Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Modelling the location decisions of manufacturing firms with a spatial point process approach

Chiara Bocci & Emilia Rocco

To cite this article: Chiara Bocci & Emilia Rocco (2015): Modelling the location decisions of manufacturing firms with a spatial point process approach, Journal of Applied Statistics, DOI: 10.1080/02664763.2015.1093612

To link to this article: http://dx.doi.org/10.1080/02664763.2015.1093612

Published online: 12 Oct 2015.

Journal of Applied Statistics, 2015
http://dx.doi.org/10.1080/02664763.2015.1093612

Modelling the location decisions of manufacturing firms with a spatial point process approach

Chiara Bocci[a*] and Emilia Rocco[b]

[a]*IRPET – Regional Institute for Economic Planning of Tuscany, Firenze, Italy;* [b]*Department of Statistics, Informatics, Applications'G. Parenti' (DiSIA), University of Firenze, Italy*

(*Received 28 January 2015; accepted 9 September 2015*)

The paper is devoted to explore how the increasing availability of spatial micro-data, jointly with the diffusion of GIS software, allows to exploit micro-econometric methods based on stochastic spatial point processes in order to understand the factors that may influence the location decisions of new firms. By using the knowledge of the geographical coordinates of the newborn firms, their spatial distribution is treated as a realization of an inhomogeneous marked point process in the continuous space and the effect of spatial-varying factors on the location decisions is evaluated by parametrically modelling the intensity of the process. The study is motivated by the real issue of analysing the birth process of small and medium manufacturing firms in Tuscany, an Italian region, and it shows that the location choices of the new Tuscan firms is influenced on the one hand by the availability of infrastructures and the level of accessibility, and on the other by the presence and the characteristics of the existing firms. Moreover, the effect of these factors varies with the size and the level of technology of the new firms. Besides the specific Tuscan result, the study shows the potentiality of the described micro-econometric approach for the analysis of the spatial dynamics of firms.

Keywords: firm demography; spatial econometrics; spatial heterogeneity; multivariate point process; georeferenced data

1. Introduction

In the last years, location and physical geography characteristics have become relevant variables in the regional econometric studies. Understanding the possible factors that influence the location decisions of new firms is of key importance for policy-makers because the local economic growth partially depends on the ability to attract new investments, and the knowledge of such determinants could be useful to implement effective policies to increase an area attractiveness. This is confirmed by 'the boost in the number of empirical studies investigating the driving forces behind the location decisions of new industrial concerns' reviewed by Arauzo-Carod *et al.* [1].

*Corresponding author. Email: chiara.bocci@irpet.it

Moreover, the increasing availability of spatially referenced data jointly with the diffusion of Geographical Information System (GIS) software has allowed the development of new methods for the analysis of the spatial dynamics of firms: the use of spatial micro-economic data can move the analysis point of view from the classic macro perspective to a micro-econometric framework.

Finally, the knowledge of the exact geographical location of firms allows us to combine the firms data with spatial-varying variables available from different data sources.

In such setting, the framework of spatial point process methods [9,16] plays a major role since spatial point processes can be used directly to model and analyse data which takes form of a spatial point pattern, such as the geographical distribution of firms [2,3,7].

Exploiting spatial point process methods, this paper explores the birth process of small and medium manufacturing firms in Tuscany and the possible determinants of their location decisions. To this end, the geographical distribution of the manufacturing firms born in Tuscany between 2005 and 2008 is defined in terms of an inhomogeneous marked point process in the continuous space and the effect of several spatial-varying factors on the location decisions of new firms is evaluated by parametrically modelling the intensity of the process.

Results are encouraging with respect to the potential of the suggested method of analysis and show that the availability of infrastructures, the accessibility, and the presence and characteristics of the existing firms are relevant in the location decision-making of the Tuscan firms. Moreover, the effect of these factors varies with the size and level of technology of the newborn firms.

The paper is structured as follows: Section 2 describes the data and Section 3 is devoted to the statistical framework of the analysis. Section 4 presents the application and the specific results while more general conclusions are drawn in Section 5.

2. Data description

As of 2004, each year the Italian Statistical Institute (Istat) issues the Statistical Register of Local Units of Active Enterprises (ASIA-UL) which comprises data on the street address, the sector of economic activity and the number of employees of each enterprise local unit. The sector of economic activity is identified according to the European classification of economic activities NACE Rev.2 [12], while the number of employees is measured in terms of annual average of job positions, considering both full- and part-time workers, either employed or self-employed. The field of observation of ASIA-UL covers all industrial, commercial and service-sector private activities.

By European Union definition [10], an enterprise local unit is an enterprise or part thereof situated in a geographically identified place. Thus the local units, simply referred to as firms in the following, are the natural units of analysis in our study. Moreover, since the focus of the analysis is on the birth process of firms, we need to define which are the newborn firms. At the time of our analysis only the 2004 and 2008 ASIA-UL archives were available to us, consequently we define the newborn firms as the local units born in Tuscany in the period 2005–2008, that is, the units not existent in ASIA-UL 2004 and included in ASIA-UL 2008.

Exploiting the databases georeferencing system of the Tuscany Region, we link the street address of each local unit with its corresponding geographical coordinates (expressed in the Monte Mario coordinate system).

The knowledge of the firms' geographical coordinates allows us to merge the ASIA-UL data set with a set of administrative and geostatistical data sets which include several spatial-varying variables. In particular, we consider the following data sources:

- the *Digital Terrain Model* (DTM) of Tuscany by the Italian Military Geographical Institute from which we derive the elevation and the terrain slope of each firm location;

- the *Survey on the urban land use in Tuscany* by the LaMMA Consortium (Environmental Modelling and Monitoring Laboratory), which classify the Tuscan territory in categories of urban land use;
- the data set of the *Agenda Digitale* of the Italian Ministry of Economic Development on the status of the broadband coverage, which classify each census track in *ADSL broadband covered*, *mobile broadband covered* and *digital divided*;
- the *Osservatorio del Mercato Immobiliare* (OMI – Observatory for the real estate market) of the Italian Revenue Agency, which divides each municipal area in sub-areas (OMI-area) that

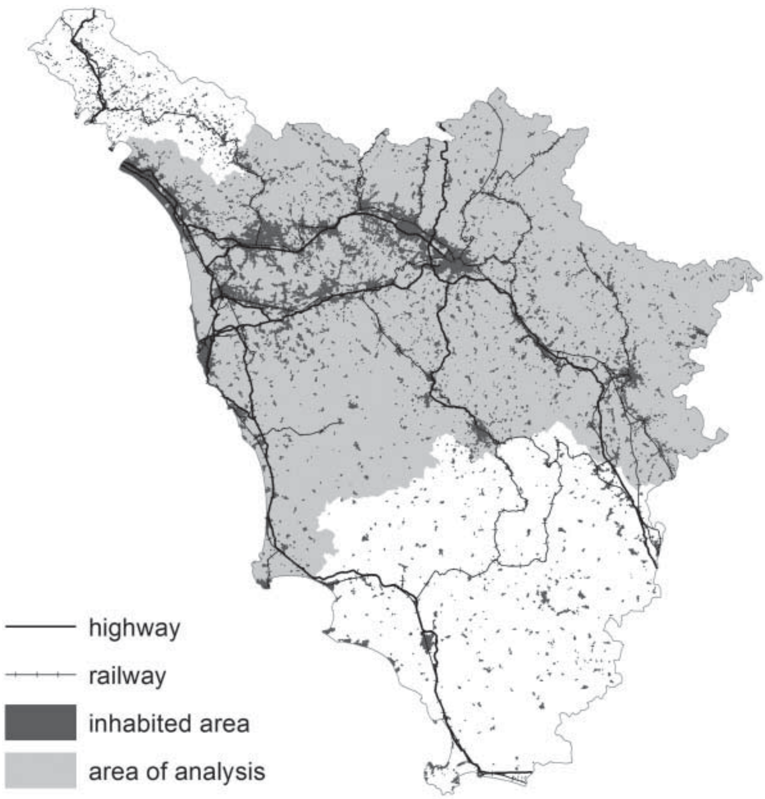

Figure 1. Inhabited areas and main transport infrastructures in Tuscany.

Table 1. Frequency and density distribution of the new manufacturing local units born in 2005–2008 divided into six groups: three levels of technology (*low*, *medium-low* and *medium-high and high*) and two levels of size (*micro* and *small and medium*).

Group	Description	Frequency	Density[a]
1	Micro firms with low technology	8750	0.6150
2	Small and medium firms with low technology	683	0.0480
3	Micro firms with medium-low technology	2550	0.1790
4	Small and medium firms with medium-low technology	286	0.0201
5	Micro firms with medium-high and high technology	721	0.0506
6	Small and medium firms with medium-high and high technology	206	0.0145

Note: [a] The density is calculated as ratio between the frequency and the area of the region of analysis.

are homogeneous with respect to the real estate market and collects the real estate prices for industrial, commercial and housing market in each sub-area.

- the *Osservatorio per la Mobilità ed i Trasporti* (Observatory for Mobility and Transport) of the Tuscany Region, which collects information on the transport infrastructures in Tuscany; the knowledge of the geographical coordinates of the train stations and highway accesses allows us to calculate for each firm the distances from the nearest train station and highway access.

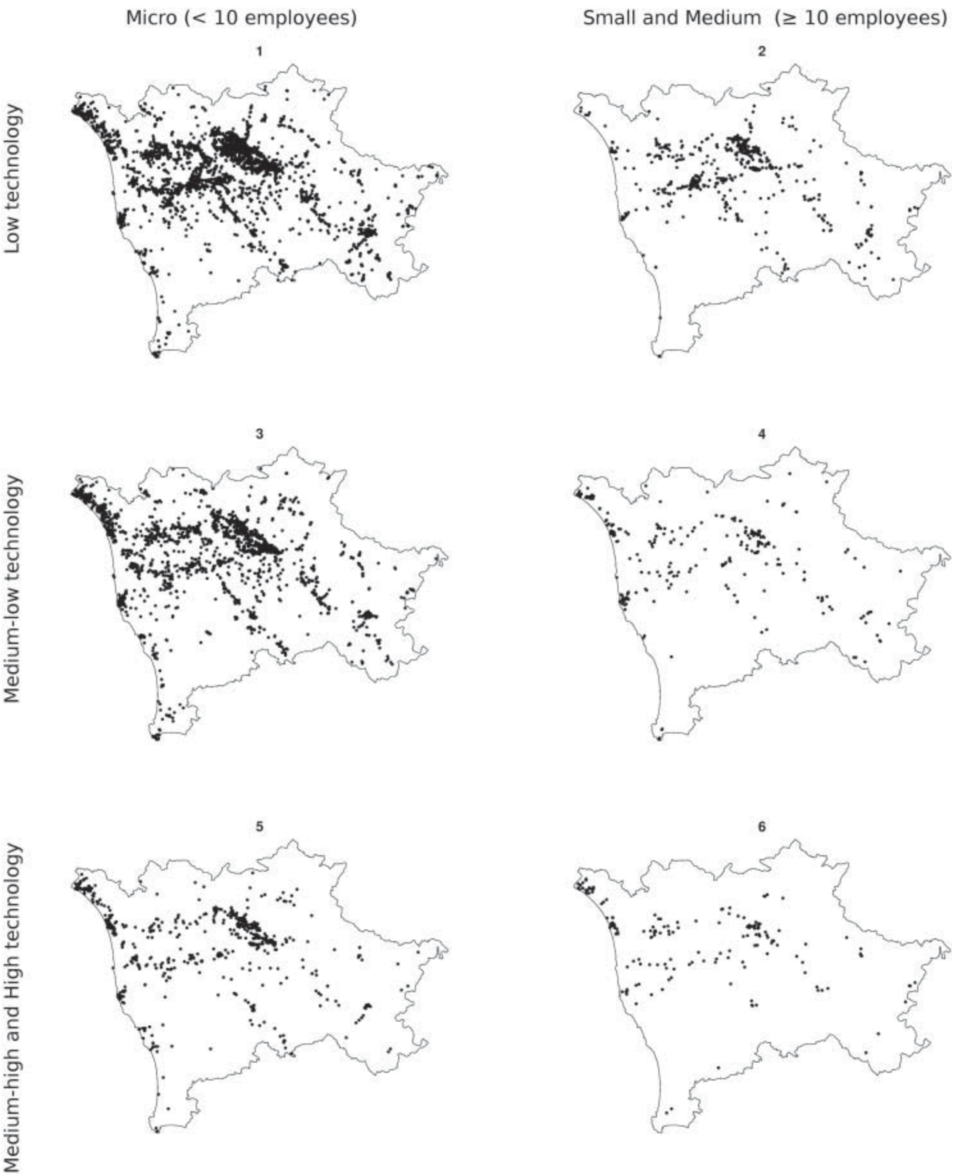

Figure 2. Spatial distribution of the 13,196 new manufacturing local units born in 2005–2008 divided into six groups: three levels of technology (*low*, *medium-low* and *medium-high and high*) and two levels of size (*micro* and *small and medium*).

All the variables included in these data sets could be useful to describe the characteristics of each firm location. To the same purpose, we calculate the Eurostat degree of urbanization index which classify each municipality in three categories: thinly populated area (rural area), intermediate density area (towns and suburbs/small urban area) and densely populated area (cities/large urban area) [11, p. 156].

In addition, we exploit the geographical coordinates of the ASIA-UL 2004 local units in order to characterize the existing economic context. For each newborn firm, we count the number of existing firms belonging to each sector of economic activity within a radius of 5 km. Finally, we calculate the density of local units dead in the period 2004–2007 (i.e. the local units included in ASIA-UL 2004 and not existent in ASIA-UL 2008).

The subject of our analysis are the small and medium newborn manufacturing firms in Tuscany. Since most of the manufacturing activity in Tuscany (92% of local units and 93% of employees) is located in the North-Central area of the region, we restrict the analysis to this geographical area. As shown in Figure 1, this is also the region where most of the inhabited areas and the main transport infrastructures are located.

Following Istat and Eurostat guidelines, the manufacturing firms can be classified accordingly to the technological intensity of their production (measured as the ratio between the research & development spending and the value added) and on their size [13, pp. 306–307]: (i) according to their technological intensity, the firms are aggregated in *high technology*, *medium-high technology*, *medium-low technology* and *low technology* based on the NACE Rev.2 classification at 2-digit level; (ii) according to their number of employees, the firms are classified into *micro* (less than 10 employees), *small* (between 10 and 49 employees) and *medium* (between 50 and 249 employees).

In our analysis, due to the small amount of manufacturing firms with high and medium-high technology, we consider them as a unique group obtaining three levels of technology (*low*, *medium-low* and *medium-high and high*) and, due to the scarcity of firms with more than 50 employees, we define only two levels of size (*micro* and *small and medium*). Combining the three levels of technology and the two levels of size we obtain six groups of firms, as described in Table 1.

An image depicting the location of the 13,196 newborn manufacturing local units for each of the six groups is presented in Figure 2. The firms show a clear tendency to locate in specific sub-areas, that correspond to the main cities and to the main infrastructures of the region (see Figure 1). In addition, as shown in both Table 1 and Figure 2, the number of units in each group varies significantly: most of them belong to the low and medium-low technology firms with less than 10 employees.

3. Model framework

An observed point pattern, like the one representing the Tuscan firms locations, can be treated as a realization of a spatial point process $\mathbf{S}$, which is a stochastic mechanism that generates the locations of some events of interest (units) within a bounded region A.

Let $\mathbf{s} = [s_1, s_2]$ be a generic point of coordinates ($\mathbf{s} \in R^2$) and $\delta(\mathbf{s})$ a region in the neighbourhood of the point $\mathbf{s}$. Then we can define the local density of the spatial point process $\mathbf{S}$ (known as intensity) [9,15] at a point $\mathbf{s}$ as

$$\lambda(\mathbf{s}) = \lim_{\delta(\mathbf{s})\to 0} \frac{E[N(\delta(\mathbf{s}))]}{\delta(\mathbf{s})},$$

where $N(\delta(\mathbf{s}))$ represents the number of points falling within $\delta(\mathbf{s})$. The intensity $\lambda(\mathbf{s})$ represents the expected number of units in the infinitesimal area $\delta(\mathbf{s})$.

Let $N(A)$ be the number of units located in A generated by $\mathbf{S}$, then $\mathbf{S}$ is a Poisson point process on R^2 if:

- $N(A) \sim \text{Poisson}(\lambda(A))$ for every A
- if $A_1, \ldots, A_h$ are disjoint regions, then $N(A_1), \ldots, N(A_h)$ are independent

where $\lambda(A) = \int_A \lambda(\mathbf{s})\, d\mathbf{s}$.

If $\lambda(\mathbf{s}) = \lambda$ is constant, the process is called homogeneous. Otherwise, the process is called inhomogeneous.

In addition, some attributes may be observable at each event location. These attributes can be continuous or discrete and are called mark variables. For a detailed description of the corresponding marked point process, we refer to [8,17]. Here, we consider only the case in which it is available a unique discrete mark variable that classifies the events within a bounded region A into M types. If the events located in a given region can be classified into two or more different types, they can be treated as a realization of a multivariate (or multitype in [14] notation) point process. In this case, the event counts for the entire process $\mathbf{S}$ is the $(M \times 1)$ vector $\mathbf{N}(A) = [N_1(A), \ldots, N_M(A)]$, where the element $N_m(A)$ is the number of events with mark m generated by the sub-process $\mathbf{S}_m$ and the local density for $\mathbf{S}_m$ is defined as

$$\lambda_m(\mathbf{s}) = \lim_{\delta(\mathbf{s}) \to 0} \frac{E[N_m(\delta(\mathbf{s}))]}{\delta(\mathbf{s})}.$$

Finally, the intensity of the point process of interest could also be influenced by some spatial covariates or some other spatial patterns observed in the same region.

As stated by Arbia [2], the birth process of the firms is spatially uneven: some locations could be more likely to be chosen than others, for various economic reasons, and this will produce an irregular pattern. Thus, to model such a situation it is useful to consider the birth process of the firms as an inhomogeneous process with spatial intensity that depends from economic characteristics and varies according to location. Higher values of $\lambda(\mathbf{s})$ indicate a higher concentration of economic activities in the infinitesimal area centred in $\mathbf{s}$. Moreover, as the newborn firms are usually classified in different types (e.g. by size or economical activity), their birth process should be treated as an inhomogeneous multivariate point pattern.

For simplicity let us consider a single spatial covariate $x(\mathbf{s})$, then the intensity of the inhomogeneous multivariate point process in $\mathbf{s}$ can be modelled as

$$\lambda_m(\mathbf{s}) = \exp[\alpha_m + \beta_m x(\mathbf{s})] \tag{1}$$

where $\alpha_1, \ldots, \alpha_M$ and $\beta_1, \ldots, \beta_M$ are mark-specific parameters and the intensity is log-linear in $x(\mathbf{s})$ with a different slope and intercept for each $m = 1, \ldots, M$.

The common approach to estimate the parameters of (1) is to use the logarithmic transformation and to maximize the log-pseudo-likelihood for $\lambda_m(\mathbf{s})$ based on the observed points of the pattern under study. The method, first proposed by Berman and Turner [6], approximates the log-likelihood by a finite sum that has the same analytical form as the (weighted) log-likelihood of a generalized linear model with Poisson responses. The obtained approximate likelihood could then be maximized using some existing software for the generalized linear models. Baddeley and Turner [4,5] extended this method and implemented it in the package `spatstat` in the `R` computing environment. Moreover, as shown by Strauss and Ikeda [18], in the case of a Poisson stochastic process the maximum pseudo-likelihood is equivalent to the maximum likelihood, therefore it is possible to test the goodness of fit of the adopted model by using the standard formal likelihood ratio criteria and the χ^2 distribution.

Following Arbia [2], the adequacy of the fitted model can be assessed with a Monte Carlo test based on the visual inspection of the empirical K-function of the observed point pattern contrasted with the 99% bands derived from simulations generated from the estimated model.

Ripley's K-function $K(r)$ measures the number of events found up to a given distance r of any particular event and, for a homogeneous process with intensity λ, it is defined as

$$K(r) = \lambda^{-1}\mathrm{E}(N_0(r)), \tag{2}$$

where $N_0(r)$ represents the number of further events up to a distance r around an arbitrary event.

One generalization of $K(r)$ to a multivariate point pattern is [9, p. 91]

$$K^{ml}(r) = \lambda_l^{-1}E[\#\text{ of events of type } l \text{ within distance } r \text{ of an arbitrary event of type } m].$$

This function is called 'cross-type' since it is computed between all pairs of types. For any particular event, it applies Equation (2) M times, one for each component of the multivariate point pattern.

Baddeley and Turner [4] extended the K-function to the case of inhomogeneous point process. The inhomogeneous K-function $K_{\text{inhom}}(r)$ has the same interpretation as Equation (2), with the only exception that the intensity of the events is no longer constant but depends on the location of the events. To account for that, every observation is weighted with the intensity of its respective location. The cross version of $K_{inhom}(r)$ is straightforward.

Since we are interested in modelling our data as an inhomogeneous multivariate point pattern, the K-function suitable to assess the adequacy of our model is the inhomogeneous cross K-function that is usually estimated by

$$\hat{K}^{ml}_{\text{inhom}}(r) = \frac{1}{|A|}\sum_{\mathbf{s}_i \in \mathbf{S}_m}\sum_{\mathbf{s}_j \in \mathbf{S}_l} w_{ij}^{-1}\frac{I\{\mathrm{d}(\mathbf{s}_i, \mathbf{s}_j) \leq r\}}{\hat{\lambda}_m(\mathbf{s}_i)\hat{\lambda}_l(\mathbf{s}_j)}, \tag{3}$$

where $|A|$ is the total surface of region A, I is the indicator function, $d(\mathbf{s}_i, \mathbf{s}_j)$ is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, $\hat{\lambda}_m$ and $\hat{\lambda}_j$ are the estimated intensity functions of the sub-processes $\mathbf{S}_m$ and $\mathbf{S}_l$ and w_{ij} is the edge correction weight which is equal to the proportion of the surface area of the circle centred on $\mathbf{s}_i$ and with radius $d(\mathbf{s}_i, \mathbf{s}_j)$ which lies within A.

4. Application

This section analyses the location decisions of the small and medium manufacturing firms born in Tuscany between 2005 and 2008. The observed spatial pattern is modelled in terms of an inhomogeneous multivariate Poisson point process. Following the approach described in Section 3, the intensity of the process is assumed to be parametrically dependent from a set of spatial-varying variables and to be different for each group of firms defined in Section 2. Thus, in our model the six groups of firms identify the components (marks) of the multivariate point process.

In the model selection process, we considered all the variables that were available from the data sources described in Section 2. Due to the great amount of observations available, a blind use of the usual goodness-of-fit tests always results in the choice of the more complex model. Thus, to choose among several possible models we considered a trade off between three desirable characteristics: (i) the value of the likelihood ratio test; (ii) the significance of each parameter (e.g. for a spatial covariate we included mark-specific coefficients if at least two of them were statistically different from the others); (iii) the economic plausibility of the model also motivated by previous literature [1].

Finally, the variables included in the selected model can be partitioned in two groups. Those belonging to the first group describe the characteristics of the spatial location and are:

x_1 : the price of industrial buildings (Euro/m^2) of the correspondent OMI-area,
x_2 : the minimum distance from the location to the train station or to the highway access (km),
x_3 : a dummy variable that assumes value 1 if the correspondent census track is covered by the ADSL broadband and 0 otherwise,
x_4 : a dummy variable that assumes value 1 if the corresponding municipality is a densely populated or intermediate area and 0 if it is a thinly populated area,
x_5 : the terrain slope.

The second group of variables characterizes the existing economic context:

x_6 : the number of existing manufacturing units located within a radius of 5 km from the new local unit and of the same sector of economic activity,
x_7 : the number of existing manufacturing units located within a radius of 5 km from the new local unit and of a different sector of economic activity,
x_8 : the number of existing commercial units located within a radius of 5 km from the new local unit,
x_9 : the number of existing tertiary sector units located within a radius of 5 km from the new local unit,
x_{10} : the density of manufacturing units dead in the period 2004–2007.

With the exception of variable x_7, the variables that count the number of neighbour existing units for each sector are included in the model with mark-specific coefficients. The influence of the presence of manufacturing units of different sector does not vary significantly among the six groups of firms; however, it shows a different behaviour in the thinly populated areas than in the other areas. In order to account for this, we included in the model the interaction term between x_4 and x_7.

Among the variables of the first group, only x_2 has been included in the model with mark-specific coefficients.

The final model specification for the intensity of the process in the point of coordinates $\mathbf{s}$ is

$$\lambda_m(\mathbf{s}) = \exp[\alpha_m + \boldsymbol{\beta}\mathbf{x}_a(\mathbf{s}) + \boldsymbol{\gamma}_m\mathbf{x}_b(\mathbf{s})], \tag{4}$$

where

- $\alpha_1, \ldots, \alpha_6$ are the mark-specific intercepts,
- $\mathbf{x}_a(\mathbf{s}) = [x_1(\mathbf{s}), x_3(\mathbf{s}), x_4(\mathbf{s}), x_5(\mathbf{s}), x_7(\mathbf{s}), x_{10}(\mathbf{s}), x_4(\mathbf{s}) * x_7(\mathbf{s})]$ is the vector of variables whose effects are constant for all marks,
- $\boldsymbol{\beta}$ is the vector of the coefficients relative to $\mathbf{x}_a(\mathbf{s})$,
- $\mathbf{x}_b(\mathbf{s}) = [x_2(\mathbf{s}),\ x_6(\mathbf{s}),\ x_8(\mathbf{s}),\ x_9(\mathbf{s})]$ is the vector of variables whose effects vary between marks,
- $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_6$ are the vectors of the mark-specific coefficients corresponding to $\mathbf{x}_b(\mathbf{s})$.

Model (4) is fitted maximizing the pseudo-likelihood by using the package `spatstat` in the `R` computing environment.

Table 2 presents the estimated parameters of our model, with the corresponding standard errors, Z-tests and exponential transformations. Most of the coefficients are highly significant, and from their values we are able to identify the influence of each variable on the location decisions of the six groups of new local units.

Table 2. Estimated parameters of the model, with the corresponding standard errors, Z-tests and exponential transformations.

Variable	Parameter Estimate	S.E.	*Z*-test	exp(est)
Intercept, $m = 1$	– 3.00407	0.08711	$***$	0.04959
Intercept, $m = 2$	– 4.65704	0.12017	$***$	0.00949
Intercept, $m = 3$	– 3.44456	0.09333	$***$	0.03192
Intercept, $m = 4$	– 5.40461	0.15904	$***$	0.00450
Intercept, $m = 5$	– 4.65897	0.12012	$***$	0.00948
Intercept, $m = 6$	– 5.68912	0.18048	$***$	0.00338
x_1	0.00054	0.00002	$***$	1.00054
x_3	1.36413	0.05761	$***$	3.91233
x_4	1.21120	0.05859	$***$	3.35749
x_5	– 0.08945	0.00218	$***$	0.91443
x_7	0.00292	0.00028	$***$	1.00293
x_{10}	0.29360	0.00611	$***$	1.34125
$x_4 * x_7$	– 0.00252	0.00028	$***$	0.99748
$x_2, m = 1$	– 0.02843	0.00656	$***$	0.97197
$x_2, m = 2$	– 0.16221	0.02393	$***$	0.85026
$x_2, m = 3$	– 0.13234	0.01205	$***$	0.87605
$x_2, m = 4$	– 0.25534	0.04335	$***$	0.77466
$x_2, m = 5$	– 0.22731	0.02728	$***$	0.79667
$x_2, m = 6$	– 0.27048	0.05240	$***$	0.76301
$x_6, m = 1$	0.00086	0.00002	$***$	1.00086
$x_6, m = 2$	0.00098	0.00007	$***$	1.00098
$x_6, m = 3$	– 0.00078	0.00006	$***$	0.99922
$x_6, m = 4$	– 0.00121	0.00022	$***$	0.99879
$x_6, m = 5$	– 0.00072	0.00011	$***$	0.99928
$x_6, m = 6$	– 0.00184	0.00035	$***$	0.99816
$x_8, m = 1$	0.00036	0.00003	$***$	1.00036
$x_8, m = 2$	0.00008	0.00010		1.00008
$x_8, m = 3$	0.00061	0.00004	$***$	1.00061
$x_8, m = 4$	0.00096	0.00012	$***$	1.00096
$x_8, m = 5$	0.00073	0.00007	$***$	1.00073
$x_8, m = 6$	0.00091	0.00013	$***$	1.00091
$x_9, m = 1$	– 0.00027	0.00001	$***$	0.99973
$x_9, m = 2$	– 0.00025	0.00006	$***$	0.99975
$x_9, m = 3$	– 0.00042	0.00002	$***$	0.99958
$x_9, m = 4$	– 0.00066	0.00007	$***$	0.99934
$x_9, m = 5$	– 0.00047	0.00004	$***$	0.99953
$x_9, m = 6$	– 0.00060	0.00007	$***$	0.99940

Note: *Z*-test significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $.p < 0.1$.

The six mark-specific parameters for the minimum distance to the train station or to the highway, x_2, indicate that the distance has a negative effect on the probability of location of a new manufacturing firm: the higher is the distance the lower is the expected number of new firms. Such relationship is relevant for all the types of firms, but it is stronger when the firm size increases or when the level of technology is higher. For example, for every 1 km increase in the minimum distance to the train station or to the highway, the intensity of new micro firms with low technology ($m = 1$) drops by nearly 3%, whereas the intensity of new small and medium firms with medium-high of high technology ($m = 6$) drops by nearly 24%. This result is in accordance with the theoretical expectations, due to the necessity of an easy access to the labour force and the ease of shipment of manufacturing goods. Similarly, the negative coefficient of the terrain slope (x_5) indicates that a friendlier environment (a more flat terrain) is preferable: when the slope increase by 1 degree the expected number of any type of firms drops by nearly 9%.

In accordance with these results is the positive influence of the availability of infrastructures, summarized by the degree of urbanization (x_4), the ADSL coverage (x_3) and the price of industrial buildings (x_1). The intensity of the new firms in the areas covered by ADSL is almost four times higher than in the areas not covered; in the same way, the intensity is more than three times higher in the densely or intermediate populated areas compared to the thinly populated areas. Focusing on the parameter of the price of industrial buildings, we should note that the multiplicative effect (measured by the exponential transformation) is nearly equal to 1. This should not mislead us about the relevance of this spatial covariate, as the parameter measures the variation corresponding to the increment of price of 1 Euro per square metre. By considering a realistic price variation, that is usually far higher than the unitary variation, the influence on the intensity appears evident: for example if the price of industrial buildings per square metre increases by 100 Euro, the intensity of the new firms rises more than 5%. Moreover, the positivity of the coefficient of x_1 should not be read as a preference of the new firms to locate in areas

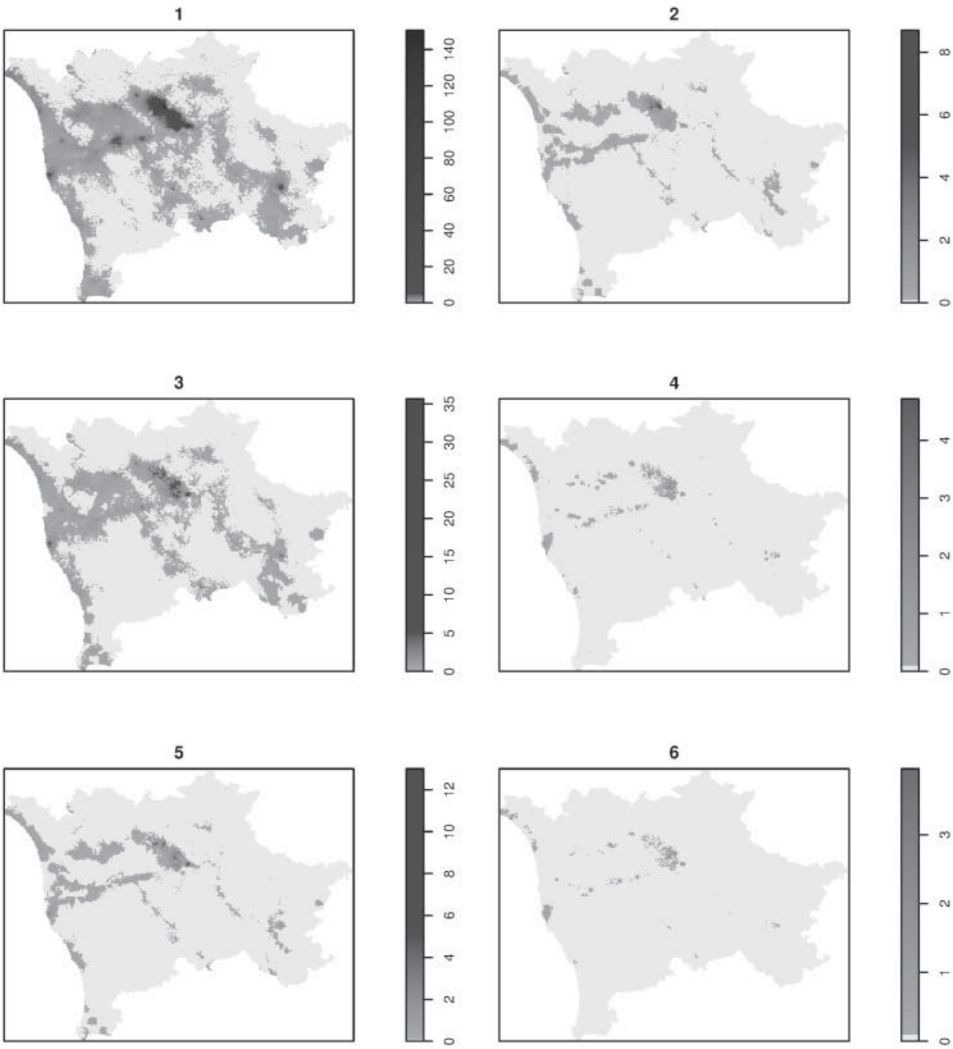

Figure 3. Predicted intensity function for each group of manufacturing firms.

with higher prices, but as a preference for areas more advantageous for the presence of services and infrastructures, which are the areas more requested and consequently more expensive.

Next, from Table 2 we note the positive effect of the spatial covariate x_{10}: newborn firms prefer to locate in areas with a higher density of manufacturing units dead in the period 2004–2007. This could be explained partly by a sort of substitution effect, that is by the choice to re-use newly available spaces and structures previously allocated to other industrial activities. It should be noted, however, that the positive value of the x_{10} parameter could be partly due to the definition of new unit in the ASIA-UL data set, where some variations in the organizational or administrative structure of a firm could be misread as the death of it and the birth of a new one.

The remaining parameters describe the influence of the underlying economic context. The presence of existing manufacturing units of the same sector of economic activity (x_6) has a positive influence for the low-technological firms, which include the more traditional manufacturing sectors that are usually aggregated in specialized districts. The effect on the remaining groups is negative, in particular for the firms with more than 10 employees, which may indicate a competitors effect. In addition, the positive relationship with the existing manufacturing units of a

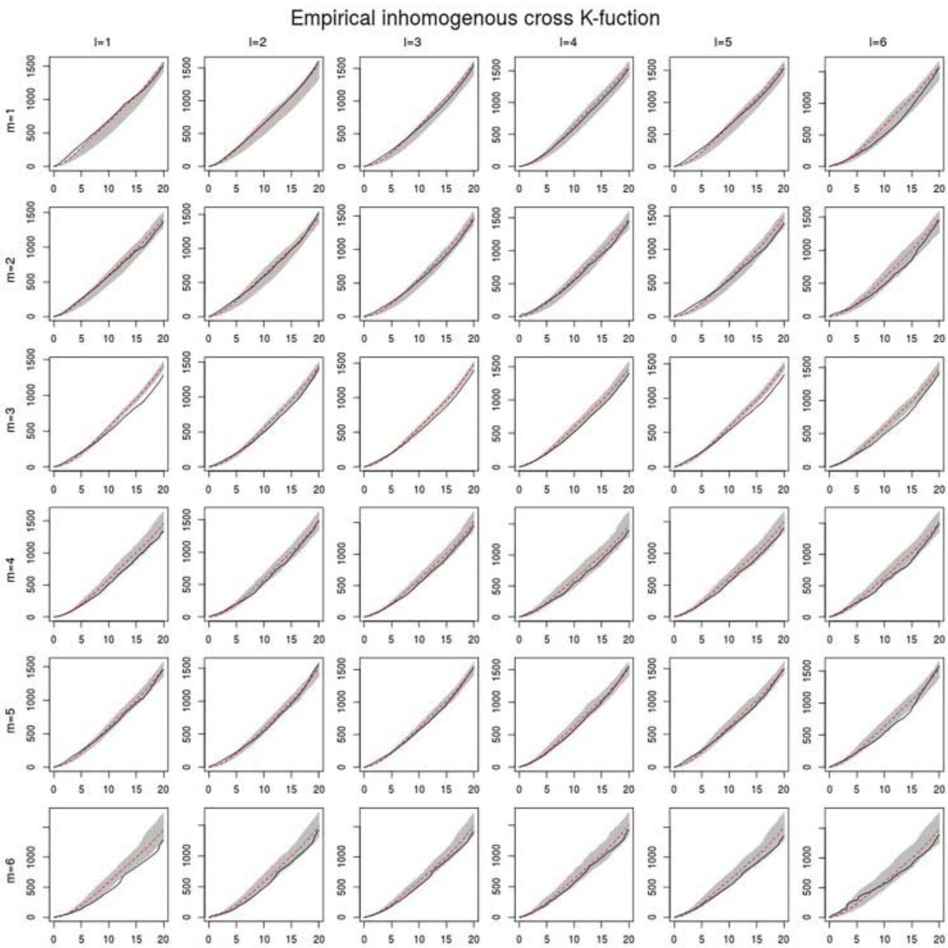

Figure 4. Empirical inhomogeneous cross K-function $\hat{K}^{ml}_{inhom}(r)$ for the observed points (continuous line) with the average (dashed line) and the 99% bands (grey area) derived from 500 simulations generated from of the estimated model.

different sector of economic activity (x_7) and with the existing commercial units (x_8), could indicate the presence of areas dedicated to industrial and commercial activities where newborn firms tend to locate to exploit positive spatial externalities. This result is confirmed also by the negative relationship with the presence of tertiary sector units (x_9), that are usually located in the urban and central areas which are more dedicated to residential and tertiary activities. Finally, the same consideration about the relevance of the multiplicative effect made for x_1 is valid for these last variables: the exponential transformation of their parameters are close to 1 as they measure the variation of intensity corresponding to the presence of one more existing firm near the chosen location. However, it is reasonable to expect that a sensible variation should not correspond to the increment of a single existing firm, but to the presence of a more dense industrial fabric.

It is important to note that model (4) can be used not only to explore the effects of several variables on the location decisions of newborn firms, but also to predict the intensity function in each location inside the study area, as shown in Figure 3. Therefore, it allows us to evaluate the attractiveness of each location for a new firm, conditional to the existing characteristics of the location and differently for each of the six types of manufacturing firms.

Finally, as mentioned in Section 3, we assess the accuracy of our model by applying the Monte Carlo test procedure. We calculate the empirical inhomogeneous cross K-function (3) on the observed data (represented by the continuous line in Figure 4) and the 99% empirical bands derived by generating 500 simulations from the estimated model (4) (indicated by the grey area in Figure 4).

Figure 4 shows the empirical cross-type K-function for the observed data contrasted with the empirical bands for all pairs of firm types. At almost all the distances ranging from 0 to 20 kilometres the empirical K-function lies between the bands, with few exceptions. This can be interpreted as an indication to accept the estimated model as a good description of reality, even if a modest variability in the location decision process seems to remain unexplained.

5. Conclusion

The aim of our paper was to point out how the availability of spatially referenced micro-data on firms can play a key role in understanding the factors that shape the location decisions of new firms. As shown by the recent literature, understanding such determinants is quite relevant to policy-makers as it could be useful to implement effective policies in order to increase an area attractiveness.

In the last years, the increasing diffusion of spatial micro-economic data and of GIS software has brought new methods for the analysis of the spatial dynamics of firms: the analysis point of view has moved to the micro perspective and the spatial information has been exploited to apply methods derived from the stochastic point processes framework [2,15]. This novel approach has been applied in our case study to model the birth process of the small and medium manufacturing firms in Tuscany in order to understand the possible determinants of their location choices.

A second implication of the availability of georeferenced data sets (particularly of georeferenced street addresses) is that it allowed us to easily merge data from different sources by their spatial location. In particular, we were able to merge the firms data with a set of administrative and geostatistical data sets which comprise several spatial-varying variables. This is actually very useful because it allowed us to formulate a model that includes a great number of explicative variables that could strongly influence the location decisions of new firms. More in detail, in our analysis we have considered the birth process of the small and medium manufacturing firms born in Tuscany between 2005 and 2008 as a realization of an inhomogeneous marked point process in the continuous space and we have analysed the effect of several spatial-varying variables on the location decisions of new firms by parametrically modelling the intensity of the process. Results show that the availability of infrastructures, the accessibility, and the presence and characteristics of other existing firms are relevant in the location decision-making of the Tuscan firms. Moreover, the effect of these factors varies with the size and level of technology of the newborn firms.

It is important to note that, differently from Arbia and Arbia *et al.* [2,3] which in their practical analysis considered a small group of firms belonging to a single economic sector, our study comprehends the whole set of small and medium manufacturing firms, characterized by different levels of technological intensity and size. The use of an inhomogeneous marked point process, instead of an unmarked process, allowed us to estimate a specific intensity function for each combination of levels of technological intensity and class of size. The analysis of a big data set of more than 13,000 units, as well as the choice of an inhomogeneous marked point process with several explicative variables, required a high computational capability which were not available a few years ago.

However, as complex as our model is, more could still be added. In particular, in our analysis we have considered only the first of the firm demographic processes, but it will be surely interesting to include a temporal component in the analysis and to consider the growth and the survival of firms, like in [3,15]. This would require a longitudinal data set for all the time-varying variables. In particular, since the ASIA-UL data set is available yearly, it could be possible to collect it for several consecutive years and to merge it with administrative data sets referred to the same time periods. This is not a simple task, as not all the data sets are updated or issued with the same temporal frequency. Future works will surely be devoted to these extensions.

In conclusion, besides the specific Tuscan result, which confirms our theoretical expectations, the study shows the potentiality of the micro-econometric approach based on stochastic point processes for the analysis of the spatial dynamics of firms, nowadays that geographical data, GIS software and computational power are quite diffuse, and even more in the near future.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] J.M. Arauzo-Carod, D. Liviano-Solis, and M. Manjón-Antolín, *Empirical studies in industrial location: An assessment of their methods and results*, J. Reg. Sci. 50 (2010), pp. 685–711.

[2] G. Arbia, *Modelling the geography of economic activities on a continuous space*, Pap. Reg. Sci. 80 (2001), pp. 411–424.

[3] G. Arbia, P. Cella, G. Espa, and D. Giuliani, *A micro spatial analysis of firm demography: The case of food stores in the area of Trento (Italy)*, Empir. Econ. 48 (2015), pp. 923–937.

[4] A. Baddeley and R. Turner, *Practical maximum pseudolikelihood for spatial point patterns*, Aust. N.Z. J. Stat. 42 (2000), pp. 283–322.

[5] A. Baddeley and R. Turner, *Spatstat: An r package for analyzing spatial point patterns*, J. Statist. Softw. 12 (2005), pp. 1–42.

[6] M. Berman and R. Turner, *Approximating point process likelihoods with GLIM*, J. R. Statist. Soc. Ser. C (Appl. Stat.) 41 (1992), pp. 31–38.

[7] F. Bonneu and C. Thomas-Agnan, *Measuring and testing spatial mass concentration with micro-geographic data*, Spat. Econ. Anal. 10 (2015), pp. 289–316.

[8] N. Cressie, *Statistics for Spatial Data*, Revised Edition, Wiley, Hoboken, NJ, USA, 1993.

[9] P. Diggle, *Statistical Analysis of Spatial Point Patterns*, Academic Press, London, UK, 1983.

[10] European Union, *Council Regulation (EEC) No. 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community*, Official Journal of the European Communities L 076 (30/03/1993), pp. 1–11.

[11] EUROSTAT, *Regions: Statistical yearbook 2006*, European Communities, 2006.

[12] EUROSTAT, *NACE Rev. 2 – Statistical classification of economic activities in the European Community*, European Communities, 2008.

[13] ISTAT, *Rapporto annuale 2012*, ISTAT, 2012.

[14] J. Moller and R.P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall/CRC, Boca Raton, FL, USA, 2003.

[15] S.L. Rathbun and N. Cressie, *A space-time survival point process for a longleaf pine forest in southern georgia*, J. Amer. Statist. Assoc. 89 (1994), pp. 1164–1174.

[16] B.D. Ripley, *Modelling spatial patterns*, J. R. Statist. Soc. Ser. B (Methodol.) 39 (1977), pp. 172–212.

[17] O. Schabenberger and C.A. Gotway, *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL, USA, 2005.

[18] D. Strauss and M. Ikeda, *Pseudolikelihood estimation for social networks*, J. Amer. Statist. Assoc. 85 (1990), pp. 204–212.