ELIT
Economic Laboratory Transition
Research Podgorica

Montenegrin Journal of Economics

Afanasyev, M.Y., Kudrov, A.V. (2021), “Economic Complexity, Embedding Degree and Adjacent Diversity of the Regional Economies”, *Montenegrin Journal of Economics*, Vol. 17, No. 2, pp. 7-22.

Economic Complexity, Embedding Degree and Adjacent Diversity of the Regional Economies

MIKHAIL Y. AFANASYEV[1] and ALEXANDER V. KUDROV[2]

[1] Professor, Central Economics and Mathematics Institute of the Russian Academy of Sciences (CEMI RAS), Moscow, Russia, e-mail: miafan@cemi.rssi.ru
[1] Professor, Central Economics and Mathematics Institute of the Russian Academy of Sciences (CEMI RAS), Moscow, Russia, e-mail: kovlal@inbox.ru

ARTICLE INFO

Received August 14, 2020
Revised from September, 27, 2020
Accepted November 28, 2020
Available online June 15, 2021

JEL classification: C12; C51; R15

DOI: 10.14254/1800-5845/2021.17-2.1

Keywords:

Regional economy,
economic complexity,
embedding index,
economic structure.

ABSTRACT

This paper presents a probabilistic model for identifying potential sectors for inclusive economic development related to existing comparative advantages of the regional economy. The model uses bipartite network representation for the economy’s structures presented in the paper (Hausmann, Hidalgo, Bustos, Coscia, Simoes, Yildirim, 2011). For the considered network representation, it is given the probabilistic interpretation and described some theoretical properties. Also, it is presented the procedure for filtering the network structure from randomness, which improves the model’s quality. As an example of the model implementation, for each region of Russia, it is estimated the probability of a specific sector's appearance as a strong one in its economy’s structure. Using the estimated probabilities, a list of sectors recommended for priority development in a region has been compiled. Based on the list of recommended sectors, groups of adjacent sectors have been formed. Also, it is studied the empirical relationships between characteristics of the considered network representation for structures of the Russian region's economies and some factors of socio-economic development factors including economic complexity.

INTRODUCTION

One of the crucial tasks in the regional economy is to study the mechanisms of knowledge diffusion. Nobel laureate Krugman (2011) notes that careful analysis of developed economies indicates that the time has come to shift the focus from material factors, such as transportation costs, to non-material factors, such as knowledge diffusion. Today there are quite a few works that attempt to describe the mechanism of external factors' influence on the processes of new knowledge penetration. A complete overview is presented in the paper (Beaudry, Schiffauerova, 2009). The transfer of knowledge within or between sectors is a driving force of economic development. There are two main, but opposite, theories describing the mechanism of knowledge creation and diffusion: localized specialization and economic diversification. As noted in (Kluge and Lehmann, 2012), the theory of localized specialization was first introduced in (Marshall, 1890) and stated that companies surrounded by other members of the same industry would grow faster due to the circulation of knowledge within the industry. This theory was further developed in work (Arrow, 1962; Romer, 1986). In the article (Glaeserl et al., 1992), this effect is called MAR-externalia by the authors' names who developed this theory. There is a series of articles presenting both positive and negative effects of specialization on the examples of different countries and industries (Henderson et al., 1995; Combes, 2000; de Lucio et al., 2002; Dekle, 2002; Blien and Suedekum, 2005). Also noted in (Kluge, Lehmann, 2012), that the opposite theory is presented in work (Jacobs, 1970). It rejects the concept that knowledge flows only diffuse the contours of the same industry. According to this theory, companies will benefit from facing a heterogeneous environment of different industries, as new ideas will not be formed endogenously, but come from an external environment. The mechanisms through which diversity leads to economic growth are usually called economic urbanization or diversification. An assessment of empirically observed effects within this theory is presented in (Glaeser et al., 1992; Blien and Wolf, 2006; Lee et al., 2005; Fuchs, 2011; Illy, Schwartz et al., 2011).

All studies on localization or diversification theories suggest that these theories work independently of each other and that one or the other should be chosen. Although half of the papers referenced in this review (Beaudry and Schiffauerova, 2009) show positive effects from both externalities, neither of these studies assesses the effects of interaction between them. The effects of localization and urbanization may complement each other and are not mutually exclusive. Ideally, the regional economy should be both specialized and diversified. A high level of specialization makes a region vulnerable to adverse changes in market conditions; too much diversity in economic activity prevents it from concentrating resources for successful competition in world markets. It is, therefore, challenging to determine the best production structure to create economic growth, specialization or diversity. The solution to this problem can be the so-called adjacent diversity, see (Frenken et al. 2007). In the paper (Frenken et al. 2007), it was shown that localized specialization contributes to employment growth while diversification constrains unemployment growth. For localized industries with overlapping interests, there is excellent potential for labor and knowledge sharing.

Moreover, innovations and other solutions used by companies in related industries can be more easily adapted. On the other hand, the presence of a large number of related economic activities in the region contributes to the generation of new combinations of existing technologies and the redistribution of risk from unfavorable market conditions through the diversity of sectors in the regional structure economy. A region that becomes dependent on one specific type of economic activity is exposed to severe risks.The paper (Neffke, Henning, Boschma, 2011) studied the dynamics of change in the regional economy structure. . It showed that the rapid economic development of regions is likely to be achieved through the emergence of industries that are technologically connected with the previously existing industries in these regions. Namely, industries that are technologically related to the previously existing industries in the region are more likely to diffuse into the region than industries that are not technologically related. The paper (Neffke, Henning, Boschma, 2011) has shown that the presence of industrial contiguity conditions the growth in the share of specific industries and that it is easier for regions to develop new industries in the structure of their economies if they are related to already existing industries in the region. Conversely, companies operating in industries that are not related to other activities are more likely to leave the region. Self-reinforcing processes arise from the concentration of related industries, creating endogenous economic growth. There are several articles (see, for example, (Hausmann, Hidalgo, 2011), (Hausmann and Klinger, 2006), Hidalgo, et al. (2007) studies in which it is presented the estimator of relatedness between products or sectors. That estimator equals the minimum of the pairwise conditional probabilities, which are obtained empirically using a different number of observations and it leads to uneven statistical reliability. For estimates obtained using a small number of observations, we get a poor statistical reliability of estimates.

This paper deals with the problem of constructing a model that would allow forecasting the emergence of new strong sectors based on the characteristics of the economy structure with better statistical reliability since it is used more information about the economies with close structures. The possibility of constructing such a model is based on the assumption that the life cycle of a sector of the regional economy combines dynamics that turn a developing sector into a "mature" one, after which a transition to the creation of new related sectors is possible. Thus, the evolution of past economic activities contributes to the emergence of new sectors on its basis. This interaction contributes to the formation of technological chains, the development process of which creates an integral economy with a high level of economic complexity, with a more stable growth structure and sufficiently diversified risk factors. Using data on the regional economies structure, which is characterized by a great variety and different degrees of value chain development, a probabilistic model is constructed, which allows obtaining probabilistic estimates for the emergence of a given sector as a strong one the structure of the region's economy. The obtained probabilistic estimations can be used to solve project management problems (Makarov, 2010), in the context of forming a list of potential strong sectors of a region associated with already formed 'mature' sectors in the structure of its economy.

In the construction of the model, it was used the aggregated indicators characterizing the embedding structures of strong sectors of regional economies. These indicators have been introduced as a result of probabilistic interpretation and description of matrix elements properties, by which economic complexity is estimated, see (Hausmann, Hidalgo, Bustos, Coscia, Simoes, Yildirim, 2011). The embedded structures of economies correspond to different degrees of common value chain development. Aggregate embedding indicators are statistically significant explanatory variables for economic complexity. Also, it is introduced the interpretation of aggregated embedding measures in the context of the economic basis, see (Aivazian, Afanasiev, Kudrov, 2018).

1. THE METHODOLOGY OF RESEARCH

Structure of strong sectors. Regional data on output in a wide range of sectors are used to describe the structure of strong sectors. To begin with, let us define the indicator RCA_{cp} of the identified comparative advantages:

$$RCA_{cp} = \frac{(y_{cp}/\sum_p y_{cp})}{(\sum_c y_{cp}/\sum_{c,p} y_{cp})}, \quad (1)$$

where y_{cp}— output of sector *p* of the regional economy *c*. Accordingly, RCA_{cp} represents the ratio of the share of production from sector *p* in the total volume of production from all sectors of the region's *c* economy to the share of production from sector *p* in all regions in the volume of production from all sectors of the economy of all regions. According to the articles in which RCA_{cp} is used to estimate of the revealed comparative advantages in the economies, see (Hausmann, Klinger, 2006), which is taken with the restriction from below. Namely, if the value of RCA_{cp} exceeds one[1], it is considered that the economy of region *c* has the identified comparative advantages in the output of the sector *p.* Otherwise, it is considered that there are no identified comparative advantages. More formally:

$$a_{c,p} = \begin{cases} 1, if\ RCA_{cp} > 1 \\ 0, if\ RCA_{cp} \le 1 \end{cases} \quad (2)$$

Matrix $(a_{c,p})$ contains data on the economy sectors that are developed in different regions at the level of the identified comparative advantages defined by the expression (1). The rows of this matrix correspond to the regions; the columns correspond to the economy sectors. Further, we will call the vector $(a_{c,p_1}, \dots, a_{c,p_m})$ as the *structure of strong sectors* of the regional economy *c or structure of the economy*. Correspondingly, we call a sector as strong in the regional economy, if its share of production in the region is higher than in the national economy.

Economic complexity and two-layer graph for strong sectors structure. Let us define a two-layer graph $G(X, Y, E)$, where X is a set of nodes of the first layer that correspond to regions, $Y -$ set of nodes of the second layer that correspond to sectors of the economy, and $E -$ graph edges between

[1] Other threshold higher than one can be used.

nodes (strong sectors in the region). Let n be the total number of regions under consideration and m – total number of sectors.

$X = \{c_1, \dots, c_n\}$

$Y = \{p_1, \dots, p_m\}$

For graph $G(X, Y, E)$ the nodes of the same layer have no connecting edges. Edges connect only a pair of nodes from different layers. The presence or absence of an edge between nodes c and p, where $c \in X, p \in Y$, is determined by the elements of matrix $(a_{c,p})$. If $a_{c,p} = 1$, then there is an edge between nodes *c* and *p*. Otherwise, there is no edge.

The concept of economic complexity of a region is further regarded as a characteristic reflecting the level of its technological development, which in turn is determined by strong sectors in the structure of its economy. Similarly, the economic complexity of a sector depends on the technological development of those regions where the sector is present in the structure as a strong one. Let us give a more formal definition of economic complexity, which corresponds to the procedure of its calculation presented in the paper (Hausmann et al., 2011).

Definition (*economic complexity*). The economic complexity of a region (ECI_c) or sector (ECI_p) is a latent characteristic, which has the following properties. The economic complexity of the region is proportional to the average economic complexities of strong sectors in the structure of its economy. Namely:

$ECI_c = a_1 \sum_p r_{c,p} ECI_p$, where $r_{c,p} = \frac{a_{c,p}}{k_{c,0}}$, $k_{c,0} = \sum_p a_{c,p}$,

where a_1- positive constant;

The economic complexity of the sector is proportional to the average economic complexities of regions in the structure of whose economies this sector is strong:

$ECI_p = a_2 \sum_c r^*_{p,c} ECI_p$, where $r^*_{p,c} = \frac{a_{c,p}}{k_{p,0}}$, $k_{p,0} = \sum_c a_{c,p}$,

where a_2- positive constant.

Further, the indicator $k_{c,0}$, which is defined as equal to the number of strong sectors in the economy structure of the region *c*, will be called as *diversification* of the region's economy *c*.

Let's enter some additional designations:

$\mathbf{c} = \left(ECI_{c_1}, ECI_{c_2}, \dots\right)^{\mathbf{T}}$ - vector-column of economic complexities for the regions $c_1, c_2, \dots$;

$\mathbf{p} = \left(ECI_{p_1}, ECI_{p_2}, \dots\right)^{\mathbf{T}}$ - vector-column of economic complexities for the sectors $p_1, p_2, \dots$;

$\mathbf{R_1} = (r_{c,p}), \mathbf{R_2} = (r^*_{p,c})$ - weight matrices.

Write down the properties (a) and (b) in a matrix form: $\mathbf{c} = a_1 \mathbf{R_1 p}, \mathbf{p} = a_2 \mathbf{R_2 c}$. It follows from that:

$\mathbf{c} = a_1 a_2 \mathbf{R_1 R_2 c}$,

$\mathbf{p} = a_1 a_2 \mathbf{R_2 R_1 p}$.

Thus, the economic complexity of a region is defined as the eigenvector of matrix $\mathbf{R_1 R_2}$, and economic complexity of the sector – the eigenvector of matrix $\mathbf{R_2 R_1}$. Note that the element at the intersection of *i*-th row and *j*-th column of the matrix $\mathbf{R_1 R_2}$, that is $(\mathbf{R_1 R_2})_{ij}$, is given by the following formula:

$$\frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}}.$$

In this paper (Hausmann and Rodrik, 2003) it is suggested to use the standardized second principal component of matrices $\mathbf{R_1R_2}$ and $\mathbf{R_2R_1}$ as the values of economic complexity estimations. It should be noted that the coordinates of the first principal component for these matrices consist of the same values, as the matrices $\mathbf{R_1R_2}$ and $\mathbf{R_2R_1}$ are stochastic, see (Hausmann, Rodrik, 2003), (Kemp-Benedict E., 2014). Let us also recall that if $\mathbf{x}$ is the eigenvector for matrix $\mathbf{A}$, which corresponds to the eigenvector λ, then the vector $r\mathbf{x}$, where r is any non-zero real number, is also the eigenvector of matrix $\mathbf{A}$, which corresponds to the same eigenvector λ (for more details, see (Afanasiev, Kudrov, 2020)).

Interpretation matrix $\boldsymbol{R_1R_2}$ elements and their properties. Let U_c — random variable corresponding to region c that takes values from the set $\{i: a_{c,p_i} = 1\}$ with equal probabilities, that is:

$P(U_c = r) = \frac{1}{k_{c,0}}$, where $r \in \{i: a_{c,p_i} = 1\}$, and $k_{c,0} = \sum_{j=1}^{m} a_{c,p_j}$. (1)

Let's also assume that V_p — random variable corresponding to sector p, which takes values from the set $\{i: m_{c_i,p} = 1\}$ with equal probabilities, that is:

$P(V_p = r) = \frac{1}{k_{p,0}}$, where $r \in \{i: a_{c_i,p} = 1\}$, and $k_{p,0} = \sum_{j=1}^{n} a_{c_j,p}$. (2)

Random variables U_c and V_p are assumed independent. For this model, we calculate the probability

$$w_{i,j} = P(on\ the\ graph\ G(X,Y,E)\ there\ is\ a\ path\ from\ node\ c_i\ to\ c_j) =$$
$$= \sum_{t=1}^{m} P(U_{c_i} = t, V_{p_t} = j) = \sum_{t=1}^{m} \frac{a_{c_i,p_t}}{k_{c_i,0}} \frac{a_{c_j,p_t}}{k_{p_t,0}} = \frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}}. \quad (3)$$

Therefore, matrix $(w_{i,j})$ matches up with the matrix $\mathbf{R_1R_2}$, which, in turn, evaluates the economic complexity of the regions. Note some properties for $(w_{i,j})$. For each $i \in \{1, \dots, m\}$ the sum of probabilities $w_{i,1}, \dots, w_{i,m}$ equals one:

$$\sum_{t=1}^{m} w_{i,t} = 1.$$

If there is at least one strong sector in region c_i then $w_{i,i} > 0$. Otherwise, $w_{i,i} = 0$.

The fairness of this statement is easy to show, since:

$$w_{i,i} = \frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t}}{k_{p_t,0}} \geq 0,$$

and zero value is reached only when $a_{c_i,p_t} = 0, t = 1, \dots, m,$ but there are no such cases in our data which will be considered further. Elements $w_{i,j}$ are equal to zero if and only if the following condition is fulfilled: $\{t: a_{c_i,p_t} = 1\} \cap \{t: a_{c_j,p_t} = 1\} = \emptyset$.

The fulfillment of this condition means the absence of common strong sectors in the economic structures of the regions c_i and c_j. In each row of the matrix $(w_{i,j})$ the maximal element corresponds to the diagonal element, i.e. $w_{i,i} = \max_{j \in \{1,\dots,n\}}(w_{i,j})$. Let's show it. Due to:

$$w_{i,j} = \frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}} \leq \frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t}}{k_{p_t,0}} = \frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t} a_{c_i,p_t}}{k_{p_t,0}}$$

we get that $w_{i,j} \leq w_{i,i}$. Moreover, equality in the latter inequality is achieved only when the following condition is satisfied:

$\{t: a_{c_i,p_t} = 1\} \subseteq \{t: a_{c_j,p_t} = 1\}$.

Fulfillment of this condition means that all strong sectors of the economic structure for the region c_i are also strong sectors in the structure of the region c_j economy. If this condition is not satisfied, we have severe inequality:

$$\frac{w_{i,j}}{w_{i,i}} < 1$$

Matrix asymmetry $(w_{i,j})$. It's easy to show that $w_{j,i} = \frac{k_{c_i,0}}{k_{c_j,0}} w_{i,j}$. If the level of diversification in region c_i coincides with the level of diversification in region c_j, then $w_{j,i} = w_{i,j}$; if the level of diversification in region c_i is more (less) than the diversification in region c_j, then $w_{j,i} > w_{i,j}$ $(w_{j,i} < w_{i,j})$. Thus, different levels of diversification of regions guarantee asymmetry of the matrix $(w_{i,j})$.

From properties (2-3) it follows that the ratio $\frac{w_{i,j}}{w_{i,i}}$ can be interpreted as characteristic of the embedding degree of the strong sectors set in region c_i into the set of strong sectors for region c_j. The lower this ratio, the fewer strong sectors in the structure of region c_i are included in the set of $\{t: a_{c_j,p_t} = 1\}$ strong sectors in the structure of the economy in region c_j.

Aggregate index

$$I_i^{(1)} = \sum_{j=1}^{n} \left(\frac{w_{i,j}}{w_{i,i}}\right)^2, i = 1, \dots, n, \qquad (4)$$

characterizes the extent to which the structure of the strong sectors of the economy structure in region c_i is embedded in the structures of the strong sectors of other regions. Let us call this indicator as the *aggregate embedding of the structure* of the region's economy. The minimum value I_i^1, which is equal to one, appears either in conditions when the structure of strong sectors in region c_i consists of unique sectors, for which $w_{i,j} = 0$ for all $i \neq j$. Aggregate index

$$I_j^{(2)} = \sum_{i=1}^{n} \left(\frac{w_{i,j}}{w_{i,i}}\right)^2, j = 1, \dots, n \qquad (5)$$

characterizes the degree of embedding in the structure of strong sectors in the structure of region c_j. We call this indicator as the *aggregate embedding in the structure* of the region's economy. When the number of strong sectors in the structure of region c_j increases, the value of $I_j^{(2)}$ does not decrease. The minimum value $I_j^{(2)}$ equals to one, appears when the structure of strong sectors for region c_j consists of unique sectors, that is when $w_{i,j} = 0$ for all $i \neq j$. The correlation of $I_j^{(2)}$ with the index of diversification of strong sectors in region c_j is 0.99, which indicates that the process of diversification in Russian regions is mainly formed not by the unique strong sectors, but by the development of embedded structures. The correlation coefficient $I_i^{(1)}$ with the index of diversification $k_{c_i,0}$ is 0.21, and the correlation between the index of embedding of the structure and the index of embedding in the structure is 0.18, which suggests that they are not correlated.

Null hypothesis: no adjacent sectors in the economy. Among the set of calculated characteristics of the embedding degree of the strong sectors $\left(\frac{w_{i,j}}{w_{i,i}}\right)$ one should distinguish the values which are statistically different from zero. For this purpose, let us consider the problem of approximating the embedding characteristic $\frac{w_{i,j}}{w_{i,i}}$ distribution function in the assumption of absence of adjacent sectors in the structure of regional economies (let us denote this hypothesis as H_0). This problem is solved by the following procedure:

– *Stage 1*: possible scenarios for the matrix $(a_{c,p})$ are generated under the fairness of H_0 hypothesis. To model these scenarios, it is proposed to use the algorithm presented below:

– *Step 1*. Set the number of iterations, $K = 1000$

– *Step 2*. Let denote through S_n the set of all the bijective mappings of the set $\{1, \dots, n\}$ into itself. The number of elements in the set S_n equals to $n!$. Let F be a random permutation that can take any values from the set S_n:

$$P\left(F = \begin{pmatrix} 1 & \dots & n \\ f(1) & \dots & f(n) \end{pmatrix}\right) = \frac{1}{n!}$$

where $(f(1), \dots, f(n)) \in S_n$.

We generate a sample of size mK of independent random permutations for the ordered set $(1, \dots, n)$, where m is the number of sectors.

$$\hat{F}_{(k-1)m+j} = \begin{pmatrix} 1 & \dots & n \\ f_{(k-1)m+j}(1) & \dots & f_{(k-1)m+j}(n) \end{pmatrix}, j = 1, \dots, m.$$

– *Step 3*. For each scenario $k = 1, \dots, K$, using the permutations $\hat{F}_{(k-1)m+j}, j = 1, \dots, m$ generated in the previous step, calculate the elements of matrix $(\hat{a}_{c,p}^{(k)})$, which are defined as follows:

$$\hat{a}_{c_i p_j}^{(k)} = a_{c_{f_{(k-1)m+j}(i)}, p_j}$$

Note that

$$\sum_{i=1}^{n} \hat{a}_{c_i p_j}^{(k)} = \sum_{i=1}^{n} a_{c_i p_j} \text{ for all } j \in \{1, \dots, m\}. \qquad (6)$$

Moreover, if for scenario k there is such region c_i for which $\sum_{j=1}^{m} \hat{a}_{c_i p_j}^{(k)} = 0$, such a scenario is excluded from consideration.

– *Stage 2*: for each of the generated scenarios the values of the characteristics of interest are calculated. For example, the embedding characteristics.

– *Stage 3*: approximation of the distribution of the characteristics of interest is constructed using the empirical function of distribution of values obtained at Stage 2.

It should be noted that the obtained distribution function approximation for the embedding characteristics distribution $\frac{w_{i,j}}{w_{i,i}}$ do not depend on *i* and *j*, where *i* and *j* correspond to regions. Besides, the distribution functions of regional diversification coefficients obtained as a result of the modeled scenarios coincide for all regions. Using the obtained empirical distribution function it is possible to check the hypothesis about equality to zero of the embedding characteristics for the structure of strong sectors of the region $c_j, j = 1, \dots, n, j \neq i$, namely:

$$H_0: \frac{w_{i,j}}{w_{i,i}} = 0 \text{ vs } H_1: \frac{w_{i,j}}{w_{i,i}} \neq 0 \qquad (7)$$

at the given significance level $\alpha = 0{,}95$. 95%-quantile for the distribution $\frac{w_{i,j}}{w_{i,i}}$ under the null hypothesis equals to 0,47.

Graphs region-region by embedding degree characteristics. Let us define matrix $L^{signif} = (l_{i,j}^{signif} | i, j \in \{c_1, \dots, c_n\})$, which is defined as follows:

$$l_{i,j}^{signif} = \begin{cases} \frac{w_{i,j}}{w_{i,i}}, \text{if } \frac{w_{i,j}}{w_{i,i}} > 0{,}47 \\ 0, \; otherwise \end{cases} \quad (8)$$

Let L^{signif} be the adjacency matrix for the graph of "significant embedding characteristics" $G^{signif} = (V^{signif}, E^{signif})$ with directed weighted edges, see Figure 1.

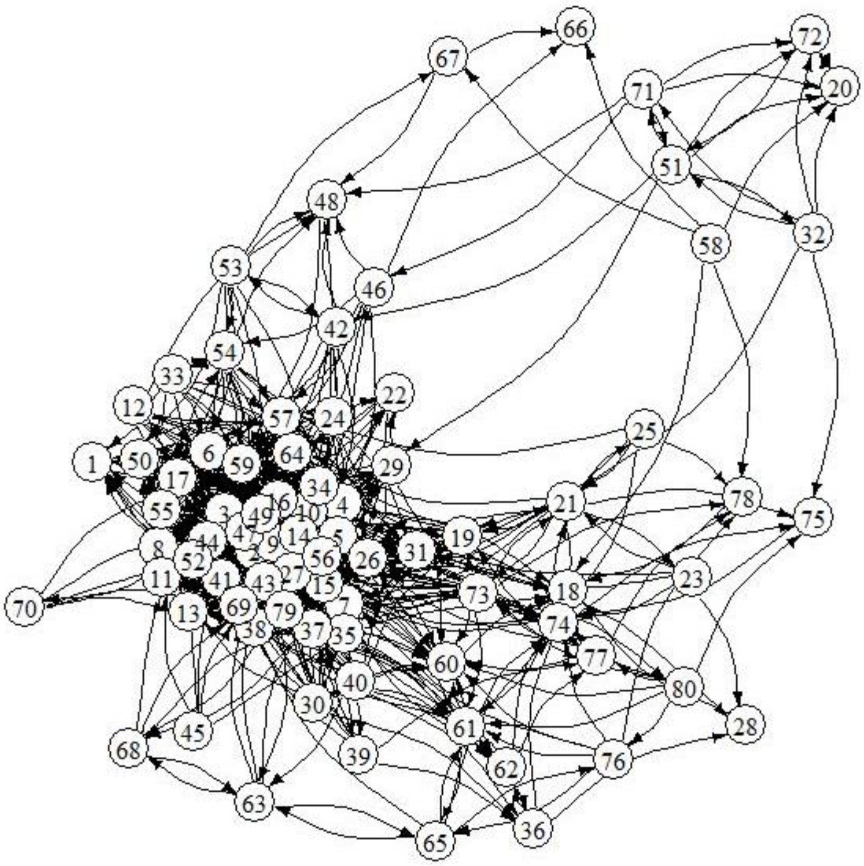

Figure 1: Graph of "significant embedding characteristics", $G^{signif} = (V^{signif}, E^{signif})$.

This graph reflects statistically significant embedding of strong sectors structures. It should be noted that the node numbers correspond to the numbering of regions used in the articles (Afanasiev, Kudrov, 2020; Ayvazyan, Afanasiev, Kudrov, 2018). The largest number of incoming edges on the graph of "significant embedding characteristics" corresponds to: 1) node 47 (Chuvash Republic) with the number of incoming edges equal to 45; 2) node 15 (Tver region) with the number of incoming edges equal to 47; 3) node 69 (Novosibirsk region) with the number of incoming edges equal to 42. These regions are characterized by the fact that their economies have the most number of adjacent sectors, which are replicated to varying degrees by regions corresponding either directly or indirectly to the incoming edges.

The model of recommendations forming for the development of the economic structure. At present, algorithms of recommendations formation are widely used in the selection process (see, for example (Glodberg et al., 2001). A very general statement of this task consists in recommending those variants of choice that correspond to the characteristics of the object for which recommendations are made to the greatest extent. In the case of the structure of strong sectors of the regional economies, recommendations give a list of potentially achievable strong sectors with the given characteristics of the region, as well as using information on strong sectors in regions similar to it. More formally, possible ways of implementing this procedure are presented below. Define the random variables:

$$A_{c_i p_k} = \begin{cases} 1, with\ probability\ \pi(f^1_{c_i p_k}, f^2_{c_i p_k}, \dots) \\ 0, with\ probability\ 1 - \pi(f^1_{c_i p_k}, f^2_{c_i p_k}, \dots) \end{cases} \qquad (9)$$

where $f^1_{c_i p_k}, f^2_{c_i p_k}, \dots$ — factors, impacting the appearance of the events $A_{c_i p_k} = 1$ and $A_{c_i p_k} = 0$. Next, we'll assume that the function $\pi(f^1_{c_i p_k}, f^2_{c_i p_k}, \dots)$ has the following form:

$$\pi(f^1_{c_i p_k}, f^2_{c_i p_k}, \dots) = \frac{\exp(\beta_0 + \sum \beta_h f^h_{c_i p_k})}{1 + \exp(\beta_0 + \sum \beta_h f^h_{c_i p_k})}, \beta_0 - constant \qquad (10)$$

Suppose that a sequence $a_{c_1 p_1}, \dots, a_{c_1 p_m}, \dots, a_{c_n p_1}, \dots, a_{c_n p_m}$ is the realization of a random sequence $A_{c_1 p_1}, \dots, A_{c_1 p_m}, \dots, A_{c_n p_1}, \dots, A_{c_n p_m}$, whose elements are assumed to be independent when $f^1_{c_1 p_1}, f^1_{c_1 p_2}, \dots$ are known. Moreover, we assume that $P(A_{c_i p_k} | f^1_{c_1 p_1}, f^1_{c_1 p_2}, \dots) = P(A_{c_i p_k} | f^1_{c_i p_k}, \dots, f^h_{c_i p_k}, \dots)$, i.e. the factors affecting the distribution of $A_{c_i p_k}$ are limited by the set $f^1_{c_i p_k}, \dots, f^h_{c_i p_k}, \dots$, with the lower pair of indices (c_i, p_k). Then likelihood function has the following form:

$$P(A_{c_1 p_1} = a_{c_1 p_1}, \dots, A_{c_1 p_m} = a_{c_1 p_m}, \dots, A_{c_n p_1} = a_{c_n p_1}, \dots, A_{c_n p_m} = a_{c_n p_m} | f^1_{c_1 p_1}, f^1_{c_1 p_2}, \dots) =$$

$$P(A_{c_1 p_1} = a_{c_1 p_1} | f^1_{c_1 p_1}, \dots, f^h_{c_1 p_1}, \dots) \cdot \dots \cdot P(A_{c_n p_m} = a_{c_n p_m} | f^1_{c_n p_m}, \dots, f^h_{c_n p_m}, \dots) =$$

$$\left[\frac{\exp(\beta_0 + \sum \beta_h f^h_{c_1 p_1})}{1 + \exp(\beta_0 + \sum \beta_h f^h_{c_1 p_1})}\right]^{a_{c_1 p_1}} \left[\frac{1}{1 + \exp(\beta_0 + \sum \beta_h f^h_{c_1 p_1})}\right]^{1 - a_{c_1 p_1}} \cdot$$

$$\dots \cdot \left[\frac{\exp(\beta_0 + \sum \beta_h f^h_{c_n p_m})}{1 + \exp(\beta_0 + \sum \beta_h f^h_{c_n p_m})}\right]^{a_{c_n p_m}} \left[\frac{1}{1 + \exp(\beta_0 + \sum \beta_h f^h_{c_n p_m})}\right]^{1 - a_{c_n p_m}}.$$

By maximizing this likelihood function, we obtain the estimates of parameters $\beta_1, \dots, \beta_h, \dots$ To assess the quality of the model, we will use the Akaike criterion as well as the following characteristics: $G = \frac{1}{mn} \sum_{i,k} I(I(\pi(f^1_{c_i p_k}, f^2_{c_i p_k}, \dots) > 0.5) = a_{c_i p_k})$,

where $I(A)$ — indicator of the event A. Note that G characterizes the proportion of cases correctly identified by the model.

Consider three possible ways to form explanatory variables $f^1_{c_i p_k}, \dots, f^h_{c_i p_k}, \dots$:

Form 1: for this form $h = 1$. Let us define the metric of difference between the structures of strong sectors of regions c_i and c_j, $i \neq j$, as follows (see (Afanasiev, Kudrov,2020)):

$$d(\mathbf{a}(c_i), \mathbf{a}(c_j)) = \sqrt{1 - \left(\frac{2(\mathbf{a}(c_i), \mathbf{a}(c_j))}{|\mathbf{a}(c_i)| + |\mathbf{a}(c_j)|}\right)^2},$$

where vectors

$\mathbf{a}(c_i) = (a_{c_i p_1}, \dots, a_{c_i p_m})$

$\mathbf{a}(c_j) = (a_{c_j p_1}, \dots, a_{c_j p_m})$

determine the structure of strong sectors in the regions c_i and c_j, respectively; the operation $(\cdot,\cdot)$ is a scalar product; the operation $|\cdot|$ applied to a vector gives the sum of its coordinates.

If the structures of strong sectors of two regions c_i and c_j coincide, then $\mathbf{a}(c_i) = \mathbf{a}(c_j)$ and $d(\mathbf{a}(c_i), \mathbf{a}(c_j)) = 0$. If the structures of strong sectors do not intersect, then the scalar product

$\left(\mathbf{a}(c_i),\mathbf{a}(c_j)\right) = 0$ and the distance between these structures is the highest possible value, $d\left(\mathbf{a}(c_i),\mathbf{a}(c_j)\right) = 1$. Now define the factors $f^1_{c_i,p_k}, \dots$ for all $i = 1, \dots, n; k = 1, \dots, m$:

$$f^1_{c_i,p_k} = \sum_{j=1, j\neq i}^{n} \left(1 - d\left(\mathbf{a}(c_i),\mathbf{a}(c_j)\right)\right) a_{c_j,p_k}.$$

Factor $f^1_{c_i,p_k}$ takes on large values in case the sector p_k is encountered quite often in regions whose strong sectors structure is close to that for region c_i.

Table 1. The parameters estimates for model of the strong sectors identification (Form 1).

	Estimate	*Standard deviation*	*t-statistic*	*p-value*
Constant	-2,87	0,06	-44,12	0,00
$f^1_{c_i,p_k}$	0,19	0,01	36,63	0,00

G	0,8
AIC	5882

As it can be seen from the results presented in Table 1, the model correctly identifies 80% of the total number of cases.

Form 2: for this form $h = 2$. Define the factors $f^1_{c_i,p_k}, f^2_{c_i,p_k}$ for all $i = 1, \dots, n; k = 1, \dots, m$:

$$f^1_{c_i,p_k} = \frac{\sum_{j=1, j\neq i}^{n} w_{i,j} a_{c_j,p_k}}{1 - w_{i,i}};$$

$$f^2_{c_i,p_k} = \frac{\sum_{j=1, j\neq i}^{n} \frac{w_{j,i}}{w_{j,j}} a_{c_j,p_k}}{\sum_{j=1, j\neq i}^{n} \frac{w_{j,i}}{w_{j,j}}}.$$

For this form of explanatory factors formation the estimates of the model parameters (9) - (10) are presented in Table 2:

Table 2. The parameters estimates for model of the strong sectors identification (Form 2).

	Estimate	*Standard deviation*	*t-statistics*	*p-value*
Constant	-2,94	0,07	-43,42	0,00
$f^1_{c_i,p_k}$	2,91	0,83	3,52	0,00
$f^2_{c_i,p_k}$	18,15	0,91	20,12	0,00

G	0,8
AIC	5681

As it can be seen from Table 2, the coefficients at f_{c_i,p_k}^1 and f_{c_i,p_k}^2 are significant and positive. The model correctly identifies 80% of the total number of cases and has nearly the same value for the Akaike criterion as for the previous Form 1.

Form 3: for this form $h = 2$. Define the factors $f_{c_i,p_k}^1, f_{c_i,p_k}^2$ for all $i = 1, \dots, n; k = 1, \dots, m$:

$$f_{c_i,p_k}^1 = \frac{\sum_{j=1, j\neq i}^{n} w_{i,j} I\left(\frac{w_{i,j}}{w_{i,i}} > 0.47\right) a_{c_j,p_k}}{\sum_{j=1, j\neq i}^{n} w_{i,j} I\left(\frac{w_{i,j}}{w_{i,i}} > 0.47\right)};$$

$$f_{c_i,p_k}^2 = \frac{\sum_{j=1, j\neq i}^{n} \frac{w_{j,i}}{w_{j,j}} I\left(\frac{w_{j,i}}{w_{j,j}} > 0.47\right) a_{c_j,p_k}}{\sum_{j=1, j\neq i}^{n} \frac{w_{j,i}}{w_{j,j}} I\left(\frac{w_{j,i}}{w_{j,j}} > 0.47\right)}.$$

Estimates of the model parameters (9) - (10) with the factors $f_{c_i,p_k}^1, f_{c_i,p_k}^2$ are presented in Table 3:

Table 3. The parameters estimates for model of the strong sectors identification (Form 3).

	Estimate	*Standard deviation*	*t-statistics*	*p-value*
Constant	-3,22	0,08	-41,78	0,00
f_{c_i,p_k}^1	2,93	0,15	19,30	0,00
f_{c_i,p_k}^2	3,26	0,15	21,24	0,00

G	0,83
AIC	4592

As it can be seen from Table 3, the coefficients at f_{c_i,p_k}^1 and f_{c_i,p_k}^2 are significant and positive. For this form of explanatory variables, the proportion of cases correctly identified by the model differs slightly from the proportion for Form 1 and Form 2, it equals to 83% of the total number of cases. However, the value of Akaike criterion is somewhere 21% less than for Form 1 and Form 2 of defining the explanatory variables. From that we could conclude that filtering and extracting significant embedding degree characteristics improve probabilistic properties of the model.

2. DISCUSSION OF THE RESULTS

Relationship between the indicators of aggregated embedding and economic basis. In the paper (Ayvazyan et al., 2018) it is presented the concept of economic basis, the components of which are characteristics of differentiation formed by theoretically justified models of regional development. It includes a scale of economy, technical efficiency of production, index of sectoral specialization (based on the first principal component of GRP structure), industrialization index (based on the second principal component of GRP structure), a trend of technical efficiency. The position of a region on the basis of differentiation characteristics determines its economic uniqueness. For a more detailed description, see the paper (Ayvazyan et al., 2018). Let us consider the regression of the logarithm of the indicator of the embedding degree of the economy's structure into the components of the economic basis.

Table 4: Regression of the logarithm of the aggregate embedding of the structure indicator on the economic basis.

	Estimate	*Standard deviation*	*t-statistic*	*p-value*
Constant	2,19	0,04	59,52	0,00
Scale of economy	-0,18	0,04	-4,51	0,00
Technical efficiency	-0,07	0,04	-1,67	0,10
Sectoral specialization index	-0,18	0,04	-4,65	0,00
Industrialization index	-0,05	0,04	-1,36	0,18
Trend of technical efficiency	0,04	0,04	1,08	0,29

R-squared	0,43
Ads.R-squared	0,40

Regression parameter estimates of logarithm $I_i^{(1)}$ on the economic basis allow us to see that a small degree of embedding of the regional economy structure is typical for large economies with specialization in the mining industry. The negative sign in the sectoral specialization index indicates that there are no adjacent sectors localized in regions with specialization in the mining industry, which forms value-added from mining. A high degree of embedding is typical for small economies with specialization in manufacturing. Thus, regions with specialization in the manufacturing industry are characterized by the formation of value chains localized in these regions.

Table 5. Regression of the logarithm of the aggregate embedding in the structure indicator on the economic basis.

	Estimate	*Standard deviation*	*t-statistic*	*p-value*
Constant	2,13	0,05	42,47	0,00
Scale economy	-0,03	0,05	-0,54	0,59
Technical efficiency	0,01	0,05	0,24	0,81
Sectoral specialization index	-0,32	0,05	-5,97	0,00
Industrialization index	0,04	0,05	0,69	0,50
Trend of technical efficiency	0,06	0,05	1,23	0,22

R-squared	0,3552
Ads.R-squared	0,3116

Regression parameter estimates of logarithm $I_i^{(2)}$ on the economic basis show that the significant explanatory variable is only sectoral specialization index. A high degree of embedding in the structure of the region's economy is typical for economies with specialization in the manufacturing industry, and a low degree of embedding in the structure is typical for economies with specialization in the mining industry. It also points to the formation of localized value chains in regions with specialization in the manufacturing industry.

Relationship of economic complexity with the indicators of aggregated embedding. As mentioned above, the economic complexity of a region is determined by strong sectors in the structure of its economy. The article (Afanasyev, Kudrov, 2020) presents the estimates of economic complexity for the Russian Federation's regions, for calculating which the data on tax receipts in 82 sectors of the economy for the regions of the Russian Federation were used. Table 6 presents estimates of parameters in the regression of economic complexity of Russian regions obtained in the paper (Afanasyev and Kudrov, 2020) on the logarithm of aggregated embedding in the structure of the economy's strong sectors and the logarithm of aggregated embedding of the structure.

Table 6. Regression for the economic complexity of regions from the article

	Estimate	Standard deviation	t-statistic	p-value
Constant	-4,18	0,45	-9,22	0,00
Log(aggregated embedding of the structure)	0,96	0,19	5,14	0,00
Log(aggregated embedding in the structure)	0,97	0,15	6,65	0,00

R-squared	0,55
Ads.R-squared	0,54

Source: Afanasyev and Kudrov, 2020

As can be seen from the results, the high level of economic complexity corresponds to the high level of embedding degree in the economy's structure simultaneously with the high level of embedding degree of the structure. In contrast, a low level of economic complexity corresponds to a low level of the index of the embedding of the economy structure simultaneously with a low level of the index of the embedding in the structure. Thus, the economic complexity index can be interpreted as a measure of the level of development value chains or adjacent sectors.

3. RECOMMENDATIONS

The constructed model allows us to explain the emergence and absence of a strong sector in the structure of the economy of each region. For each region, the probability of emergence of a specific sector as a strong one in its structure is estimated. On this basis, a list of sectors recommended for priority development in a region has been compiled. Based on the list of recommended sectors and embedding structure shown in Figure 1, groups of adjacent sectors have been formed according to the following procedure, which is presented below in details for the agricultural sector:

a. Form a list of regions with specialization in agriculture according to the methodology presented in the paper (Aivazian et al., 2016).

b. Using the model (10) estimated in Form 3, calculate the probability of occurrence of each sector of the nomenclature under consideration $p_1, \dots, p_m$ for each region from the list formed at the previous step.

c. For each sector calculate the average probability of its occurrence by the regions from the list formed at step 1.

d. Take only those sectors where the average probability of occurrence exceeds 50%.

The resulting set of sectors will be called adjacent sectors for the agricultural sector. It includes the following sectors:

- crop and livestock breeding, hunting;
- beverage industry;
- dairy production;
- food production;
- processing and preservation of meat and meat-based food products;
- textile manufacturing;
- leather and leather goods production;
- wood processing and manufacturing;
- transportation and storage;
- wholesale and retail;

– electricity, gas and steam supply; air conditioning;
– water treatment and distribution.

For the majority of regions specializing in agriculture, the development of the "Transportation and Storage" sector is required, which will allow for greater territorial coverage in the supply of manufactured products. In the future, national and international commodity flows are expected to increase, and improved accessibility and physical expansion of transport infrastructure are necessary to make optimal use of this potential. Given other infrastructure problems for agricultural development, the recommendations include the development of the sector "Water treatment and distribution", which is necessary for the functioning of irrigation systems, as well as the sector "Electricity, gas and steam supply; air conditioning" for the development of greenhouse complexes, which require high energy consumption in creating a microclimate. Besides, in some regions with specialization in agriculture, the manufacturing sectors for agricultural products are not well developed: beverage industry (Vladimir region, Pskov region, Voronezh region, Kostroma region, Orel region, Smolensk region), processing and preservation of meat and meat food products (Ivanovo region, Republic of Adygeya), food production (Kostroma region, Republic Dagestan), dairy products (Smolensk region, Republic Dagestan). It should also be noted that the development of agriculture contributes to the demand for the development of the agricultural machinery sector, as well as many other sectors. Including sectors that offer solutions that promote process automation both at the level of agricultural work and at the level of the manufacturing industry health and food safety are also essential factors that should determine the technological features of production.

The advantages of complementarities between adjacent sectors can be enhanced by the advantages of being located within the same region. Stimulating the development of adjacent sectors requires not only economic policy solutions, but also education, labor mobility, and housing. It also implies close cooperation and coordination to the maximum extent possible between state authorities, entrepreneurs and educational institutions. Connected sectors stimulate cooperation and innovation, in the sense that proximity within a region ensures mutual diffusion of knowledge at relatively low costs. Also, strategic coalitions between economically similar regions and cooperation with universities and scientific institutions can be very useful, facilitating the diffusion of new, additional knowledge from outside.

CONCLUSION

It is shown a probabilistic interpretation of the matrix elements used to estimate the economic complexity (Hausmann, Hidalgo, Bustos, Coscia, Simoes, Yildirim, 2011). Their properties are given, based on which aggregated indicators are introduced, characterizing the embedding of strong sectors structures of regional economies and explaining the economic complexity. Economic complexity is higher in those regions that have a high degree of embedding. On the other hand, embedding indices prove to be statistically significantly related to some components of the economic basis presented in the paper (Ayvazyan, Afanasiev, Kudrov, 2018). The most significant relationship is revealed with the index of sectoral specialization, which indicates that large values of embedding indicators correspond to regions with specialization in the manufacturing industry.

Restrictions on embedding characteristics that meet the scenarios of random embedding of strong sectors of regional economies are obtained. Random embedding structure formation corresponds to the scenario of the absence of adjacent sectors linking several sectors. Using these constraints, a graph of embedding structures of regional economies was constructed, where insignificant embedding was excluded. This graph is a convenient way to visually present information about embedding structures of strong sectors of regional economies.

It is considered a model with the different sets of explanatory variables which explains the emergence and absence of a strong sector in the structure of the region's economy. On its basis the condition of the appearance of a certain strong sector in the structure of the economy of a particular region with the probability exceeding 50% has been obtained. This condition has been used for forming the list of sectors recommended for priority development in the region. For each region, the probability of a specific sector's appearance as a strong one in its economy's structure is estimated. Based on ordering the sectors by the value of these probabilities and estimations of their potential contribution to social and economic development, an expert estimation of expediency of development of a new strong sector in the region can be made.

REFERENCES

Aivazian, S.A., Afanasiev, M.Y., Kudrov, A.V. (2016), "Clustering methodology of the Russian Federation regions with account of sectoral structure of GRP", *Applied Econometrics*, Vol. 41, pp. 24–46 (in Russian).

Aivazian, S.A., Afanasiev, M.Y., Kudrov, A.V. (2018), "Indicators of economic development in the basis of the characteristics of regional differentiation", *Applied Econometrics*, Vol. 50, pp. 4–22. (in Russian).

Afanasiev, M.Y., Kudrov, A.V. (2020), "Estimates of Economic Complexity in the Structure of the Regional Economy", *Montenegrin Journal of Economics,* Vol. 16, No. 4, pp. 61-72.

Arrow, K.J. (1962), "The Economic Implications of Learning by Doing", *The Review of Economic Studies*, Vol. 29, No. 3, pp. 155–173.

Beaudry, C., Schiffauerova, A. (2009), "Who's right, Marshall or Jacobs? The localization versus urbanization debate", *Research Policy*, Vol. 38, pp. 318–337.

Blien, U., Suedekum, J. (2005), "Local Economic Structure and Industry Development in Germany, 1993-2001", *Economics Bulletin*, Vol. 15 (17), pp. 1–8

Blien, U., Wolf, K. (2006), "Local employment growth in West Germany: A dynamic panel approach", *Labour Economics*, Vol. 13, No. 4, pp. 445–458.

Combes, P. (2000), "Economic Structure and Local Growth: France, 1984-1993", *Journal of Urban Economics*, Vol. 47, No. 3, pp. 329–355.

Dekle, R. (2002), "Industrial concentration and regional growth: evidence from the prefectures", *Review of Economics and Statistics*, Vol. 84, No. 2, pp. 310–315.

de Lucio, J.J., Herce, J.A., Goicolea, A. (2002), "The effects of externalities on productivity growth in Spanish industry", *Regional Science and Urban Economics*, Vol. 32, No. 2, pp. 241–258.

Frenken, K, Van Oort, F.G., Verburg, T. (2007), "Related Variety, Unrelated Variety and Regional Economic Growth", *Regional Studies*, Vol. 41, No. 5, pp. 685-697.

Fuchs, M. (2011), "The determinants of local employment dynamics in Western Germany", *Empirical Economics*, Vol. 40, No. 1, pp. 177–203.

Glaeser, E., Kallal, H., Scheinkman, J., Shleifer, A. (1992), "Growth in Cities", *Journal of Political Economy*, Vol. 100, No. 6, pp. 1126–1152.

Glodberg K., Roeder T., Gupta D., Perkins C. (2001), "Eigentaste: A Constant Time Collaborative Filtering Algorithm", *Information Retrieval*, Vol. 4, pp. 133–151.

Hausmann, R., Hidalgo C. et al. (2011), *The Atlas of Economic Complexity: Mapping Pathsto Prosperity*, Center for International Development, Harvard University, MIT, Cambridge.

Hausmann R., Klinger B. (2006). "Structural Transformation and Patterns of Comparative Advantage in the Product Space". *CID Working Paper*, No. 128.

Hausmann, R., Rodrik, D. (2003), "Economic development as selfdiscovery", *Journal of Development Economics*, Vol. 72, No. 2, pp. 603–633.

Hidalgo C. A., Klinger B., Barabási A.-L., Hausmann R. (2007) "The Product Space Conditions the Development of Nations", *Science*, Vol. 317, No. 5837, pp. 482-487

Henderson, V., Kuncoro, A., Turner, M. (1995), "Industrial Development in Cities", *Journal of Political Economy*, Vol. 103, No. 5, pp. 1067–1090.

Illy, A., Schwartz, M., Hornych, C., Rosenfeld, M. (2011), "Local economic structure and sectoral employment growth in German cities", *Journal of Economic and Social Geography*, Vol. 102, No. 5, pp. 582–593.

Kemp-Benedict. E. (2014), "An interpretation and critique of the method of reflections", *MPRA Paper*, No. 60705, https://mpra.ub.uni-muenchen.de/60705/1/MPRA_paper_60705.pdf

Kluge, J., Lehmann, R. (2012), Marshall or Jacobs? Answers to an unsuitable question from an interaction model, *Ifo Working Paper*, No. 124, Institute for Economic Research at the University of Munich, Munich

Krugman, P. R. (2011), "The New Economic Geography, Now Middle-aged", *Regional Studies*, Vol. 45, No. 1, pp. 1–7.

Lee, B., Sosin, K., Hong, S. (2005), "Sectoral Manufacturing Productivity Growth in Korean Regions", *Urban Studies*, Vol. 42, No. 7, pp. 1201–1219.

Makarov, V.L. (2010), *Social'nyj klasterizm. Rossijskij vyzov*. Biznes Atlas, Moskow (in Russian

Marshall, A. (1890), *Principles of Economics*, MacMillan, London.

Neffke, F., Henning, M., Boschma, R. (2011), "How do regions diversify over time? Industry relatedness and the development of new growth paths in regions", *Economic Geography*, Vol. 87, No. 3, pp. 237–265.

Romer, P.M. (1986), "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, Vol. 94, No. 5, pp. 1002–1037.