DETECTION OF FIRMS′ CLUSTERING BY LOCAL SCALING

Ilona BERKOVÁ
University of South Bohemia in České Budějovice, Faculty of Economics
berkova@ef.jcu.cz

Tomáš MRKVIČKA
University of South Bohemia in České Budějovice, Faculty of Economics
mrkvicka@ef.jcu.cz

Renata KLUFOVÁ
University of South Bohemia in České Budějovice, Faculty of Economics
klufova@ef.jcu.cz

Radim REMEŠ
University of South Bohemia in České Budějovice, Faculty of Economics
remes@ef.jcu.cz

Abstract

The paper analyses locations of headquarters of companies and their interactions by inhomogeneous point process, especially local scaling principles, because companies choose their locations according to the number of the local population. Relationships of the companies within economic sectors are studied using the analysis of locally scaled L function. The inhomogeneity was modelled using the local population, then the company's size was included. Lastly the level of clustering in each sector was computed. The companies are located in three regions in the Czech Republic. It was found out that the companies tend to cluster when the population or the companies' size is taken into account.

Keywords: Inhomogeneous point process, L-function, Global envelope test, Spatial clusters, Agglomeration

JEL classification: C21, L60, O18, R12

1. Introduction

The localization of corporate activities is the most important decision for a company and is the result of many factors and decision-making of actors such as companies, households and the public sector. Due to this fact the location of companies has been a part of economics for many years. For decision making about the companies' location, it is necessary to take into account the external and internal factors that influence activities of the companies and the optimal combination of these factors leads to optimal localization. One of these factors is the location where the company operates. This fact has led to the development of location theory that is one of the oldest theories dealing with regional economics (Alonso, 1972). Localization theory is focused on the geographic location of economic activities and it tries to answer the question why and where economic activities are placed and try to define location factors which can explain the decisions of firms (North, 1955). Each author of location theory introduces different location determinants. For example, Weber (1929) identifies transportation costs, labour costs and consumer agglomerations as the main localization factors. In addition, Weber differentiates factors according to the company sector on the socio-political, natural-technical and socio-cultural factors and dispersion of economic activities on regional and agglomerative factors. Ježek (2002) considers the natural resources, labour, suppliers of goods and services, information and access to information as the main location factors. According to Kuşluvan (2013) the location is determined by technological, economical and geographical, political and social factors. Christofakis (2014) considers, as the most important determinant, transport and infrastructure costs and costs associated with transport services. However, the division of location factors into soft and hard can be regarded as crucial. In general, it can be said that hard localization factors are those that directly affect business activities and can be directly calculated (acquisition of property, labour force, etc.).

Soft localization factors, on the contrary, have an indirect or very small impact on an enterprise and are not recorded in accounting documents (quality of life, population education...) (Damborský and Wokoun, 2010). Investors, however, attribute different weight and motivation to these factors.

Localization theories have undertaken huge development due to various changes concerning the world economy, the environment, and also globalisation (Fujita, 2010). The origin of localization theories can be dated to the early 19th century to Germany, when there was a great development of industry and agriculture. As the founders of the theories are considered Johann-Heinrich von Thünen (1826), William Alonso (1972) and Alfred Weber (1929), whose models served as a basis for localization theories, and were broadened to suit the needs of geographers, economists and regional scientists. The first localization theories focused on agriculture activities (e.g. Stevens, 1968; Alonso, 1972; Berry and Harris, 1970) which were then expanded to industry (e.g. Weber, 1929; Krugman and Lawrence, 1993).

The further development of localization theories has expanded since the second half of the 20th century with the use of multi-criteria approach and modelling (Rumpel et al., 2008) and due to the large-scale globalization. There is much research dealing with cluster modelling of economics activities. The localization theories have been expanded in particular by the features of foreign activities such as the exchange rate, political risk, transnational policy and politics, and cultural differences (Popovici, Călin, & others, 2014). K-function has become a popular tool used for the recognition of spatial behaviour (Dixon, 2002). This field is the interest of Espa et al. (2010), Marcon and Puech (2003), Duranton and Overman (2005), Quah and Simpson (2003) and Arbia et al. (2008) who popularized the use of K-function when analysing companies' locations. Arbia et al. (2010) expanded the cluster analysis of companies adding time perspective using space-time K-function.

All localization theories have been the subject of criticism over the years as it is based on many assumptions that lead to generalization, unrealism, etc. However, removing these assumptions is unrealistic due to the variation of environment and conditions and many influential factors.

The spatial distribution of companies is not homogeneous, because the probability of hosting companies is not geographically constant. Arbia et al. (2012) consider inhomogeneous space and assume that a company's location choice depends on physical or administrative constraints. In our research, we presume that taking homogeneous economy space into account is not realistic. We assume that the main source of inhomogeneity is caused by the given population. Until this time, the inhomogeneity with respect to the companies' location has been solved using different methods. For example, Sweeney, Feser (1998) and Marcon, Puech (2003) used D-function that considers density variations. They tested whether small companies were more concentrated than large ones. Another approach to solve non-stationarity used inhomogeneous K-function (e.g. Arbia et al., 2012) that is a generalization of Ripley's function which assumes second-order intensity-reweighted stationarity (Baddeley et al., 2000). In this paper, we propose to solve the inhomogeneity by the method of local scaling (Hahn et al., 2003). We believe that modelling the inhomogeneity using local scaling is, in the case of companies' locations more realistic, because the local scaling deforms the original Euclidian space into space where the distances are governed by the assumed inhomogeneity. In our case it is governed by population and then by size of firms. Particularly, it allows us to vary the range of interactions in uneven dense space of companies. Although, the second order reweighting stationarity is frequently used to model inhomogeneity, we prefer the use of the L-function, which is a variance stabilizing transformation of K-function, as a tool for our modelling purposes since it is one of the most commonly used summary characteristic of point patterns.

The main aim of this paper is to reveal whether the positions of companies' headquarters are spatially dependent or independent. We tackle the problem of spatial heterogeneity with the population given in observed window, because homogeneity leads to unreality due to existence of natural features and settlement. Thus as the first step of the analysis, we test if the data can be modelled using locally scaled point process with inhomogeneity governed by population. In the other step, we add to the inhomogeneity also the size of the companies and perform the same test.

This paper is organized as follows. First of all, in Section 2, we introduce methodological statistical framework, especially methodology of Ripley's K-function (Ripley, 1976), in empirical analysis preferred Besag´s L-function (Besag, 1977), inhomogeneous spatial point processes and Global Envelopes. Chapter 3 contains data description and empirical application of the methodology in inhomogeneous case. At the end of paper in Section 4 there are the conclusion and discussion of research and our next steps in future studies in this field.

2. The statistical methodological framework

Companies can be established at different location. To find out a spatial phenomena of the companies, we have to introduce a statistical test that provides information about behaviour of the companies in space. In this section, we introduce Ripley’s K-function and its derived Besag’s L-function that is used for determination of the distribution of the companies in our research. Then, we explain the inhomogeneous point process, especially the method of local scaling.

2.1. K-function analysis

It was considered that companies’ positions form a point process. The most important activity in point processes is to summarize data sets by numerical and functional characteristics. The second-order characteristics offer a way to present statistical information about interactions among the points in different distances. Probably the most commonly used and the most popular functional second-order summary characteristics for the analysis of point patterns are Ripley’s K function K(r), Besag’s L-function L(r) and the pair correlation function g(r). Illian et al. (2008) believe that these distance-based functions are more powerful than the other summary characteristics because of their way of statistical presentation of distributional information of point patterns. Further L function provides the easiest interpretation because of its linear form.

Ripley’s K-function was proposed by B. D. Ripley and describes the spatial dependence between events in point patterns (Ripley, 1976). This function calculates the expected number of additional events located in a ball surrounding a randomly chosen event and quantifies spatial dependence and clustering (e.g. Diggle (1983) and Ripley (1976)).

In homogeneous case the density, denoted λ, is considered to be constant The K-function (Ripley, 1976) is defined as:

$$K(r) = \lambda^{-1} E(number\ of\ points\ falling\ at\ a\ distance \leq r\ from\ an\ arbitrary\ point)$$

, where E(.) indicates the expectation operator and λ (intensity) represents the mean number of events per area. λK(r) can be interpreted as the expected number of points within a distance r of an arbitrary point of the process. The empirical homogeneous K-function is defined as

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} 1\{d_{ij} \leq r\},$$

where $|W|$ is total study area, d_{ij} Euclidean spatial distance between the i^{th} and j^{th} observed points.

For complete spatial randomness (points are distributed completely randomly and independently in the area, abbreviated by CSR), K-function is equal to $K(r) = \pi r^2$, for $r > 0$. Significant deviations from this hypothesis represent alternative hypothesis e.g. clustering for $K(r) > \pi r^2$, for $r > 0$ or inhibition for $K(r) < \pi r^2$, for $r > 0$ (Ripley, 1976).

To determine whether the distribution of companies is significantly different from CSR, L-function is commonly used. The L-function is a transformation of K-function proposed by Besag (1977) and presents the same information as K-function and has graphical advantages. The L-function in two-dimensional case is:

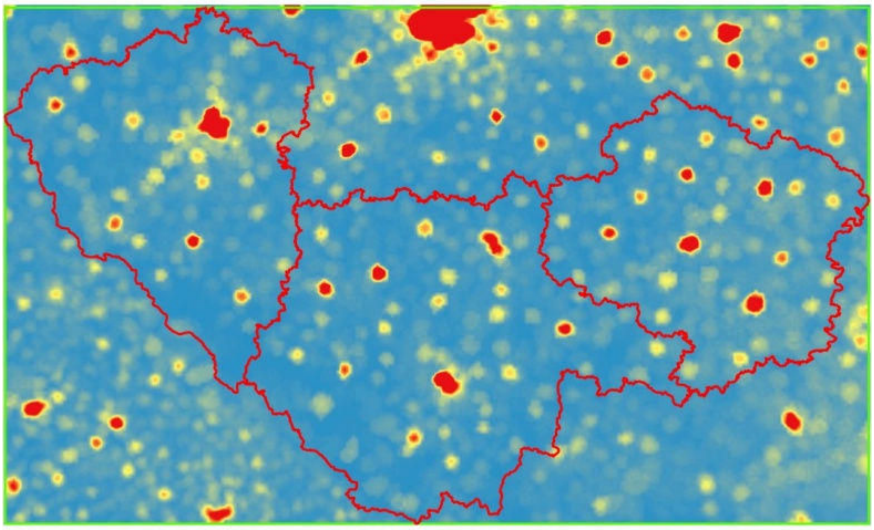
$$L(r) = \sqrt{\frac{K(r)}{\pi}} \text{ for } r \geq 0.$$

The practical interpretation of *L(r)* is *L(r) = r* leads to CSR, *L(r) > r* indicates clustering of point pattern while *L(r) < r* indicates dispersion of point (Illian et al., 2008) in the interpoint distance *r*.

2.2. Inhomogeneous spatial point patterns

We cannot consider that the density of companies is the same in the whole observation window so we suppose that the location of the companies depends on the population (Figure 1) and the size of the companies (Figure 2) in the given area. This approach is more realistic in large observation areas and areas with geographical features like mountains where concentration of companies is not as common. From this reason, it was necessary to use tools for inhomogeneous analysis.

Figure 1. Population density



Source: Own processing

The Population density in the given area is depicted in Figure 1. The observation area is highlighted with the red curve. The values with low population density are blue and with high population density are displayed in red. The size of the companies belonging to the sector of agriculture, forestry and fishing and their positions are depicted in Figure 2 below.

For inhomogeneous point processes, various models differing in the specification of how the interactions between points depend on the local intensity of points have been suggested. We will use local scaling for modeling the inhomogeneity (Hahn et al., 2003). This approach yields models for patterns that are homogeneous up to the local scale factor. The inhomogeneity is obtained by local scaling of the template process with a location-dependent scaling factor (in our study it is the population). If the scaling factor is constant, then the point process behaves like a template.

The main aim of local scaling is to find global summary characteristics which are adapted to variable point density by a mechanism of rescaling distances relative to local point density. This is achieved by replacing distance measures used in the density with locally scaled analogs defined by a location dependent scaling function (Hahn et al., 2003). Due to local scaling pattern, distances become shorter in the regions with low population density and longer in the regions with high population density.

Locally scaled version of K-function modifies distances for each pair of points x_i, x_j by rescaling factor $s(x_i, x_j)$. The rescaled distance for each pair of data points xi, xj is defined as $d_{ij}^* = \frac{\|x_i - x_j\|}{s(x_i, x_j)}$, where the rescaling factor is computed as $s(x_i, x_j) = \frac{1}{2}\left(\frac{1}{\sqrt{\hat{\lambda}(x_i)}} + \frac{1}{\sqrt{\hat{\lambda}(x_j)}}\right)$ (Baddeley et al., 2015).

The most common way to find out differences of the empirical distribution of companies from a given null model is by using an exploratory tool called envelope tests that are often used in spatial statistic and were introduced by Besag (1977) and Ripley (1976). However, in our study, Global envelope tests are used as they are more exact and also offer a graphical interpretation (Myllymäki et al., 2017). These tests generate an acceptance band by computing L-function for n simulated patterns of the null model, i.e. inhomogeneous Poisson processes with the same intensity and the same number of points as the observed pattern. The Global envelope tests reject the null hypothesis if the observed L function is not completely inside the envelope. Their undeniable advantages are that they allow the selection of α and they yield p-values and provide graphical representation. The significance level α = 0.05 was used in the analysis.

Global envelope tests have two approaches primary depending on a selected number of simulations (Myllymäki et al., 2017). The first approach Global rank envelope test as having a better performance because its bounds are constructed directly from the functions. On the other hand, it is necessary to use it with an appropriate number of simulations. The second approach Global scaled maximum absolute difference (SMAD) envelope test is not as accurate as the first approach because bounds are parameterized by the r-wise variance or quantiles. The advantage of this approach is that it does not need a large number of simulations. In this paper, we used Global SMAD envelope test, concretely Direction quantile MAD envelope test with 99 simulations because of a time limitation.

In case of Global SMAD envelope test (Myllymäki et al., 2017), the critical bounds were calculated as follows $T_{low}^{u}(r) = T_0(r) - u \times |\underline{T}(r) - T_0(r)|$ and $T_{upp}^{u}(r) = T_0(r) - u \times |\overline{T}(r) - T_0(r)|$, where $\overline{T}$ and $\underline{T}$ denote the r-wise 2.5% lower and upper quantiles of the distribution of T(r) under null hypotheses. T(r) denotes functional statistics in our case $L^*(r)$, i.e. the locally scaled version of L-function. The critical bounds are parametrized with respect to u, where u is found to correspond to required global level of significance α.

Lastly, we wanted to compare the tendencies towards clustering between the sectors so the level of clustering was determined in both analyses. The level of clustering was defined as $\frac{L^*(r)-L^*_{central}(r)}{L^*_{upp}(r)-L^*_{central}(r)}$, where $L^*_{central}(r)$ is the value of the estimated L-function in a given argument r obtained for the null model, i.e. inhomogeneous Poisson process of given sector and $L^*_{upp}(r)$ is the value of the simulated upper band of envelope for the null model in a given argument R. The argument of interest was chosen to be equal to the rescaled distance equal to 0.25.

3. Results

3.1. Data description

In our empirical analysis, we used a set of companies in three regions located in the Czech Republic, i.e. Jihocesky, Plzensky and Vysocina regions. These regions were chosen not to effect the results because of the similar characteristics in these regions. The main source of economic wealth is primary sector, further economic-social level in these regions is almost the same, especially the dynamic of development and the quality of life (Martinčík, 2008). The data set was collected in 2015 by database Albertina Gold and contains information from the financial reports of the companies from the year 2013.

The classification of the companies into the given sectors was selected using the CZ-NACE methodology according to the core business that is the main product of the companies. Based on their economic activities, the companies were divided into 13 sectors. The data set contains 10 201 companies and their full addresses. Some descriptive analysis to understand the localization pattern is displayed in the table 1 below.

Table 1. Descriptive analysis of the observed window

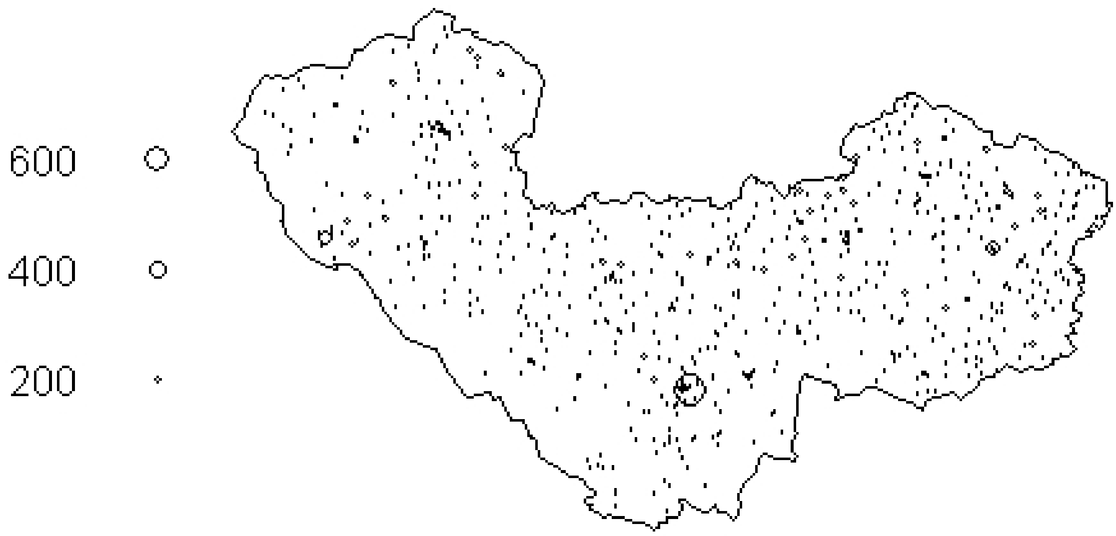
	Region	Number of companies	Mean of employees
Agriculture, forestry and fishing	Total	747	35,1
	Jihocesky	312	29,5
	Plzensky	163	34,6
	Vysocina	272	41,3
Mining and quarrying	Total	20	72,8
	Jihocesky	10	32,3
	Plzensky	7	123,4
	Vysocina	3	62,7
Manufacturing industry	Total	2193	83,3
	Jihocesky	862	66,5
	Plzensky	622	92,6
	Vysocina	709	90,8
Production and distribution of electricity, gas and water	Total	245	34,5
	Jihocesky	114	40,0
	Plzensky	70	32,8
	Vysocina	61	30,6
Construction	Total	1234	18,0
	Jihocesky	570	18,2
	Plzensky	325	17,2
	Vysocina	339	18,7
Wholesale and retail trade, repair of motor vehicles	Total	2127	15,8
	Jihocesky	919	16,1
	Plzensky	586	19,3
	Vysocina	622	11,9
Transport, storage and communication	Total	439	8,6
	Jihocesky	227	7,6
	Plzensky	130	10,8
	Vysocina	82	7,3
Accommodation and food service activities	Total	653	32,3
	Jihocesky	277	31,4
	Plzensky	213	35,0
	Vysocina	163	30,4
Financial intermediation	Total	75	7,6
	Jihocesky	26	4,7
	Plzensky	28	11,8
	Vysocina	21	6,2
Real estate activities, renting and business activities	Total	1737	12,8
	Jihocesky	914	8,6
	Plzensky	447	20,9
	Vysocina	376	9,0
Education	Total	102	9,4
	Jihocesky	50	7,4

	Region	Number of companies	Mean of employees
	Plzensky	19	8,3
	Vysocina	33	12,4
Health and social care, veterinary activities	Total	458	28,4
	Jihocesky	181	53,8
	Plzensky	172	18,8
	Vysocina	105	12,6
Other community, social and personal services	Total	171	17,3
	Jihocesky	77	16,7
	Plzensky	63	24,9
	Vysocina	31	10,3

Source: Own processing

We analysed all the sectors in the observed area but in this paper, we introduce results only for the sector of agriculture, forestry and fishing which has 746 companies. In this sector, there is 286 micro companies (less than 10 employees), 291 small companies (less than 50 employees), 167 medium companies (less than 250 employees) and 3 big companies (more than 251 employees). The size of the companies is displayed in Figure 2. This sector was chosen as it is the most important sector in the observed area.

Figure 2. Size and positions of companies belonging to sector agriculture, forestry and fishing



Source: Own processing

The spatial distribution of the companies together with their sizes in this sector is displayed in Figure 2. In the first visual inspection, it is clear that the companies tend to make clusters by concentration on some specific positions in the observed window.

To take inhomogeneity into account, we generated a matrix population for each 2x2 km in the observed window in Geographic Information System. The population values were given for the year 2011 and interpolated by Inverse Distance Weighing Method with the power 0.3. The population density is displayed in Figure 1.

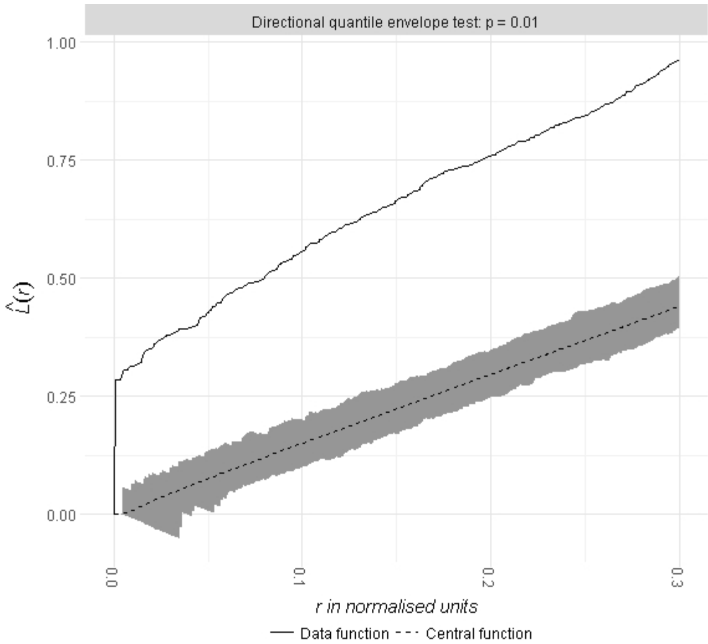
4. Analysis

In the analysis we considered these three regions as one observation window due to similar conditions in these regions mentioned above. The observed area is characterized by variant natural features like mountains, rivers, forests etc. At the first glance we can see that companies are more concentrated in areas of towns. From that reason, we tested spatial behaviour of the companies in inhomogeneous case where inhomogeneity was firstly given by population density. Secondly, we assumed that bigger companies drain away more workers from population and it leads to less available workers for establishment of a new company. Further according to Ježek (2002) factors affecting the foundation of enterprises rank the sectoral structure of the local or regional economy, the prevailing size of enterprises, the educational level of the workforce and regional business traditions. Therefore, we spatially smoothed the size of companies and tested inhomogeneity which was given by dividing of population density with the smoothed size.

We used L-function was as a tool for testing geographical interaction of the companies (Besag, 1977) and this function was tested in inhomogeneous case. Inhomogeneity was tested by local scaling where the locally scaled factor was given by population density in the first case and by the spatially smoothed size of the companies in the second case. The models' significances were identified by Directional quantile MAD envelope test (Myllymäki, et al., 2017), which were computed using 99 simulations. The null hypothesis of the test was CSR under the given inhomogeneous function. The analysis was carried out for all sectors but in this paper there are shown only results of agriculture, forestry and fishing sector because other sectors tends to same tendency of clustering.

Generally, values of locally scaled L-function outside the envelopes represent the distance where the spatial concentration or dispersion is significant. The result of Directional quantile MAD envelope test is shown in Figure 3. At first glance, we can reveal a strong phenomenon of spatial clustering in each distance *r*.

Figure 3. Directional quantile MAD envelope test computed using 99 simulated realizations of inhomogeneous Poisson process with intensity equal to the intensity of population in Figure 1

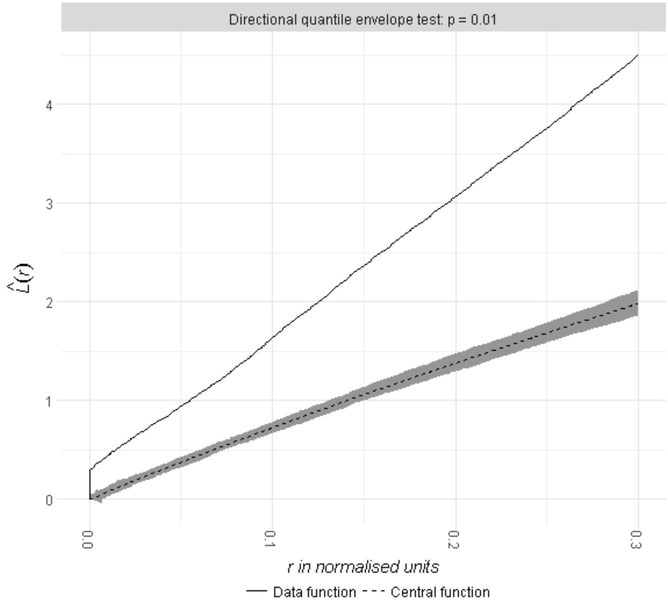


Source: Own processing

It is necessary to reject (p-value = 0.01) the hypothesis that companies are completely spatially random when the population is taken into account (Figure 3). Spatial concentration cannot be explained by the population given in the observed area.

In the second step, we tried to find out if the size of the companies can drive clustering or not thus the inhomogeneity was modelled with the size of the companies. The scaling factor was, for inhomogeneous process, calculated at every grid point as a ratio of the population and the size of the closest company. This formulation of inhomogeneity corresponds to the hypothesis that greater companies repel more employees than small companies. In Figure 4, there is depicted Directional quantile MAD envelope test where the scaled factor is calculated as a ration of the population and the companies' size. As before, the locally scaled L-function is outside the envelopes so it is necessary to reject the hypothesis of CSR (p-value = 0.01).

Figure 4 Directional quantile MAD envelope test computed from 99 simulated realizations of inhomogeneous Poisson process with intensity derived from the size of companies in Figure 2



Source: Own processing

The both analysis show positive external agglomeration effects in which spatial interaction of businesses arise. Positive agglomeration effects could be caused by many factors such as concentration in the place of natural resources, creation of special supply sector, creation of a specialized labour market, existence of special research and development facilities, special infrastructure, etc. We found out that the agglomeration effect can be explained by population only partly and there must be effect of another factors.

The last level of clustering was detected. In our case, the level of clustering was found out for distance $r = 0.25$ in each sector.

Table 2. Level of clustering in the distance $r = 0.25$ for all sectors with inhomogeneity given by population and size of companies

Sector	Level for population	Level for size	Number of companies
Agriculture, forestry and fishing	7.518	37.198	748
Mining and quarrying	1.437	3.160	21
Manufacturing industry	4.805	15.084	2192
Production and distribution of electricity, gas and water	10.215	20.263	245
Construction	131.172	247.633	1234
Wholesale and retail trade, repair of motor vehicles	376.509	696.577	2125
Transport, storage and communication	68.339	90.661	439
Accommodation and food service activities	48.175	68.454	652
Financial intermediation	12.988	13.430	73
Real estate activities, renting and business activities	534.979	708.294	1735
Education	12.759	14.143	100
Health and social care, veterinary activities	33.467	48.015	456
Other community, social and personal services	18.293	21.783	169

Source: Own processing

Considering the first case, the level of clustering was 7.518. When considering the second case, the value was 37.198 regarding the sector of agriculture, forestry and fishing. The level of clustering is lower in all sectors for the inhomogeneity given only by population (see Table 2). There is higher tendency to make clusters when the size of companies is taken into account so the companies' size does not clarify the clustering.

The highest concentration of companies is located in the town even when the population and size of companies is taken into an account. Geographic concentration is stronger than population because it helps to amplify production and innovation benefits, specifically to reduce transaction costs, increase information flow, improve specialized needs and be stronger in competitive environment. Many companies probably realise the advantages lie in clustering where the whole is more than the sum of its parts.

5. Discussion and Conclusion

The choice of a suitable location for a company and its economic activities is one of the most important decisions in the company. For this reason, the location approach had been solved in deep history when a settlement was dependent on accessible livelihoods and suitable climatic conditions and was focused on the choice of location for economic activities with optimal resources. These theories are considered as the starting point for regional science that is based on discovering specific characteristics that affect the location of activities. The first location theories were focused on agriculture and originated from the time when this sector was the most widespread. The development of industry gave rise to industrial location theories. These theories were followed by modern location theories that are based on multi-criteria approaches and modelling as in our case.

In the research we tried to find out if companies are spatially independent or if there is some spatial dependence between them. Our aim was to compare the level of clustering between sectors, not the causes of clustering. There are a lot of papers trying to explain the economic mechanism of firms' clustering by applying different methods. For example, to identify clustering of firms D-function was firstly used by Sweeney and Feser (1998) on companies in the southeast of USA. They were followed for example by Marcon and Puech (2003) with companies in Paris, France or Albert, Casanova, and Orts (2012) who analysed firms in Madrid, Spain. The bivariate K function to study location of companies in Italy was used by Arbia, Espa and Quah (2008). Sweeney and Gómez-Antonio (2016) used Gibbs models as a framework for studying industry localization. For example, Espa et al. (2010), Marcon and Puech (2003), Duranton and Overman (2005), Quah and Simpson (2003) and Arbia et al. (2008) popularized the use of K-function on location of companies.

There are existing few studies focused on inhomogeneous space of companies. The inhomogeneous K function to analyse spatial concentration of companies was solved by Arbia et al. (2012) who was concentrated on spatial concentration on five sectors of high-tech manufacturing in Milan, Italy. Further Mori, Nishikimi and Smith (2005) studied companies' localization by D-index. In the analysis they removed the effect of regional population size. Sweeney, Feser (1998) and Marcon, Puech (2003) used D-function that considers density variations to analyse if small companies are more concentrated than big ones. There is the main difference between our paper and the existing literature. Although we were concentrated on the study of location of companies as others, we used local scaling, the method nobody used on companies before.

We assumed that the location of companies could be caused by the clustering of firms in towns where a higher population exists. To remove this circumstance, we put population as a variable into our analysis. Due to the application of population we tried to explain the clustering of firms. We have shown that the clustering of firms is not completely driven by population and there has to be an influence of other factors. Therefore, the influence of the size of companies was taken into account. We found out that even after implementing population density and size of companies in the model, the companies of the studied sector still tend to make clusters in the space. We confirm the hypothesis of Porter (2000), who claims that new companies are mostly established in areas where there are other companies. The main reasons for this behaviour can be cluster advantages such as better information about opportunities. Porter and Porter (1998) found out that clustering of firms is source of economic growth and prosperity in the area because clusters increase the current (static) productivity of constituent firms or industries, the capacity of cluster participants for innovation and productivity growth, stimulate new business formation that supports innovation and expands the cluster. There were another authors who found the advantages of clusters. Krugman (1997) said that the idea of clustering of producers in given locations generates benefits. The main reason for concentration of firms for Marshall (2009) were location savings, creation of specialized workforce stock and transfer of knowledge and technical progress among firms. According to Kovárník and Stejskal (2009) the main reason for the formation of clusters is the solution of the implementation of innovation and knowledge in sectors.

Variables for inhomogeneity (population and size of companies) were determined based on another studies, for example Mori, Nishikimi and Smith (2005) who take population in their analyses into the account too. Then Porter and Porter (1998) claimed that strong clusters are often concentrated in particular geographic areas, especially in a single city or metropolitan region. In the study of Klier's (2006) was found out that headquarters of companies disproportionately locate in large metropolitan areas where the size of company plays the important role. Further Ježek (2002) found out that the role of location effects depends on the size of the enterprise.

At first sight, the mining and quarrying sector has the lowest level of clustering. Similar results are shown by the agriculture, forestry and fishing sector too. Generally, the Raw Materials Sector (Primary Sector), which covers all sectors of human activities that transform natural resources into basic products, has lower values of clustering than the Industry Sector (Secondary Sector) and the Service Sector (Tertiary Sector). The highest level of clustering is for the wholesale and retail sector and the sectors for the trade and repair of motor vehicles and real estate activities, renting and business activities. In both cases, these are companies from the Tertiary Sector. It confirms the ideas of Sweeney and Gómez ‐ Antonio (2016) who claims that the clustering will be very strong among companies from the high-tech industry and knowledge sectors.

Investors' decisions are not only dependent on geographical factors. The location of business activities is primarily dependent rather on macro and microeconomic factors. Cohen (2000) revealed major influences on location decision-making of companies about a city for the company. He revealed there are three main factor that can explain the decision about location, e.g. technology, business organisation and government policies (education, speeding-up the permitting process and simplifying bureaucracy and the (un)importance of tax incentives. From that reason we would like to add marks to the point process describing microeconomic factors especially health of companies that is indirectly united with business organization.

Acknowledgements

Authors state that the presented work was supported by the grant No. GA ČR 19-04412S of the Grant Agency of Czech Republic.

References

Albert, J. M., Casanova, M. R., and Orts, V. 2012. "Spatial location patterns of Spanish manufacturing firms." Papers in Regional Science. 91(1). 107-136. doi: 10.1111/j.1435-5957.2011.00375.x.

Alonso, W. 1972. "A theory of the urban land market." Readings in urban economics. New York. 104-112. doi: 10.1111/j.1435-5597.1960.tb01710.x.

Arbia, G., Espa, G., and Quah, D. 2008. "A class of spatial econometric methods in the empirical analysis of clusters of firms in the space." Empirical Economics, 34(1). 81-103. doi: 10.1007/s00181-007-0154-1.

Arbia, G., Espa, G., Giuliani, D., and Mazzitelli, A. 2010. "Detecting the existence of space–time clustering of firms." Regional Science and Urban Economics. 40(5). 311-323. doi: 10.1016/j.regsciurbeco.2009.10.004

Arbia, G., Espa, G., Giuliani, D., and 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.

Baddeley, A. J., Møller, J., and Waagepetersen, R. 2000. „Non ‐ and semi ‐ parametric estimation of interaction in inhomogeneous point patterns." Statistica Neerlandica. 54(3). 329-350. doi: 10.1111/1467-9574.00144.

Baddeley, A., Rubak, E., and Turner, R. 2015. "Spatial point patterns: methodology and applications with R." CRC Press. doi:10.1201/b19708.

Berry, B. J., and Harris, C. D. 1970. "Walter Christaller: an appreciation." Geographical Review. 116-119.

Besag, J. E. 1977. "Comments on Ripley's paper." Journal of the Royal Statistical Society B. 39(2). 193-195.

Christofakis, M. 2014. "Transport cost in location practice and economic geography: traditional theories, some new dimensions and policy implications." Bulletin of Geography. Socio-economic Series. 25(25). doi:10.2478/bog-2014-0029.

Cohen, N. 2000. "Business location decision-making and the cities: bringing companies back." Washington, DC: Brookings Institution Center on Urban and Metropolitan Policy.

Damborský, M., and Wokoun, R. 2010. "Lokalizační faktory malého a středního podnikání v podmínkách ekonomiky ČR." http://dspace5.zcu.cz/handle/11025/17329 (accessed May 9, 2018).

Diggle, P. J. 1983. "Statistical analysis of spatial point processes." Academic-Press. London.

Dixon, P. M. 2002. "Ripley's K function." Encyclopedia of environmetrics.

Duranton, G., and Overman, H. G. 2005. "Agglomeration and the adjustment of the spatial economy." Papers in Regional Science. 84(3). 311-349. doi: 10.1111/j.1435-5957.2005.00038.x.

Espa, G., Giuliani, D., and Arbia, G. 2010. "Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration." Department of Economics. University of Trento. doi: 10.1177/0160017612461357.

Fujita, M. 2010. "The evolution of spatial economics: from Thünen to the new economic geography". Japanese Economic Review. 61(1). 1-32. doi:10.1111/j.1468-5876.2009.00504.x.

Hahn, U., Jensen, E. B. V., Van Lieshout, M. C., and Nielsen, L. S. 2003. „Inhomogeneous spatial point processes by location-dependent scaling." Advances in Applied Probability. 319-336. doi:10.1239/aap/1051201648.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. 2008. "Statistical analysis and modelling of spatial point patterns." John Wiley & Sons. doi: 10.1002/9780470725160.

Ježek, J. 2002. "Prostorová a regionální ekonomika." Plzeň: ZČU.

Klier, T. H. 2006. "Where the headquarters are: Location patterns of large public companies, 1990-2000." Economic Development Quarterly. 20(2). 117-128. doi: 10.1177/0891242405286158.

Kovárník, J., and Stejskal, J. 2009. "Regionální politika a její nástroje." Praha: Kosmas.

Krugman, P. R. 1997. "Development, Geography, and Economic Theory." MIT Press. doi: 2389.001.0001.

Krugman, P., and Lawrence, R. 1993. "Trade, jobs, and wages." National Bureau of Economic Research. doi:10.3386/w4478.

Kuşluvan, S. 2013. "A review of theories of multinational enterprises." Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 163–180. doi: 10.1142/9789813238459_0002.

Marcon, E., and Puech, F. 2003. "Evaluating the geographic concentration of industries using distance-based methods." Journal of Economic Geography. 3(4). 409-428. doi:10.1093/jeg/lbg016.

Marshall, A. 2009. "Principles of economics." New York: Cosimo Classics.

Martinčík, D. 2008. "Ekonomicko-sociální úroveň krajů: komplexní srovnávací analýza."

Mori, T., Nishikimi, K., and Smith, T. E. 2005. "A divergence statistic for industrial localization." Review of Economics and Statistics. 87(4). 635-651. doi: 10.1162/003465305775098170.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. 2017. "Global envelope tests for spatial processes." JRSS Series B 79/2. 381-404. doi:10.1111/rssb.12172.

North, D. C. 1955. "Location theory and regional economic growth." The Journal of Political Economy. 243-258. doi:10.1086/257668.

Popovici, O. C., Călin, A. C., and others. 2014. "FDI theories. A location-based approach." Romanian Economic Journal. 17(53). 3–24.

Porter, M. E. 2000. "Location, competition, and economic development: Local clusters in a global economy." Economic development quarterly. 14(1). 15-34. doi:10.1177/089124240001400105.

Porter, M. E., and Porter, M. P. 1998. "Location, clusters, and the" new" microeconomics of competition." Business Economics. 7-13.

Quah, D., and Simpson, H. 2003. "Spatial cluster empirics." London School of Economics Working Paper Series.

Ripley, B. D. 1976. "The second-order analysis of stationary point processes." Journal of applied probability. 255-266.

Rumpel, P., Slach, O., and Koutský, J. 2008. „Měkké faktory regionálního rozvoje." University of Ostrava.

Stevens, B. H. 1968. “Location theory and programming models: The von Thünen case.” Papers in Regional Science. 21(1). 19-34. doi: 10.1111/j.1435-5597.1968.tb01438.x.
Sweeney, S. H., and 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.
Sweeney, S., and Gómez ‐ Antonio, M. 2016. “Localization and industry clustering econometrics: an assessment of Gibbs models for spatial point processes.” Journal of Regional Science. 56(2). 257-287. doi: 10.1111/jors.12238.
Thünen, J. H. 1846. “Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie.”
Weber, A. 1929. “Theory of the Location of Industries.” University of Chicago Press.