SIS 2014

47th Scientific Meeting of the Italian Statistical Society

Cagliari - June 11/13, 2014

PROCEEDINGS

Editors:
S. Cabras, T. Di Battista and W. Racugno

CUEC EDITRICE

SIS 2014

47th Scientific Meeting of the Italian Statistical Society

Cagliari - June 11/13, 2014

PROCEEDINGS

Editors:
S. Cabras, T. Di Battista and W. Racugno

CUEC EDITRICE

ISBN: 978-88-8467-874-4
47th SIS Scientific Meeting of the Italian Statistical Society
Cagliari, June 11-13, 2014.

CUEC Cooperativa Universitaria Editrice Cagliaritana
via Is Mirrionis 1, 09123 Cagliari
Tel. e Fax 070271573
www.cuec.eu
info@cuec.eu

ISBN 978-88-84-67-874-4

9 788884 678744

Modeling the location decisions of firms

Un modello per le scelte localizzative delle imprese

Chiara Bocci, Emilia Rocco, and Patrizia Lattarulo

Abstract The paper analyzes the birth process of small and medium manufacturing firms in Tuscany, an Italian region. In particular it explores, through a spatial micro-econometric approach, the possible determinants of the location decisions of the new firms. The geographical distribution of the manufacturing firms born in Tuscany between the 2005 and the 2008 is defined in terms of a inhomogeneous marked point process in the continuous space and we evaluate the effect of space-varying factors, both exogenous and endogenous, on the location decisions of new firms by parametrically modeling the intensity of the process. Results show that the choice is influenced on the one hand by the availability of infrastructures and the level of accessibility, and on the other by the presence and characteristics of existing firms.
Abstract *Il lavoro analizza il processo di nascita delle piccole e medie imprese manifatturiere in Toscana. In particolare, si esplorano le possibili determinanti delle scelte localizzative delle nuove imprese mediante un approccio micro-econometrico spaziale. La distribuzione geografica delle imprese manifatturiere nate in Toscana tra il 2005 ed il 2008 viene definita come un processo di punto marcato inomogeneo nello spazio continuo e viene valutato l'effetto di fattori spazialmente variabili, sia esogeni che endogeni, sulle scelte localizzative costruendo un modello parametrico per l'intensita del processo. I risutati mostrano come tale scelta sia influenzata da una parte dalla disponibilità di infrastrutture e dalla facilità di accesso e dall'altra dalla presenza e dalle caratteristiche delle imprese già esistenti sul territorio.*

C. Bocci · P. Lattarulo
IRPET - Regional Institute for Economic Planning of Tuscany,
Via Dazzi, 1 - 50141 Firenze, Italy
e-mail: chiara.bocci@irpet.it, patrizia.lattarulo@irpet.it

E. Rocco
Department of Statistics, Informatics, Applications "G. Parenti" (DiSIA), University of Firenze,
Viale Morgagni, 59 - 50134 Firenze, Italy
e-mail: rocco@ds.unifi.it

1 Introduction

In the last years location and physical geography characteristics have become relevant factors in economic studies. Moreover, the increasing availability of spatially referenced data jointly with the diffusion of GIS softwares, has allowed the development of new methods for the analysis of spatial dynamics using micro-economic data. In such setting, the framework of spatial point process methods (Diggle 2003) plays a major role since spatial point processes can be used directly to model and analyze data which takes form of a spatial point pattern, such as the geographical distribution of firms (Arbia 2001).

In this paper we adopt this approach to analyze the birth process of small and medium manufacturing firms in Tuscany and the possible determinants of their location decisions. In order to achieve this aim we structure the paper as follows. Section 2 describes the data while Section 3 is devoted to the statistical framework and the parametric model formulated for the analysis. Finally, results and conclusions are presented in Section 4.

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 location, sector of economic activity and number of employees of each enterprise local unit. The field of observation of ASIA-UL covers all industrial, commercial and service-sector private activities.

For our analysis we consider the local units born in Tuscany in the period 2005-2008, that is, the units not existent in ASIA-UL 2004 and registered in the period 2005-2008. We limit the analysis to the roughly 13000 manufacturing firms, classified accordingly to the technological intensity of their production[1] and on their size. Three levels of technology (Low, Medium-low and Medium-high) and two levels of size (less than 10 employees, 10 or more employees) are combined to obtain six groups of firms. We restrict the geographical area of analysis to the North-Central area of Tuscany, that comprehend most of the manufacturing activity in Tuscany (92% of the local units and 93% of employees).

A map depicting the location of the 13196 new manufacturing local units for each of the 6 groups is presented in Figure 1. There is a clear tendency of the firms to locate in specific sub-areas, that correspond to the main cities and to the main infrastructures of the region. In addition, the number of units in each group varies

[1] Istat and Eurostat classify the manufacturing firms according to their technological intensity (R&D spending/value added) using the classification of economic activities in the European Community (NACE) at the 2-digit level. They define four manufacturing groups: High technology, Medium-high technology, Medium-low technology and Low technology. In our analysis, due to the small amount manufacturing firms with High and Medium-high technology, we consider them as a unique group.

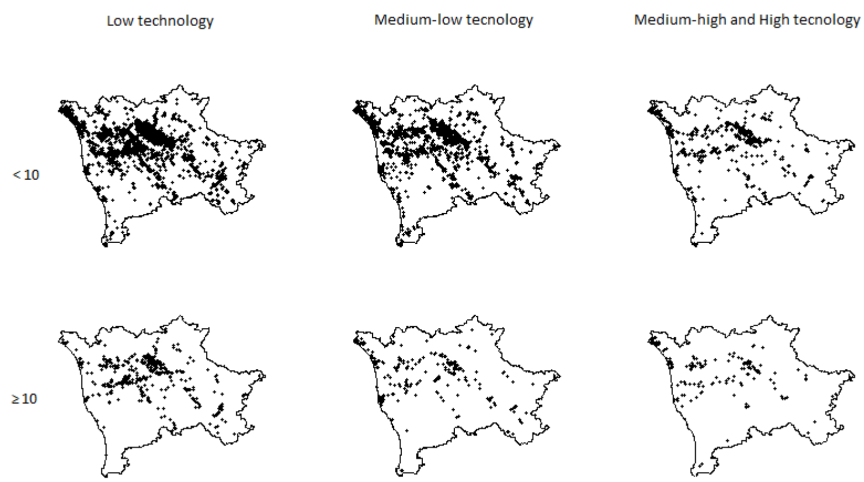

Fig. 1 Spatial distribution of the 13196 new manufacturing local units born in 2005-2008 divided in 6 groups: three levels of technology (Low, Medium-low and Medium-high) and two levels of size (less than 10 employees, 10 or more employees).

significantly: most of the observations belongs to the low and medium-low technology firms with less than 10 employees.

In order to evaluate the possible determinants of the location decisions of the new firms, we linked the geographical coordinates of our observations to a set of spatial-varying variables, collected from administrative datasets. A first group of variables describes the characteristics of the location: the price of industrial buildings (euro/m^2) (x_1), the minimum distance to the train station or to the highway (km) (x_2), the adsl coverage (dummy) (x_3), the degree of urbanization of the corresponding municipality[2] (dummy) (x_4) and the terrain slope (x_5). A second group of variables characterizes the existing economic context: considering a radius of 5 km, for each new local unit we count the number existing manufacturing units of the same sector of economic activity (x_6), the number of existing manufacturing units of a different sector of economic activity (x_7), the number of existing commercial units (x_8) and the number of existing tertiary sector units (x_9); in addition, the density of manufacturing units dead in the period 2004-2008, that is the units that exist in ASIA-UL 2004 but not in ASIA-UL 2008, is also considered (x_{10}).

[2] Eurostat defines the "degree of urbanization" concept that distinguishes three types of zones: densely populated area (high), intermediate area (medium) and thinly populated area (low). For a description of how these types are identified we refer "Regions: Statistical Yearbook 2006" published by Eurostat. In our analysis consider a classification in two levels: medium-high and low urbanization.

3 Model framework

Let us define A as the study area, $\mathbf{s} = [s_1, s_2]$ ($s_i \in R^2$) as the spatial coordinates of the units, N as the number of units located in A, M_i ($i = 1, \ldots, N$) the mark of the i-th unit (that indicates to which of the classes of our data it belongs) and $\delta(\mathbf{s})$ as a region in the neighborhood of the point of coordinates $\mathbf{s}$.

Then we define the local density of the spatial process (know as intensity) (Diggle 2003, Ratbun and Cressie 1994) at a point $\mathbf{s}$ as

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

which represents the expected number of units in the infinitesimal area.

If the process is stationary, then $\lambda(\mathbf{s})$ will be constant. However in real situations some places could be more likely to be chosen than others, for various economic reasons, and this will produce an irregular pattern. To model such a situation it is useful to consider the birth process of the firms as a non-stationary process with spatial intensity that depends from economic characteristics and varies according to location (Arbia 2001). High values of $\lambda(\mathbf{s})$ indicate a concentration of economic activities in the infinitesimal area centered in $\mathbf{s}$. Moreover, by considering a marked point process we can estimate a different intensity for each class of our data.

More explicitly, the geographical distribution of the manufacturing firms born in Tuscany between the 2005 and the 2008 is assumed to be a realization of a inhomogeneous marked Poisson point process (Diggle 2003). On the basis of this assumption, we model the intensity of the process in the point of coordinates $\mathbf{s}$ defining the following parametric structure:

$$\lambda(\mathbf{s}) = \exp(\alpha_0 + \alpha_1 d_2(\mathbf{s}) + \alpha_2 d_3(\mathbf{s}) + \alpha_3 d_4(\mathbf{s}) + \alpha_4 d_5(\mathbf{s}) + \alpha_5 d_6(\mathbf{s}) + \beta \mathbf{X}_a(\mathbf{s}) +$$
$$+ \gamma_1 d_2(\mathbf{s})\mathbf{X}_b(\mathbf{s}) + \gamma_2 d_3(\mathbf{s})\mathbf{X}_b(\mathbf{s}) + \gamma_3 d_4(\mathbf{s})\mathbf{X}_b(\mathbf{s}) + \gamma_4 d_5(\mathbf{s})\mathbf{X}_b(\mathbf{s}) + \gamma_5 d_6(\mathbf{s})\mathbf{X}_b(\mathbf{s}))$$

where $d_2(\mathbf{s})$, ..., $d_6(\mathbf{s})$ are dummy variables that indicates the mark of the unit located in $\mathbf{s}$ (mark 1, that correspond to the low technology manufacturing with less than 10 employees, is the base reference); $\mathbf{X}_a(\mathbf{s}) = [x_1(\mathbf{s}), \ldots, x_{10}(\mathbf{s}), x_{11}(\mathbf{s})]$ is the vector of the 10 explicative variables described in the previous section plus the interaction term $x_{11}(\mathbf{s}) = x_4(\mathbf{s}) * x_7(\mathbf{s})$ and $\beta = [\beta_1, \ldots, \beta_{11}]$ is the vector of the relative coefficients; $\mathbf{X}_b(\mathbf{s}) = [x_2(\mathbf{s}), x_6(\mathbf{s}), x_8(\mathbf{s}), x_9(\mathbf{s})]$ is the vector of the variables that interact with the marks and γ_1, ..., γ_5 are the corresponding vectors of coefficients.

The logarithmic transformation allows to fit the model by maximizing the log-pseudo-likelihood for $\lambda(\mathbf{s})$ based on the observed points of the pattern under study. To maximize the log-pseudo-likelihood and to obtain the estimates of parameters, we adopt the method proposed by Berman and Turner (1992) and discussed in detail by Baddeley and Turner (2000) which is implemented in the package `spatstat` in the `R` computing environment. Moreover, in the case of a Poisson stochastic process maximum pseudo-likelihood is equivalent to maximum likelihood, therefore it is possible to test the goodness of fit of the adopted model by using standard formal

Table 1 Estimated parameters.

Parameter	Estimate	S.E.	Z-test	Parameter	Estimate	S.E.	Z-test
α_0	-3.00407	0.08711	***	$\gamma_{1,8}$	-0.00028	0.00010	**
α_1	-1.65298	0.09283	***	$\gamma_{1,9}$	0.00002	0.00006	
α_2	-0.44049	0.05289	***	$\gamma_{2,2}$	-0.10391	0.01329	***
α_3	-2.40054	0.13954	***	$\gamma_{2,6}$	-0.00164	0.00007	***
α_4	-1.65490	0.09252	***	$\gamma_{2,8}$	0.00024	0.00004	***
α_5	-2.68506	0.16353	***	$\gamma_{2,9}$	-0.00016	0.00002	***
β_1	0.00054	0.00002	***	$\gamma_{3,2}$	-0.22691	0.04372	***
β_2	-0.02843	0.00656	***	$\gamma_{3,6}$	-0.00207	0.00022	***
β_3	1.36413	0.05761	***	$\gamma_{3,8}$	0.00059	0.00012	***
β_4	1.21120	0.05859	***	$\gamma_{3,9}$	-0.00039	0.00007	***
β_5	-0.08945	0.00218	***	$\gamma_{4,2}$	-0.19888	0.02785	***
β_6	0.00086	0.00002	***	$\gamma_{4,6}$	-0.00158	0.00011	***
β_7	0.00292	0.00028	***	$\gamma_{4,8}$	0.00036	0.00007	***
β_8	0.00036	0.00003	***	$\gamma_{4,9}$	-0.00020	0.00004	***
β_9	-0.00027	0.00001	***	$\gamma_{5,2}$	-0.24206	0.05270	***
β_{10}	0.29360	0.00611	***	$\gamma_{5,6}$	-0.00270	0.00035	***
β_{11}	-0.00252	0.00028	***	$\gamma_{5,8}$	0.00055	0.00013	***
$\gamma_{1,2}$	-0.13378	0.02458	***	$\gamma_{5,9}$	-0.00033	0.00007	***
$\gamma_{1,6}$	0.00012	0.00007					

likelihood ratio criteria and the χ^2 distribution. In our study we implemented several possible models for $\lambda(\mathbf{s})$, then we applied the likelihood ratio test in order to select the best model in terms of parsimony and accuracy of the estimates.

4 Results

Table 1 presents the estimated parameters of our model. Most of the coefficients are highly significant, and from their values we are able to identify the influence of each variables on the location decisions of the six groups of new local units.

β_2, $\gamma_{1,2}$, $\gamma_{2,2}$, $\gamma_{3,2}$, $\gamma_{4,2}$ and $\gamma_{5,2}$ indicates the negative relationship with the distance to the train station or to the highway. Such relationship is important for all the groups, but it become even stronger for the medium-high technological firms and for the units with more than 10 employees. This result is in accordance with 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 value of β_5 indicates that a friendlier environment (a more flat terrain) is preferable. Moreover, the values of β_1, β_3, β_4 show that the availability of infrastructures (summarized by the degree of urbanization, the adsl coverage and the price of industrial buildings) has a positive influence on the location choices of all the manufacturing firms.

The remaining parameters describe the influence of the underlying economic context: the presence of existing manufacturing units of the same sector of economic activity 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

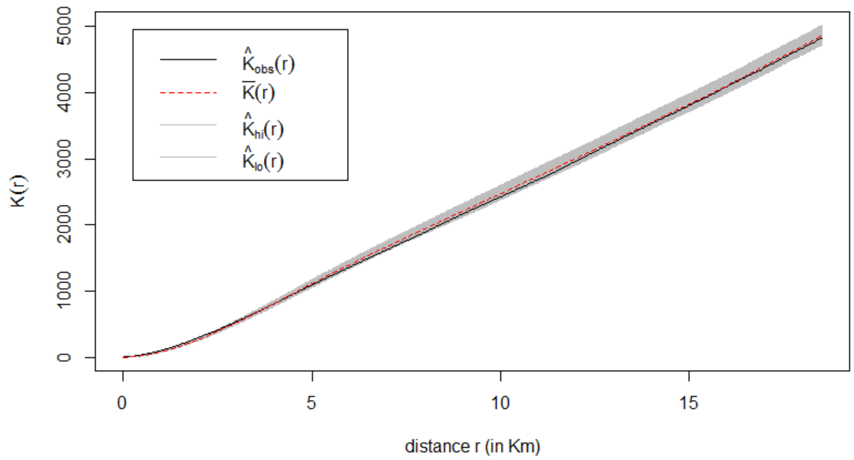

Fig. 2 Empirical K function for the observed points contrasted with the 99% bands derived from 200 simulations generated from of the estimated model

with more than 10 employees, which may indicate a competitors effect. On the other hand, the presence of manufacturing units of a different sector of economic activity and of commercial units has a positive influence, that could be due to the presence of industrial areas where the new firms tend to locate, due to the action of positive spatial externalities, in contraposition to urban and central areas that are more dedicated to residential and tertiary activities.

The accuracy of the fit of our model can be assessed with a Monte Carlo test. We used a procedure based on the visual inspection of the empirical K function for the observed points contrasted with the 99% bands derived from 200 simulations generated from the estimated model (Arbia 2001). At all the distances ranging from 0 to 20 kilometers the empirical K function lies between the bands. We interpret this as an indication to accept the estimated model as a good description of reality.

References

1. Arbia, G.: Modelling the geography of economic activities on a continuous space. Papers Reg. Sci. **80**, 411–424 (2001)
2. Baddeley, A., Turner, R.: Practcal maximum pseudolikelihood for spatial point patterns. Austalian and New Zealand Journal of Statistics **42**, 283–322 (2000)
3. Berman, R., Turner, T.R.: Approximating point process likelihoods with GLIM. Applied Statistics **41**, 31–38 (1992)
4. Diggle, P.J.: Statistical analysis of spatial point patterns,2nd edition. Edward Arnold, London (2003)
5. Ratbun, S.L., Cressie, N.: A space time survival point processfor a longleaf pine forest in southern Georgia. Journal of the american Statistical Association **89**, 1164–1174 (1994)