of cases and controls are randomly permuted among the observed points. For each of the m randomly labelled patterns, a different $\hat{D}(s)$ can be computed. The $1/(m + 1) \times 100\%$ significance envelopes for the null hypothesis of no spatial interaction among cases can then be obtained from the highest and lowest values among the resulting m $\hat{D}(s)$'s functions.

To illustrate the D-function-based approach to detecting relative spatial interaction of firms within an inhomogeneous space, we refer to an example concerning the spatial distribution of firms exiting the lodging industry in the main island of Sicily (Italy). Specifically, we focus on the 164 tourist accommodation services that entered the market in 2010 and that may have ceased to operate during the following period 2011-2015. The lodging industry is here composed of the following types of service: hotels, resorts, youth hostels, mountain refuges, holiday homes, farm stays, campgrounds and other short-stay accommodations[1]. Data were made available by the Italian National Institute of Statistics (ISTAT).

Figure 2 displays the map of the spatial distribution of the 164 lodging sector firms. The map also highlights the distinction between the 109 firms

[1] We consider the following NACE codes: I55100, I55201, I55202, I55203, I55204, I55205, I55300, I55902.

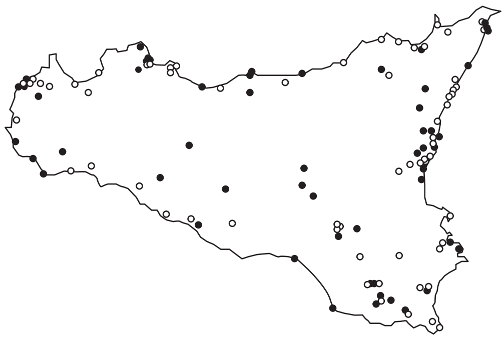

Figure 2: Locations of the 164 lodging sector firms that started to operate in 2010 in main Island of Sicily (Italy). Empty circles represent the 109 firms that survived between 2011 and 2015. Solid circles represent the 55 firms that ceased to operate between 2011 and 2015.
Source: Authors' elaboration.

that stayed in the market for the entire period 2011-2015 (empty circles) and the 55 firms that exited the market before the end of the same period (solid circles). Visual inspection of the map suggests that the space is very inhomogeneous since the lodging sector firms appear to be mostly located near the seaside. This implies that it is not realistic to consider the first-order intensity function as a constant.

With these data it may be interesting to detect the occurrence of some spatial pattern in the firm exit process, such as a contagion effect or local co-opetitive relationships. Due to the apparent inhomogeneity of the given territory, the *K*-function should not be employed to detect spatial dependence among firms' exits because it may lead to spurious results in which spatial interactions are confounded with spatial inhomogeneity. Since the location choices of firms which did and did not survive are likely to have concerned the same tourist destinations, we can properly detect significant spatial dependence among firm exits by verifying if firms that ceased to operate were more or less spatially aggregated than the firms surviving for the entire observed period. It is therefore possible to apply the *D*-function-based test by labelling the firms that exited from market as «cases» and the firms that stayed in the market as «controls».

The result of the test is shown in Figure 3 and reveals that, at small distances (under 9 km), firms that ceased to operate were less spatially aggregated then those which instead survived, thus indicating that firms' exits in the lodging sector in Sicily tend to occur at a certain distance from the other firms' exits. This evidence points towards a phenomenon of spatial local competition, instead of a contagion effect, where the failure of a lodging firm positively affects the survival of the others lodging firms located within 9 km.

The *D*-function-based test, as it has been presented here or in slightly modified versions, has been used in various regional science studies. For

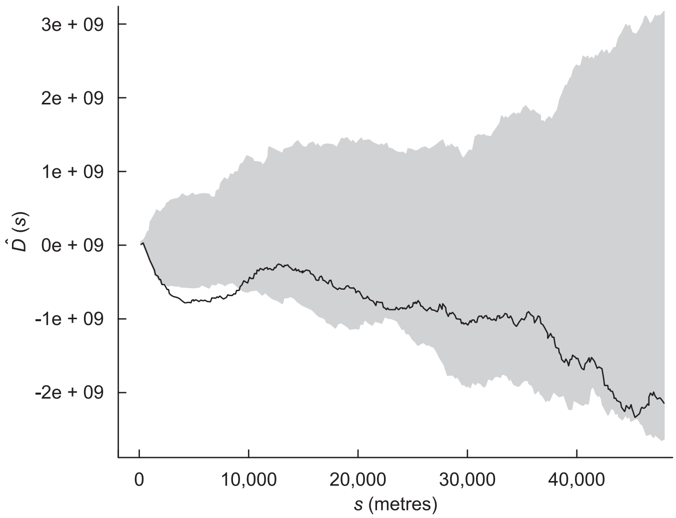

Figure 3: Empirical D-function (continuous line) and the corresponding 99.9% significance bands (shaded area) for the firms' exits in the lodging sector in Sicily (Italy), 2011-2015.
Source: Authors' elaboration.

instance, Sweeney and Feser (1998) used this inferential framework to assess how the spatial distribution of firms in North Carolina depends on their size; Feser and Sweeney (2000) focused on the spatial concentration of industrial linkages; Marcon and Puech (2003) measured the level of geographic concentration of industries in France; Arbia *et al.* (2008) assessed the occurrence of knowledge spillovers through the analysis of the spatial pattern of patents and Kosfeld *et al.* (2011) studied the conditional concentration of industries in Germany.

4.2. Duranton-Overman's K-density and Marcon-Puech's M function methods

An important stream of literature in the field of economic geography has focused on developing measures of geographic concentration of industries for geo-referenced firm-level data. These measures represent an attempt to overcome the methodological limits of the more traditional indices, such as the popular Location Quotient or Ellison-Glaeser Index (Ellison, Glaeser, 1997), which make use of regional level data. The most established and frequently employed amongst these measures are perhaps the *K*-density and the *M* function, developed by Duranton and Overman (2005) and Marcon and Puech (2010), respectively.

Duranton and Overman (2005)'s *K*-density is logically related to the pair correlation function, $g(s) = \lambda_2(s)/\lambda(x)\lambda(y)$, in that it can be regarded as an estimator of the probability density function of finding a point located at a given distance s from a generic point. Indeed, for an observed point pattern,

K-density calculates the average number of pairs of points at each distance s and then normalizes by means of some smoothing operation so as to obtain a continuous function that sums to 1. Formally, the K-density for a point pattern of n firms is defined as:

$$\hat{K}_{density}(s) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} k(s_{ij}, s; h)$$

where

$$k(s_{ij}, s, h) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(s_{ij} - s)^2}{2h^2}\right)$$

is a Gaussian kernel function *à la* Silverman (1986) with standard deviation h, which represents the smoother bandwidth. According to Duranton and Overman (2005) the optimal value of the h parameter can be set as per Equation [3.31] of Section 3.4.2 of Silverman (1986).

In order to assess the significant deviations $\hat{K}_{density}$ of from the null hypothesis of no spatial concentration of firms, Duranton and Overman (2005) suggested using significance envelopes obtained by randomly reassigning the firms among their locations. In particular, they suggest running a number of Monte Carlo simulations, in each of which the firms of a single industry (or of any given interesting typology) are randomly allocated among the locations of all firms of all industries (or of all the given typologies). In this setting, the null hypothesis of no spatial concentration of the firms of the same industry (or the same typology) is provided by the centre of the Monte Carlo simulated significance envelopes. Therefore, like the D-function, K-density can detect relative spatial concentration of firms of a specific typology. Duranton and Overman (2005) show that K-density can be weighted to also deal with the size of firms, as it may proxied, for example, by the value-added, the number of employees or the capital. See Duranton and Overman (2005) for further details.

In the past ten years K-density has been widely used to measure the spatial concentration of single industries. See, among others, Duranton and Overman (2008), Klier and McMillen (2008), Vitali *et al.* (2013), Koh and Riedel (2014), Kerr and Kominers (2015) and Behrens and Bougna (2015). Antonietti *et al.* (2013) used an approach similar to that of K-density to investigate the relationship between co-localization and vertical disintegration of the firms belonging to the knowledge intensive business service sector in the metropolitan area of Milan. Interestingly, Coniglio *et al.* (2018) have applied K-density to a non-geographical space; in particular, while referring to the level of relatedness between products in the product-space, they studied the export baskets of Italian provinces.

Marcon and Puech (2010)'s M function can be regarded as the cumulative counterpart of K-density, as it is logically related to the cumulative, instead of the density, distribution of the distances between pairs of points in a point pattern of firms. It has been developed as a micro-geographic data-based version of the popular Location Quotient of industrial specialization, which is based on regional data. The *M* function is indeed an empirical function of the distance *s* that gives the proportion of firms of a given typology (e.g. a given industry) that are located within distance *s* from a generic firm of the same typology divided by the same proportion computed with respect to all firms of all typologies (e.g. of all industries). Considering an observed point pattern of *n* firms belonging to different typologies, the *M* function can be formally defined as

$$\hat{M}(s) = \sum_{i=1}^{n} \frac{\sum_{j_c \neq i} I(s_{ij^c} \leq s) w_{j^c}}{\sum_{j \neq i} I(s_{ij} \leq s) w_j} \Bigg/ \sum_{i=1}^{n} \frac{W_c - w_{j^c}}{W - w_j}$$

where j^c is another firm of the same typology of firm i, while j denotes another firm of any typology. The terms and represent weights associated with firms j and j^c, respectively; while W_c and W represent the total weight of firms of the same typology of j^c and the total weight of all n firms, respectively. The weights make it possible to account for the differing sizes of firms. Therefore, if the firms' size is not considered, the weights should be all set to 1.

Analogously to the Location Quotient, $\hat{M}(s) = 1$ represents the null hypothesis of no concentration. As a consequence, if $\hat{M}(s)$ computed for a given typology of firms is significantly greater (or lower) than 1 at some distance s, we have significant relative spatial concentration (or dispersion) of firms of that typology. As with the *K*-density measure, the significance envelopes can be obtained by means of Monte Carlo simulations based on random reallocations of firms.

A number of empirical studies in the field of regional sciences and spatial economics has used the *M* function: see for example Jensen and Michel (2011) and Moreno-Monroy and García (2016). Araldi and Fusco (2019), in particular, use a local version of the *M* function to investigate the local spatial pattern of retail activities in the French Riviera metropolitan area, in Southern France.

To give an example of the application of both the *K*-density and *M* function, here we analyse the spatial pattern of the single-plant firms of the metallurgical industry in the province of Trento (Italy) in 2009. In particular, the data of this example are a subset of the Statistical Register of Active Enterprises (ASIA), managed and updated by the Italian National Institute of Statistics (ISTAT). Figure 4 shows the spatial distribution of these firms together with that of all manufacturing industries of the economy in 2009. In the logic of both the *K*-density and the *M* function, to control for the inhomogeneity of space, the spatial distribution of all manufacturing single-plants (Figure 4b) is used as the null distribution. The plots depicted in Figure 5 show the behaviour of the empirical *K*-density and *M* function, together with the corresponding 99% significance envelopes.

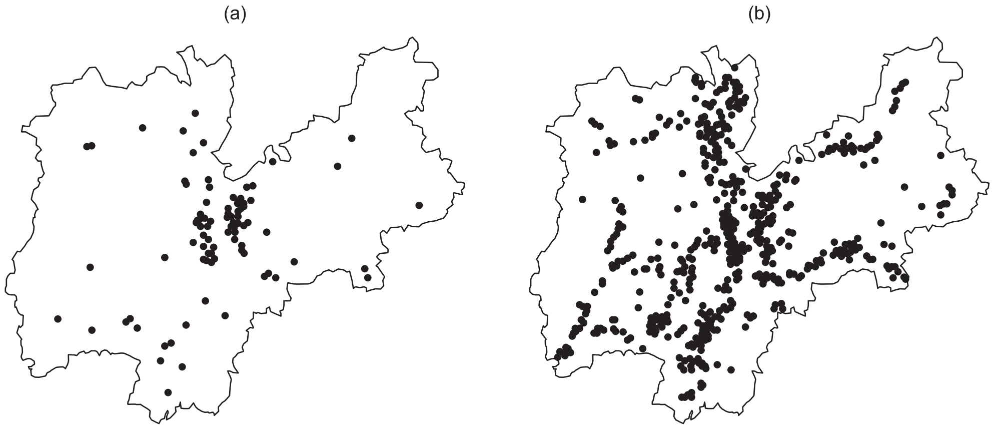

Figure 4: Spatial distribution of manufacturing firms in the province of Trento (Italy) in 2009: (a) the 98 single-plants firms of the metallurgical industry; (b) all 1007 manufacturing single-plants.
Source: Authors' elaboration.

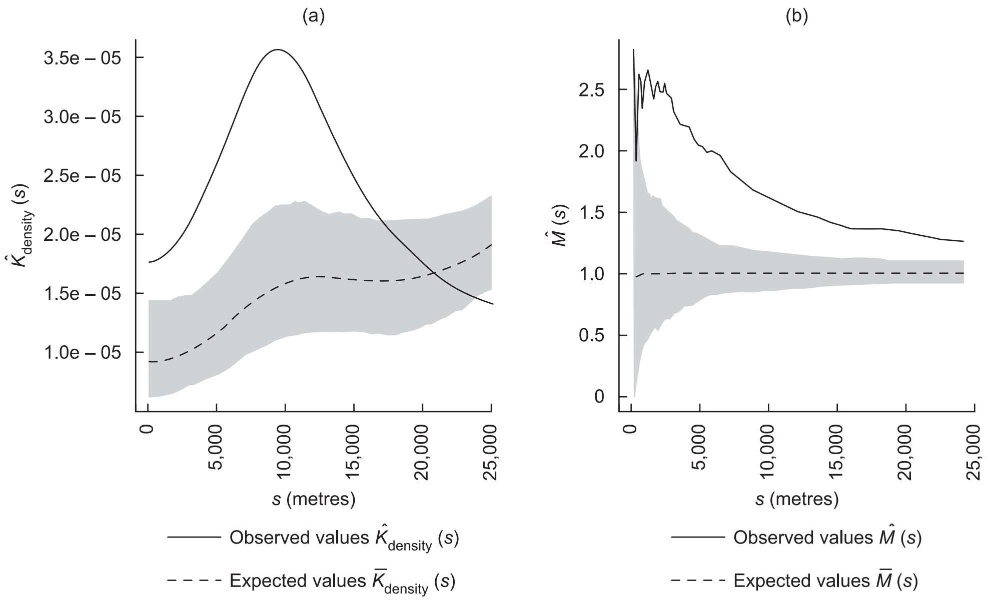

Figure 5: Empirical *K*-density and *M* function (continuous lines) and the corresponding 99% significance bands (shaded areas) for the metallurgy sector single-plant firms in the province of Trento (Italy) in 2009.
Source: Authors' elaboration.

Both measures show that metallurgical single-plant firms are significantly more spatially concentrated than the firms of the whole manufacturing industry, until at least 25 km. As a cumulative function which assesses spatial interactions among firms up to a given distance, the *M* function can detect the global pattern of spatial concentration more clearly. In contrast, as a density function which assesses spatial interactions among firms at a certain distance, the *K*-density can detect the occurrence of local clusters of firms more clearly. In this respect, Marcon and Puech (2010) show that the two functions constitute complementary, instead of alternative, ways to assess spatial concentration of industries. In fact, both measures should be employed to obtain an exhaustive picture of a spatial pattern of economic activities (Marcon, Puech, 2010). A further interesting discussion on the aspect of complementarities among alternative measures of localization can be found in Fratesi (2008).

5. Spatial location patterns of economic agents in an inhomogeneous space: the absolute approaches

As clarified above, the approaches based on the use of the *D*-function, *K*-density and *M* function are able to control for spatial inhomogeneity by assessing *relative* spatial interaction. Indeed, they all find positive (or negative) spatial dependence between the economic activities of interest when their locations are seen to be more aggregated (or dispersed) than the trend in another reference point pattern observed in the same space. This implies that the results for different spatial patterns cannot be compared, over the same distances, if the reference point pattern is not the same. Indeed, if the reference point pattern changes, then the benchmark null distribution capturing spatial inhomogeneity changes as well, thus invalidating any comparison.

Baddeley *et al.* (2000) developed the so-called K_{inhom}-function to detect and measure *absolute*, instead of *relative*, spatial interaction within an inhomogeneous space, thus allowing comparisons to be made between different point patterns observed in different spaces. It consists, essentially, in a generalization of Ripley's *K*-function, which works with homogeneous point processes, to the case of inhomogeneous processes. An inhomogeneous point process is characterized by a spatially varying first-order intensity function, $\lambda(x)$. Provided that $\lambda(x)$ is bounded away from zero, Baddeley *et al.* (2000) have shown that the second-order properties of an isotropic inhomogeneous point process can be described by the K_{inhom}-function: that is, by

$$K_{\text{inhom}}(s) = 2\pi \int_0^s g(t)\, t\, dt, \text{ with } s > 0, \text{ with } s > 0$$

where $g(s) = \lambda_2(s)/\lambda(x)\lambda(y)$ is the of the pair correlation functions introduced in Section 2. In the empirical circumstances where the underlying generating point process is inhomogeneous, the K_{inhom}-function measures the mean level of spatial interactions between points up to each distance *s*, while adjusting for spatial inhomogeneity, as described by $\lambda(x)$. In contexts where the space is inhomogeneous, a spatial point pattern characterized by the absence of spatial interactions can be considered as a realization of the so-called *inho-*

mogeneous Poisson process which, informally, is like an homogeneous Poisson process where the constant first-order intensity, λ, is substituted with a spatially varying first-order intensity, $\lambda(x)$. As a consequence, assessing whether an observed spatial point pattern is consistent with the hypothesis of no spatial interaction, under spatial inhomogeneity, corresponds to verifying whether the observed points may have been generated by an inhomogeneous Poisson process. As shown by Baddeley *et al.* (2000), if the underlying point process is an inhomogeneous Poisson process with first-order intensity $\lambda(x)$ and without spatial interactions between points then $K_{\text{inhom}}(s) = \pi s^2$. Therefore, if $K_{\text{inhom}}(s) > \pi s^2$ (or $K_{\text{inhom}}(s) < \pi s^2$), the underlying point process tends to generate point patterns that are more aggregated (or more spread) than point patterns that are realizations of an inhomogeneous Poisson process with first-order intensity $\lambda(x)$ (Diggle *et al.*, 2007). To help the interpretation, the same linear transformation used for Ripley's K-function can be employed, i.e. $L_{\text{inhom}}(s) = \sqrt{K_{\text{inhom}}(s)/\pi} - s$, which makes the null hypothesis of no spatial interactions represented by 0.

According to Baddeley *et al.* (2000), provided that $\lambda(x)$ is known, the approximately unbiased estimator of $K_{\text{inhom}}(s)$ for a spatial point pattern with n points, observed in a study region A, is

$$\hat{K}_{\text{inhom}}(s, \lambda(s)) = \frac{1}{|A|}\sum_{i=1}^{n}\sum_{j\neq i}\frac{w_{ij} I(s_{ij} \leq s)}{\lambda(x_i)\lambda(x_j)} \qquad [7]$$

where the symbols have the same meaning as in Equation [4].

In most of the practical applications $\lambda(x)$ is not known *a priori* and it is not possible to identify the proper theoretical economic model specifying its functional form. Therefore, it is necessary to provide an estimation of the first-order intensity. The literature has suggested both parametric and nonparametric estimation approaches. In those situations where a nonparametric estimate is a better choice, Baddeley *et al.* (2000) provided an estimator of $\lambda(x)$, which is an aptly modified version of the Berman and Diggle (1989)'s kernel estimator:

$$\hat{\lambda}_h(x_i) = \sum_{i\neq j} h^{-2} k\left(\frac{x_j - x_i}{h}\right) \Big/ C_h(x_j) \qquad [8]$$

where $k(\cdot)$ is a radially symmetric bivariate probability density function (generally of a Gaussian kind), h indicates the smoothing bandwidth and $C_h(x_j) = \int k_h(x_j - u)\,du$ is an edge effects correction factor[2].

Incorporating Equation [8] into Equation [7] makes it possible – in principle – to obtain approximately unbiased estimates of the K_{inhom}-function. Nevertheless, Diggle *et al.* (2007) have argued that estimating $\lambda(x)$ and $K_{\text{inhom}}(s)$ using the same data, i.e. using only one realization of the underlying

[2] If geo-referenced street network data are available, it may be appropriate to consider the Network Density Estimator of the first-order intensity proposed by Borruso (2008).

point process, may produce spurious estimates because we cannot properly discriminate the effects due to spatial inhomogeneity and spatial interaction without relying on some assumptions about the characteristics of the point process. For example, one may assume that the spatial scale of the first-order intensity is greater than that of the second-order intensity, and thus be able to separate the actual spatial interaction from spatial inhomogeneity by choosing a relatively high value for the bandwidth h (Diggle *et al.*, 2007). In order to make this kind of assumption, it is necessary to have prior knowledge or theoretical prescriptions about the extent of spatial interactions amongst points.

Otherwise, following the example of Espa *et al.* (2013) among others, in some real data applications it may be appropriate to treat $\lambda(x)$ as a function of spatially referenced covariates proxying the inhomogeneity of space. The values of $\lambda(x)$ may then be estimated parametrically through a regression model. For example, in the cases of spatial patterns of firms, the covariates could be common localization factors shared by the firms operating in the same area, such as the locations of communication routes, infrastructures or firms of different vertically related industries. A computationally convenient model for $\lambda(x)$ is the log-linear specification:

$$\lambda(x) = \exp\left\{\sum_{j=1}^{k} \beta_j z_j(x)\right\} \qquad [9]$$

where $z_j(x)$ is one of k spatially referenced covariates and β_j is the corresponding regression parameter. While assuming that the observed point pattern is the realization of an inhomogeneous Poisson process with intensity function $\lambda(x)$, the model in Equation [9] can be fitted to the data using maximum likelihood-based methods.

As with the K-function-based CSR test, described in Section 2, in order to assess the statistical significance of the deviations of $\hat{K}_{\text{inhom}}$ from the null hypothesis of no spatial interactions, we can rely on Monte Carlo significance envelopes obtained with simulations of the inhomogeneous Poisson process with $\lambda(x)$ estimated parametrically or non-parametrically using Equation [8] or Equation [9].

To illustrate how the K_{inhom}-function can be employed in practice, we refer to the problem of assessing the pattern of spatial competition in the distribution of supermarkets located in the municipality of Trento (Italy) in 2004 (Figure 6a). In this case, it is obviously not realistic to treat the city's territory as homogeneous since shops are naturally inclined to locate as close as possible to the potential market demand, which is clearly not equally distributed in the area. In this empirical circumstance it may be reasonable, for example, to use the number of households by census tract (Figure 6b) to proxy the spatial distribution of potential customers. Data were made available by the Italian National Institute of Statistics (ISTAT).

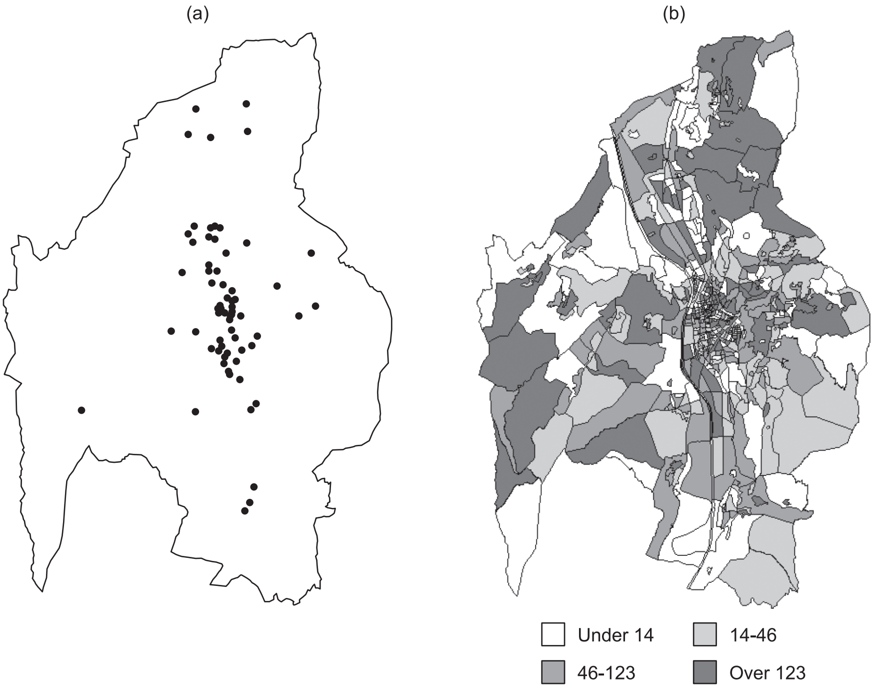

Figure 6: (a) spatial distribution of the 82 supermarkets in the city of Trento in 2004; (b) quartile distribution of the number of households by census tract in 2004.
Source: Authors' elaboration.

To assess spatial interactions among the 82 supermarkets operating in the city's territory, while adjusting for the spatial inhomogeneity of the potential market demand, the K_{inhom}-function-based test can be conducted by estimating the first order intensity function using Equation [9] and the number of households by census tract as a regressor capturing spatial inhomogeneity. The consequent estimation of the first order intensity provides the following results:

$$(x) = \exp\{-15.033 + 0.007h(x)\}$$

where $h(x)$ is the number of households in the census tract of location x. The estimated regression parameter for $h(x)$ is positive and significant (at the 5% level according to both the Wald and likelihood ratio tests). This shows that supermarkets tend to locate in the relatively more crowded census tracts. With $\hat{\lambda}(x)$ we can then estimate $\hat{K}_{inhom}(s)$ and hence $\hat{L}_{inhom}(s)$ using Equation [8]. Figure 7 shows the plot of $\hat{L}_{inhom}(s)$ and the corresponding 99.9% significance envelopes. The graphical test provides evidence of a multifaceted localization phenomenon occurring at different spatial scales. Indeed, it is apparent that the L_{inhom}-function is significantly greater than zero up to small distances (below 1.5 km), while it is significantly lower

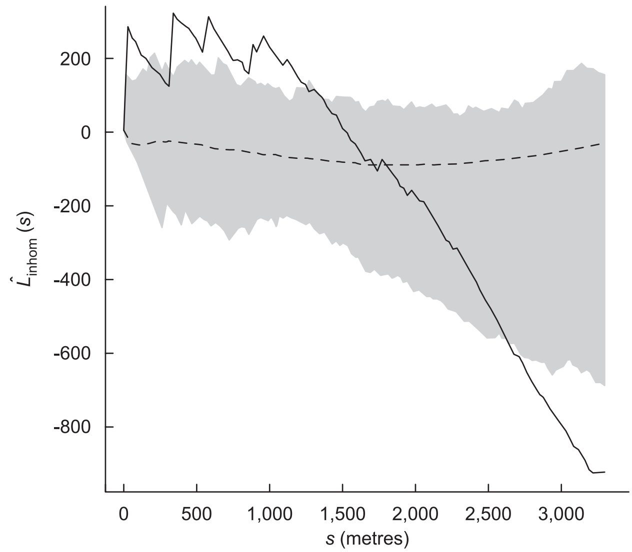

Figure 7: Empirical L_{inhom}-function (continuous line) and the corresponding 99.9% significance bands (shaded area) for the 82 supermarkets in the municipality of Trento (Italy), 2004.
Source: Authors' elaboration.

than zero up to distances greater than 2.5 km. This result suggests that, given the spatial distribution of potential market demand, the joint action of both positive and negative spatial externalities has led to the occurrence of spatial clusters of supermarkets with a spatial extension no greater than 1.5 km that, on average, are located at least 2.5 km away from the other clusters of supermarkets.

In this example, the K_{inhom}-function-based test has enabled empirical disentanglement of spatial inhomogeneity from spatial interaction in the spatial distribution of supermarkets in Trento. In particular, it has detected a strong tendency of stores to locate in secluded short-range clusters also because of real interactions among economic agents and not only as a consequence of a natural propensity to locate in the most crowded areas.

The K_{inhom}-function-based approach has been used, for example, in regional science studies by Arbia *et al.* (2012), Bonneu (2007) and Espa *et al.* (2013).

6. Conclusion

This article has provided a non-exhaustive, but hopefully representative, overview of the main empirical approaches to assessing spatial patterns of firms using micro-geographic data. Several studies (Duranton, 2008 and Combes *et al.*, 2008 among others) have argued that these approaches, as opposed to the methods based on regionally aggregated data, should now be preferred by researchers in the field of regional sciences. The use of geo-referenced micro-data is indeed a necessary requirement in analyses that consider interaction and contagion effects across units (e.g. consumers, firms), especially in the case of units operating in the same geographical area. However, despite the increasing availability of data of this kind of their use is, unfortunately, still quite limited. This is probably for two main reasons. The first is that there is no clear link between the empirical evidence that can be obtained from the use of micro-geographic data based-methods and the economic theories on industrial location. In this regard, some studies are moving in the direction of relating empirical analysis and theory (see e.g. Ellison *et al.*, 2010 and Kerr, Kominers, 2015). The second reason concerns the fact that geo-referenced micro-data researchers typically have to deal with are affected by localization errors that can occur both because of faults in the geo-coding procedures and because of deliberate choices made by the data supplier for reasons related with privacy issues (Zimmerman, 2008). Learning how to cope with this source of errors is an important future methodological challenge, following the way opened by Arbia *et al.* (2017).

In line with the advances in spatial econometrics for regional data (see Fingleton, 2017), another methodological development that could make the use of geo-referenced micro-data even more fruitful concerns the extension from a pure cross-sectional setting to a spatio-temporal modelling framework. An attempt in this direction has been made by Arbia *et al.* (2014) who proposed a method to assess the spatio-temporal concentration of firms' locations. We argue that the development of a more comprehensive methodology that includes the dynamic component may benefit from the theory of spatio-temporal point processes (González *et al.*, 2016), and we expect that it will be pursued in some future studies.

References

Antonietti R., Cainelli G., Lupi C. (2013), Vertical Disintegration and Spatial Co-Localization: The Case of Kibs in the Metropolitan Region of Milan. *Economics Letters*, 118, 2: 360-363. DOI: 10.1016/j.econlet.2012.11.031.

Araldi A., Fusco G. (2019), Retail Fabric Assessment: Describing Retail Patterns within Urban Space. *Cities*, 85: 51-62. DOI: 10.1016/j.cities.2018.11.025.

Arbia G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*. Dordrecht: Kluwer.

Arbia G., Espa G., Giuliani D., Mazzitelli A. (2012), Clusters of Firms in an Inhomogeneous Space: The High-Tech Industries in Milan. *Economic Modelling*, 29, 1: 3-11. DOI: 10.1016/j.econmod.2011.01.012.

Arbia G., Espa G., Giuliani D., Dickson M. M. (2014), Spatio-Temporal Clustering in the Pharmaceutical and Medical Device Manufacturing Industry: A

Geographical Micro-Level Analysi. *Regional Science and Urban Economics*, 49: 298-304. DOI: 10.1016/j.regsciurbeco.2014.10.001.
Arbia G., Espa G., Giuliani D., Dickson M. M. (2017), Effects of Missing Data and Locational Errors on Spatial Concentration Measures Based on Ripley's *K*-function. *Spatial Economic Analysis*, 12, 2-3: 326-346. DOI: 10.1080/17421772.2017.1297479.
Arbia G., Espa G., Quah D. (2008), A Class of Spatial Econometric Methods in the Empirical Analysis of Clusters of Firms in the Space. *Empirical Economics*, 34, 1: 81-103. DOI: 10.1007/s00181-007-0154-1.
Baddeley A. J., Møller J., Waagepetersen R. P. (2000), Non- and Semi-Parametric Estimation of Interaction in Inhomogeneous Point Patterns. *Statistica Neerlandica*, 54, 3: 329-350. DOI: 10.1111/1467-9574.00144.
Barff R. A. (1987), Industrial Clustering and the Organization of Production: A Point Pattern Analysis of Manufacturing in Cincinnati, Ohio. *Annals of the Association of American Geographers*, 77, 1: 89-103. DOI: 10.1111/j.1467-8306.1987.tb00147.x.
Behrens K., Bougna T. (2015), An Anatomy of the Geographical Concentration of Canadian Manufacturing Industries. *Regional Science and Urban Economics*, 51, 1: 47-69. DOI: 10.1016/j.regsciurbeco.2015.01.002.
Berman M., Diggle P. J. (1989), Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Process. *Journal of the Royal Statistical Society B*, 51, 1: 81-92. DOI: 10.2307/2345843.
Besag J. (1977), Contribution to the Discussion of Dr. Ripley's Paper. *Journal of the Royal Statistical Society B*, 39, 2: 193-195. DOI: 10.1111/j.2517-6161.1977.tb01616.x.
Besag J., Diggle P. J. (1977), Simple Monte Carlo Tests for Spatial Pattern. *Applied Statistics*, 26, 3: 327-333. DOI: 10.2307/2346974.
Bocci C., Rocco E. (2016), Modelling the Location Decisions of Manufacturing Firms with a Spatial Point Process Approach. *Journal of Applied Statistics*, 43, 7: 1226-1239. DOI: 10.1080/02664763.2015.1093612.
Bonneu F. (2007), Exploring and Modeling Fire Department Emergencies with a Spatio-Temporal Marked Point Process. *Case Studies in Business, Industry and Government Statistics*, 1, 2: 139-152.
Borruso G. (2008), Network Density Estimation: A GIS Approach for Analysing Point Patterns in a Network Space. *Transactions in GIS*, 12, 3: 377-402. DOI: 10.1111/j.1467-9671.2008.01107.x.
Cainelli G., Ganau R. (2018), Distance-Based Agglomeration Externalities and Neighbouring Firms' Characteristics. *Regional Studies*, 52, 7: 922-933. DOI: 10.1080/00343404.2017.1360482.
Cliff A. D., Ord J. K. (1972), *Spatial Autocorrelation*. London: Pion.
Combes P. P., Mayer T., Thisse J. F. (2008), *Economic Geography, the Integration of Regions and Nations*. Princeton: Princeton University Press. DOI: 10.2307/j.ctvcm4h9k.
Coniglio N. D., Lagravinese R., Vurchio D., Armenise M. (2018), The Pattern of Structural Change: Testing the Product Space Framework. *Industrial and Corporate Change*, 27, 4: 763-785. DOI: 10.1093/icc/dty009.

Cressie N. (1993), *Statistics for Spatial Data*. New York: Wiley.
Cuzick J., Edwards R. (1990), Spatial Clustering for Inhomogeneous Populations (with Discussion). *Journal of the Royal Statistical Society B*, 52: 73-104. DOI: 10.2307/2345652.
Diggle P. J. (2003), *Statistical Analysis of Spatial Point Patterns*. London: Edward Arnold, 2nd edition.
Diggle P. J., Chetwynd A. J. (1991), Second-Order Analysis of Spatial Clustering for Inhomogeneous Populations. *Biometrics*, 47: 1155-1163. DOI: 10.2307/2532668.
Diggle P. J., Gomez-Rubio V., Brown P. E., Chetwynd A. G., Gooding S. (2007), Second-Order Analysis of Inhomogeneous Spatial Point Processes Using Case-Control Data. *Biometrics*, 63, 2: 550-557. DOI: 10.2307/4541367.
Duranton G. (2008), Spatial Economics. In: Durlauf S. N., Blume L. E. (eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan. DOI: 0.1057/978-1-349-95121-5.
Duranton G., Overman H. G. (2005), Testing for Localization Using Micro-Geographic Data. *Review of Economic Studies*, 72, 4: 1077-1106. DOI: 10.2307/3700701.
Duranton G., Overman H. G. (2008), Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data. *Journal of Regional Science*, 48, 1: 213-243.
Ellison G., Glaeser E. L. (1997), Geographic Concentration in US Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105: 889-927. DOI: 10.1086/262098.
Ellison G., Glaeser E. L., Kerr W. R. (2010), What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100, 3: 1195-1213. DOI: 10.1257/aer.100.3.1195.
Espa G., Giuliani D., Arbia G. (2013), Conditional versus Unconditional Industrial Agglomeration: Disentangling Spatial Dependence and Spatial Heterogeneity in the Analysis of ICT Firms' Distribution in Milan. *Journal of Geographical Systems*, 15, 1: 31-50. DOI: 10.1007/s10109-012-0163-2.
Feser E. J., Sweeney S. H. (2000), A Test for the Coincident Economic and Spatial Clustering of Business Enterprises. *Journal of Geographical Systems*, 2, 4: 349-373. DOI: 10.1007/PL00011462.
Fingleton B. (2017), Dynamic Spatial Econometrics Models. A Critical Review. *Scienze Regionali – Italian Journal of Regional Science*, 16, 1: 7-30. DOI: 10.14650/85677.
Fratesi U. (2008), Issues in the Measurement of Localization. *Environment and Planning A*, 40: 733-758. DOI: 10.1068/a39223.
González J. A., Rodríguez-Cortés F. J., Cronie O., Mateu J. (2016), Spatio-Temporal Point Process Statistics: A Review. *Spatial Statistics*, 18, Part B: 505-544. DOI: 10.1016/j.spasta.2016.10.002.
Jensen P., Michel J. (2011), Measuring Spatial Dispersion: Exact Results on the Variance of Random Spatial Distributions. *The Annals of Regional Science*, 47, 1: 81-110. DOI: 10.1007/s00168-009-0342-3.
Kerr W. R., Kominers S. D. (2015), Agglomerative Forces and Cluster Shapes. *The Review of Economics and Statistics*, 97, 4: 877-899. DOI: 10.1162/REST_a_00471.

Klier T., McMillen D. P. (2008), Evolving Agglomeration in the US Auto Supplier Industry. *Journal of Regional Science*, 48, 1: 245-267. DOI: 10.1111/j.1467-9787.2008.00549.x.
Koh H. J., Riedel N. (2014), Assessing the Localization Pattern of German Manufacturing and Service Industries: A Distance-Based Approach. *Regional Studies*, 48, 5: 823-843. DOI: 10.1080/00343404.2012.677024.
Kosfeld R., Eckey H.-F., Lauridsen J. (2011), Spatial Point Pattern Analysis and Industry Concentration. *Annals of Regional Science*, 47: 311-328. DOI: 10.1007/s00168-010-0385-5.
Marcon E., Puech F. (2003), Evaluating the Geographic Concentration of Industries Using Distance-Based Methods. *Journal of Economic Geography*, 3, 4: 409-428. DOI: 10.2307/26160499.
Marcon E., Puech F. (2010), Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods. *Journal of Economic Geography*, 10, 5: 745-762. DOI: 10.1093/jeg/lbp056.
Møller J., Waagepetersen R. P. (2004), *Statistical Inference and Simulation for Spatial Point Processes*. Volume 100 of Monographs on Statistics and Applies Probabilities. London: Chapman and Hall.
Moreno-Monroy A. I., García G. A. (2016), Intra-Metropolitan Agglomeration of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia. *Tijdschrift Voor Economische en Sociale Geografie*, 107, 4: 389-406. DOI: 10.1111/tesg.12163.
Ó hUallacháin B., Leslie T. F. (2007), Producer Services in the Urban Core and Suburbs of Phoenix, Arizona. *Urban Studies*, 44, 8: 1581-1601. DOI: 10.1080/00420980701373453.
Ripley B. D. (1976), The Second-Order Analysis of Stationary Point Processes. *Journal of Applied Probability*, 13: 255-266. DOI: 10.2307/3212829.
Ripley B. D. (1977), Modelling Spatial Patterns (with discussion). *Journal of the Royal Statistical Society B*, 39, 2: 172-212. DOI: 10.2307/2984796.
Silverman B. W. (1986), *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.
Sweeney S. H., Feser E. J. (1998), Plant Size and Clustering of Manufacturing Activity. *Geographical Analysis*, 30, 1: 45-64. DOI: 10.1111/j.1538-4632.1998.tb00388.x.
Vitali S., Napoletano M., Fagiolo G. (2013), Spatial Localization in Manufacturing: A Cross-Country Analysis. *Regional Studies*, 47, 9: 1534-1554. DOI: 10.1080/00343404.2011.625006.
Zimmerman D. L. (2008), Estimating the Intensity of a Spatial Point Process from Locations Coarsened by Incomplete Geocoding. *Biometrics*, 64, 1: 262-270. DOI: 10.2307/25502044.