QUANTITATIVE METHODS FOR IDENTIFICATION OF REGIONAL CLUSTERS IN ROMANIA[1]

Adriana Elena REVEIU[2]
PhD, Associate Professor, Department of Economic Informatics and Cybernetics
University of Economic Studies, Bucharest, Romania

E-mail: reveiua@ase.ro

Marian DARDALA[3]
PhD, University Professor, Department of Economic Informatics and Cybernetics
University of Economic Studies, Bucharest, Romania

E-mail: dardala@ase.ro

Abstract:
The main goal of this paper is to describe some methods, that make use of employment data and that allow to measure to what the companies are spatially proximate. Specifically, we will outline the most prominent spatial concentration quotients that have been suggested in the literature to analyse the degree to which companies of the same sector are proximate (spatial concentration). We apply the methods on the employment statistics available for Romania's counties.

Key words: *regional clusters; quantitative methods; location coeficient method; shift-share analysis; Gini's location quotient; Ellison and Glaeser's agglomeration index*

1. Regional Clusters

The economic activity is concentrated in space, and therefore there is an increased attention over the forces of agglomeration and the role of location in economic development. Porter [1][4] defines clusters as a group of interconnected companies and associated institutions, close from geographical point of view, working in a particular field and linked by common and complementary elements.

Because of the proximity among them, both in terms of geography and of activities, the clusters constituent enjoy the economic benefits of several types of positive location-specific externalities.

Knox [2] defines a spatial cluster as a geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance. This is a useful operational definition, but there are very few situations when phenomena are expected to be distributed randomly in space. In most cases an implicit assumption in spatial cluster analysis is that the researcher has accounted for all the factors known, to influence the variable of study.

From a functional perspective, clusters are defined as networks of independent producers of powerful firms, including specialized suppliers, linked to each other in the value-added production channel. [3]

Spatial proximity has grown rapidly in importance, the cluster literature have made the distinction between industrial complexes and industrial clusters on spatial agglomeration of these industrial groups. Spatial proximity of the industrial activities interconnected assumed to influence the performance of these sectors and regional clusters on short and long term [4].

Clusters differ in many dimensions, such as: the type of products and services they produce, the location dynamics they are subject to, their stage of development, and the business environment that surrounds them.

Clusters can be **classified** by the type of product and services they provide. There are clusters in automotive, in financial services, in tourism, in a specific industrial area, and so on. Researches have pointed out how different locations play different roles. The development of clusters has discouraged many regions with no realistic chance of achieving a similar level of performance as the top level clusters.

From location point of view, the *local industries* are serving only local markets and are distributed across space approximately according to population size. They might be a kind of a cluster in a more narrow geographical sense, like a part of a city, due to the complementarities in attracting customers, but these effects are not strong enough to influence the development of clusters across regions.

On the other hand, the *natural resource dependent industries* serve global markets and are concentrated across space, in areas in which there are natural resources presented.

Finally, there are many industries that choose their location according to the quality of the cluster-specific business environment. There are the case of *traded industries* which serve markets in many regions and countries, and concentrate across various geographic locations. The cluster belongings to one industry is strong and its presence is a key part of the attractiveness of a specific location. Understanding the differences between these types of industries is important, because it affects the types of policies that are relevant to upgrade them.

From another point of view, clusters can be classified by the stage of development they have reached. The stage of development depends on two dimensions: on the quality of the external business environment the cluster operates in, and on the progress the cluster has made in mobilizing the potential of its business environment through active cooperation and other internal activities. [5] Researchers have looked at clusters in less developed economies as well as in less developed regions of advanced economies, such as rural regions or inner cities. Most of the theoretical literature suggests that clusters are a factor at every stage of economic development, but that in weaker environments clusters will tend to be weaker and more narrow as well. [6] Researchers have focused on the role of cultural factors, institutions, and individual leadership. There is strong view in the literature that cluster dynamics do not occur automatically, but that they depend on and can be reinforced by purposeful action. [6]

2. Quantitative Methods Used for Indentification of Regional Clusters

In the literature there are several ways of grouping the industries into clusters. To have a credible image of the cluster construction process, we can use different types of statistics and databases and various ways to collect information. Generaly, the choice of method for cluster representation depends on the type of cluster.

Location Coeficient Method is designed to group local industries into clusters, using regional data about employees. In order to identify the leading regions of a spatially concentrated industry, Kim (1995) and Hoover (1936), suggest to calculate for each locational unit in a given sample industries' employment shares, with respect to each industry's total employment in the aggregated locational unit.

$$\text{Location quotient} = \frac{\frac{n_{A,R}}{N_R}}{\frac{n_{A,T}}{N_T}}$$

where: $n_{A,R}$ is the number of employees in industry **A,** in region **R**,

N_R is the whole number of employees, in the region **R**,

$n_{A,T}$ is the number of employees, in industry **A,** at the national level,

N_T is the whole number of employees, from national level.

A region is considered to be specialized in one industry if the location quotient calculated for that region is greater than or equal to 1.5.

The method is structured as following:

1. The target geographical area is divided into regions.
2. Identification of global industries, based on the location quotient calculate for each industry. Using this quotient, the industries from each regions, could be classified in three groups: local industries, global industries and dependent by the resources industries. If there are several regions specialized in an industry, the methodology assumes that the industry is oriented globally. An industry is considered to be globally or global oriented if it exports the products outside the region or country. These are very important industries for a region because they are promoting economic growth for other industries. Local industries are the industries without export outside the region or country. Dependent by resources industries are those for which the location is defined by the resources availability.
3. Location quotients are analysed to identify patterns of clustering. Clustering algorithm is used to browse the different ways of grouping the industries to identify the best solution for grouping industries, based on the location quotient. It is used as a cluster quotient when the same group of industries is over represented in some different regions.

The choice of regions, industries and group identification are parts of an iterative process. In each step can be made refinements, until the definition of clusters match the reality. To do this, the resulted clusters are verified by various qualitative assessments.

The method has been applied in many countries because it uses only employment data, which are relatively easy available.

The main shortcoming of the method is the large dependence by the regions bounds choosing. The choice of regions must be a priori to identify clusters.

To resolve the problem of choosing the size of regions used in location quotient method and to have a more flexible method for clusters mapping, the **Ripley's K method** can be used**.**

Ripley's K method consider clusters mapping like an optimization problem of the distances between the companies. In this situation, it is not necessary to choose the regions in advance because the method identifies the optimal size of each cluster without predetermined geographical boundaries.

The methodology consists in:

1. Designing the locations of all the companies in each industry and compute the distance between the companies for each industry. Geographical concentrations of each industry can now be compared to measure the performance and the distribution of all the employees. The comparison shows whether a particular industry has a local over representation and if it can be considered as it is globally oriented. Geographical concentrations are identified by optimizing the distances between the companies, which is the size of specialized areas. This issue solve the problem of predefined chooses of regions of the location quotient method.
2. The patterns relating to the location of global industries are evaluated using statistical features. A cluster algorithm tries to identify the locations for each industry, in order to identify systematic patterns of clustering among industries. Like in the location quotient method, the mapping is an iterative process to identify the best clustering corresponding to reality.

The main shortcoming of this method is a greater dependence on the details about the location of each company, data hardly available. More than this, in the case of Ripley's K method, the volume of calculations to be made is extremely high.

Shift-share analysis decomposes in factors the changes in value of an indicator, such as number of employees, income, added value and so on. Decomposition is done in three parts and expresses the effect of absolute change of the indicator and the effect of changes in its structure. The method uses the assumption that regional economic growth can be explained by a combined effect of three components: increasing at national level, growth in the structure of the branch and growth due to other factors, the local factors. The last so-called competitive component is rated as most important; it emphasizes the region's top branches. Mathematically, the decomposition can be expressed by the equation:

$$ZP = ZN + ZO + ZR$$

where: ZP = changes in share of the selected index,

ZN = changes of the selected index, at the national level,

ZO = changes in share of the branch structure of the selected index,

ZR = regional changes in share of the selected index.

Changes of the index value is compared for two time periods, not necessarily two consecutive years, but rather is recommended a longer period (3-5 years). To perform the computations, the available values of the index should cover a larger area and the region for the two selected years should be divided according to NACE specifications. The individual components of the equation are determined by the relations established according with the following equations:

$$ZN = \frac{CR^t}{CR^{t-n}}$$

$$ZO = \frac{CR_i^t}{CR_i^{t-n}} - ZN$$

$$ZR = \frac{R_i^t}{R_i^{t-n}} - \frac{CR_i^t}{CR_i^{t-1}}$$

where: CR^t – the average number of employees of the national economy, in year *t,*

CR^{t-n} – the average number of employees of the national economy, in year *t-n*,

CR_i^t – the average number of employees of branch *i*, in year *t*,

CR_i^{t-n} – the average number of employees of branch *i*, in year *t-n*,

R_i^t – the average number of employees, from the region, in branch *i*, in year *t*,

R_i^{t-n} – the average number of employees, from the region, in branch *i*, in year *t-n*,

n- the length of analysed perriod.

Ellison and Glaeser's agglomeration index

The index defines the share of total geographical concentration for the branch *i*.

$$G_{EG}^k = \frac{\sum_i \left(l_i^k - x_i\right)^2}{1 - \sum_i x_i^2},$$

where: $x_i = \frac{\sum_k z_i^k}{\sum_i \sum_k z_i^k}$ – is the share of region *k* in the employment of the whole industry.

$l_i^k = \frac{z_i^k}{Z_i}$, where:

l_i^k – the share of employment of region *k*, from branch *i,*

z_i^k – the number of employees in branch *i*, in region *k*,

Z_i – the number of employees in branch *i*, at the national level.

The index is based on the comparation of the shares of employees in the selected branch, in the region and in the whole manufacturing branch.

If the index values are less than 0, then the branch is dispersed across the whole territory and cannot be described as geographically concentrated.

For the index value in a range between 0 to 0.02, it is an insignificant, very weak geographical concentration of the branch. For an index between 0.02 and 0.05 it is a medium-strong geographical concentration, and above 0.05 of strong geographical concentration [7]. In order to determine industry concentration, the modified Herfindahl index must be determined for branch *i* through application of the equation: $H^i = \sum_j \left(z_j^i\right)^2$ where: z_j^i – share of employees in enterprise *j* in the whole share of employees in the branch *i*.

The aglomeration index calculed using the formula: $\gamma_{EG}^{i} = \frac{G_{EG}^{i} - H^{i}}{1 - H^{i}}$ express the degree of additional geographic concentration of the relevant industrial branch.

Ellison and Glaeser argue that there aren't savings in agglomerations where the territorial units there are equally attractive, to a certain branch. In such a situation, the gross geographic concentration is the same with the industry concentration express by the Herfindahl. The γ_{EG}^{i} index reflects the additional concentration of the branch, developed by the region's competitive margin.

Gini's location quotient

For an assessment of the overall spatial concentration of an industry compared to other industries, Krugman (1991) suggested to compute Gini's location quotient.

This method involves the following steps to determine the location quotient:

1. It determines the share of employees in a particular branch, in total employment at the national level, using the following equations: $l_i^n = \frac{z_i^n}{Z_i}$, where:

l_i^n – the share of employment in the branch *i*, of the region *n*,

z_i^n – the number of employees in industry *i*, from the region *n*,

Z_i – the number of employees in industry *i*, at the national level.

2. The regions must be descending order to ensure that: $l_i^1 \geq l_i^2 \geq \cdots \geq l_i^N$

The whole number of regions is equal with N.

3. It is necessary the cumulative share of the employees in the branch *i* and the cumulative share of the employment in the whole branch. The cumulative shares could be represented by so-called Lorenz curves. Gini's location quotient is represented by the sourface between the straight line and an angular quotient of 45° and Lorenz curve, and could be determined using the equation:

$$GC = \frac{1}{2}\sum_{n=1}^{N}(u_{n-1} + u_n)\beta_n - \frac{1}{2} \text{ unde } GC \in [0; 0.5]$$

The more geographically concentrated the branch of industry is, the higher the value of GC is. The maximum GC value is 0.5. On the contrary, the branch exhibiting the same spatial distribution as that of the entire industry will have a GC equal to 0.

where: $u_n = \sum_{n=1}^{N} l_i^n$ – cumulative amount of share of employees of branch *i*, in region *n*,

β_n – share of region *n* in the employment in the entire manufacturing industry.

3. Application of the methods employment data from Romania

We are going to demonstrate the usage of the presented methods on a statistical data set related to the employment, in all the Romania's counties, in 2009. The data source is the Romania's Statistical Yearbook, from 2010. In figure 1, we exemplify the results of applying the location quotient model in the mining and quarrying branch. There are 8 counties: Bacau, Mures, Hunedoara, Gorj, Valcea, Dambovita, Prahova and Teleorman, which are considered to be specialised in this branch.

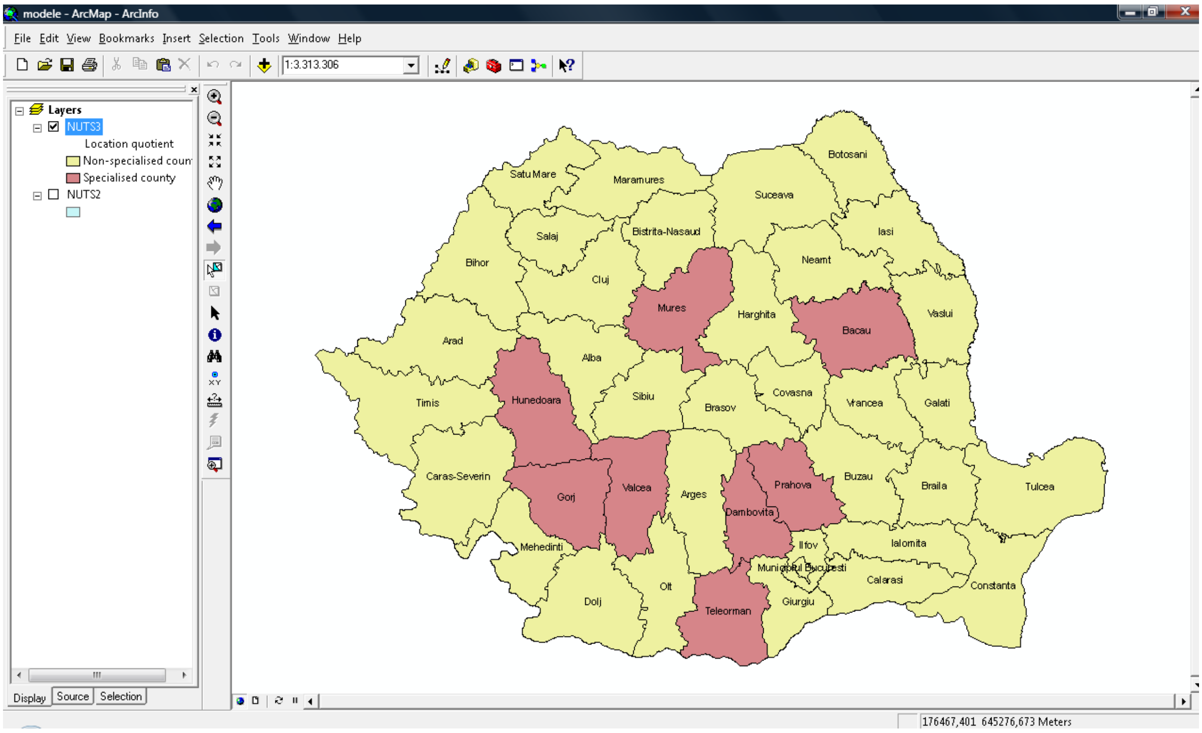

Figure 1. Location quotient method results- the mining and quarrying branch

The **shift-share analysis** used employment data from the Bucharest City and from Romania, in 2004 and 2009.

Table 1. The results of shift-share analysis applied on the Bucharest City employment data

	SS	ZN	ZR	ZP	ZO
Agriculture, forestry and fishing	0.24	1.02	-0.67	0.24	-0.11
Industry	0.77	1.02	-0.09	0.77	-0.16
Mining and quarrying	1.35	1.02	0.73	1.35	-0.40
Manufacturing	0.68	1.02	-0.15	0.68	-0.19
Electricity, gas, steam and air conditioning production and supply	0.74	1.02	0.19	0.74	-0.47
Construction	1.75	1.02	0.27	1.75	0.46
Wholesale and retail; repair of motor vehicles and motorcycles	1.35	1.02	0.15	1.35	0.18
Transport and storage	3.15	1.02	0.11	3.15	2.01
Hotels and restaurants	0.29	1.02	-0.01	0.29	-0.72
Financial intermediation and insurance	1.74	1.02	0.41	1.74	0.31
Real estate activities	0.10	1.02	-0.03	0.10	-0.89
Public administration and defence; social insurance of public sector	1.48	1.02	0.07	1.48	0.39
Education	0.93	1.02	-0.03	0.93	-0.06
Health and social assistance	1.20	1.02	0.10	1.20	0.08
Other service activities	0.59	1.02	-0.02	0.59	-0.41

The table 1 shows that the number of employees in Romania increased by 1.02% on average, during the period under consideration (ZN component). The ZO component compares the change in employment in the branch, to the average change in employment in the entire country. The most interesting is the ZR component, which compares the relative change in the number of staff in the branch and region to the relative change at the national level. The ZR component represents that section of the branch development in the region, which is explained away by regional factors, namely, the nature of local conditions for development of economic activity. From this point of view, good condition have been generated in Bucharest Region, during 2004 and 2009, for the development of the following branchs: *Mining and quarrying*, *Financial intermediation and insurance*, *Construction, Electricity, gas, steam and air conditioning production and supply, Wholesale and retail; repair of motor vehicles and motorcycles, Transport and storage,* and *Health and social assistance*.

Ellison and Glaeser's agglomeration index

The table 2 contains the values of Ellison and Glaeser's agglomeration index and the type of resulted geographial concentration, calculated based on employment statistical data available for the main branches from Romania.

Table 2. The Ellison and Glaeser's agglomeration index results

NACE	Ellison and Glaeser index value	Type of geographical concentration
Agriculture, forestry and fishing	0.01997	very weak
Industry	0.08482	Strong
Mining and quarrying	0.02170	medium-strong
Manufacturing	0.04904	medium-strong
Electricity, gas, steam and air conditioning production and supply	0.03698	medium-strong
Water supply; sewerage, waste management and decontamination activities	0.06339	Strong
Construction	0.03936	medium-strong
Wholesale and retail; repair of motor vehicles and motorcycles	0.04110	medium-strong
Transport and storage	0.05531	Strong
Hotels and restaurants	0.31619	Strong
Information and communication	0.19040	Strong
Financial intermediation and insurance	0.07051	Strong
Real estate activities	0.16910	Strong
Professional, scientific and technical activities	0.13824	Strong
Activities of administrative services and of support services	0.04569	medium-strong
Public administration and defence; social insurance of public sector	0.03474	medium-strong
Education	0.03675	medium-strong
Health and social assistance	0.10843	Strong
Shows, culture and recreation activities	0.07344	Strong
Other service activities	0.00000	very weak

Gini's location quotient

In the table 3 have been presented the Gini's location quotients. According to Gini's, the most concentrated industries are: *Information and communication*, *Financial intermediation and insurance, Professional, scientific and technical activities, Activities of* *administrative services and of support services* and *Shows, culture and recreation activities*. But the concentration level is very soft, the values being very small.

One of the main problem of the usage of Gini's location quotient is the fact that it does not control for industrial concentration. Gini's quotient consider an industry localized, if it is strongly concentrated in a limited number of geographical units.

Table 3. The Gini's location quotient results

NACE	Gini quotient
Information and communication	0.220
Financial intermediation and insurance	0.187
Professional, scientific and technical activities	0.181
Activities of administrative services and of support services	0.173
Shows, culture and recreation activities	0.165
Other service activities	0.151
Construction	0.150
Real estate activities	0.149
Hotels and restaurants	0.141
Wholesale and retail; repair of motor vehicles and motorcycles	0.138
Electricity, gas, steam and air conditioning production and supply	0.134
Transport and storage	0.134
Public administration and defence; social insurance of public sector	0.130
Health and social assistance	0.129
Education	0.127
Water supply; sewerage, waste management and decontamination activities	0.126
Industry	0.118
Manufacturing	0.118
Mining and quarrying	0.105
Agriculture, forestry and fishing	0.092

4. Conclusions

The range of quantitative methods used to identify potential clusters is very large. In this paper we tackle only a few. It is not possible to declare that a certain method is generally better or worse in comparison with other methods. The selection of a specific method depends on the type of cluster and the links between its members we seek to identify.

In practice, we find out the most widely used are the location quotients. They involve a rather undemanding method suitable for searching of local and regional clusters. The strong points of the method include the fact that the recalculations may generally use the available statistical resources. But, the location quotients cannot capture the mutual links between companies.

The shift-share analysis specify the branches that are successful in the region, in terms of the trends in employment. One disadvantage of the method is the fact that favourable results may be reached by branches where the share in overall employment in the region is largely negligible and where the region does not reveal any specialisation. On the contrary, for important branches, the results of the shift-share analysis may be ambiguous, where the given branch in the region is passing through a stage of growing at a slower speed than in the other parts of the country.

While identifying national clusters the usage of the locational Gini coefficient or Ellison and Glaeser's agglomeration index is recommended. These methods may be used to determine whether a certain branch is geographically concentrated on a national scale.

But, a simple concentration of a certain industry in the region does not necessarily mean that a cluster is present.

It is important to establish links between branches, also. That may be done by applying a broad range of methods with the objective to measure the importance of purchasing and sales flows. Based on the established links, the initial cluster map may be outlined.

References

1. Bertinelli, L. and Decrop, J. **Geographical Agglomeration: the Case of Belgian Manufacturing Industry**, in: Working Paper 14-02. Federal Planning Bureau, Brussels, 2002
2. Ellison, G. and Glaeser, E. **Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach**, Journal of Political Economy, 105, 1997
3. Enright, M. **Regional clusters and economic development: A research agenda, Business Networks: Prospects for Regional Development**, Walter de Gruyter, Berlin, 1996
4. Kim, S. **Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987**, Quaterly Journal of Economics, 110, 1995
5. Knox, E.G. **Detection of clusters**, in Elliott, P. (ed.) "Methodology of enquiries into disease clustering", Small Area Health Statistics Unit, London, pp.17-20
6. Krugman, P. **Geography and trade,** MIT Press, Cambridge, 1991
7. Maskell, P. **Towards a Knowledge-based Theory of the Geographical Cluster**, Industrial and Corporate Change, 2001, pp. 919-941
8. Porter, M. **Clusters and the New Economics of Competition**, Harvard Business Review, 1998
9. Porter, M. **The Competitive Advantage of the Inner City, In: On Competition**, Boston, Harvard Business School Press, 1998
10. Roelandt, T. J.A. and den Hertog, P. **Boosting innovation; the cluster approach**, Paris, OECD, 1999

---

[1] **Acknowledgments**
This work was supported by the project "Performance and excellence in postdoctoral research in Romanian economics science domain", contract POSDRU/89/1.5/S/59184.

[2]Adriana REVEIU has graduated the Faculty of Cybernetics, Statistics and Economic Informatics. She holds a PhD diploma in Economic Cybernetics and Statistics. She is associate professor in Economic Informatics field and branches within Department of Economic Informatics at faculty of Cybernetics, Statistics and Economic Informatics from Academy of Economic Studies. She is the author and co-author of 10 books and over 50 articles in journal and proceedings of national and international conferences, symposiums, workshops in the fields of multimedia, communications, learning systems and data management.

[3] Marian DÂRDALĂ has graduated the Faculty of Cybernetics, Statistics and Economic Informatics. He holds a PhD diploma in Economic Cybernetics and Statistics. He is professor in Economic Informatics field and branches within Department of Economic Informatics at faculty of Cybernetics, Statistics and Economic Informatics from Academy of Economic Studies. He is the author and co-author of 15 books and over 50 articles in journal and proceedings of national and international conferences, symposiums, workshops in the fields of multimedia, human computer interaction - HCI, GIS, data bases and object oriented programming.

[4] **Codification of references in text:**

[1]	Porter, M. **Clusters and the New Economics of Competition**, Harvard Business Review, 1998
[2]	Knox, E.G. **Detection of clusters**, in Elliott, P. (ed.) "Methodology of enquiries into disease clustering", Small Area Health Statistics Unit, London, pp.17-20
[3]	Roelandt, T. J.A. and den Hertog, P. **Boosting innovation; the cluster approach**, Paris, OECD, 1999

[4]	Maskell, P. **Towards a Knowledge-based Theory of the Geographical Cluster**, Industrial and Corporate Change, 2001, pp. 919-941
[5]	Enright, M. **Regional clusters and economic development: A research agenda, Business Networks: Prospects for Regional Development**, Walter de Gruyter, Berlin, 1996
[6]	Porter, M. **The Competitive Advantage of the Inner City, In: On Competition**, Boston, Harvard Business School Press, 1998
[7]	Bertinelli, L. and Decrop, J. **Geographical Agglomeration: the Case of Belgian Manufacturing Industry**, in: Working Paper 14-02. Federal Planning Bureau, Brussels, 2002
[8]	Krugman, P. **Geography and trade,** MIT Press, Cambridge, 1991
[9]	Ellison, G. and Glaeser, E. **Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach**, Journal of Political Economy, 105, 1997
[10]	Kim, S. **Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987**, Quaterly Journal of Economics, 110, 1995