Forecast of the appearance of new strong sectors in the regional economy based on a probabilistic model

Mikhail Afanasiev*, and *Aleksander* Kudrov

Central Economics and Mathematics Institute of the Russian Academy of Sciences, Moscow, Russia

Abstract The problem of forecasting the appearance of new strong sectors in the region shall be considered. Based on the methods of probabilistic and statistical modeling, a model has been built that makes it possible to assess the occurrence probability of a new strong sector in the region, taking into account the characteristics of the structure of the economy. The possibility of building such a model is based on the assumption that the appearance and development of sectors are largely due to the evolution of past economic activity. The model uses the indicators of nesting of structures of strong sectors of regional economies entered by the authors. These values are based on the probabilistic interpretation and properties of the matrix elements, which assesses the economic complexity, in accordance with the traditional approach. The condition for the appearance of a certain strong sector in the structure of the economy of a particular region with a probability exceeding 0.5 shall be obtained. This condition used to form a list of sectors recommended for priority development in the region. For each region in its structure, the occurrence probability of a specific sector as a strong one was estimated. On the basis of ordering the sectors by the value of these probabilities and assessing their potential contribution to socio-economic development, an expert assessment of the feasibility of developing a new strong sector in the region can be given. Research is focused on the development of theories of localized specialization and economic diversification.

1 Introduction

Let's define an indicatorRCA_{cp} of the revealed comparative advantages:

$$RCA_{cp} = \frac{y_{cp} / \Sigma_p y_{cp}}{\Sigma_c y_{cp} / \Sigma_{c,p} y_{cp}} \quad (1)$$

where y_{cp}— the output of the sector p of economy of the region c. Value RCA_{cp} is the ratio of the share of production from the sector p in the total volume of production from all sectors of the region's economy c to the share of production of the sector p for all regions in the volume of production from all sectors of the economy of all regions. According to the

* Corresponding author: mi.afan@yandex.ru

work (Hausmann & Klinger, 2006), to identify comparative advantages in economies, an indicator is usedRCA_{cp} for which a condition of the type of restriction from below is checked [7]:

$$a_{c,p} = \begin{cases} 1, if\ RCA_{cp} \geq 1 \\ 0, if\ RCA_{cp} < 1 \end{cases}$$

The matrix $A = (a_{c,p})$ contains the data on economic sectors that are developed in various regions at the level of revealed comparative advantages, determined using expression (1). Lines of this matrix correspond to regions, columns — to sectors of economy. Let us call further a vector$\left(a_{c,p_1}, \dots, a_{c,p_m}\right)$ the *structure of strong sectors* of economy of the region$c$.

Let $R_1 = (r_{c.p})$ be a matrix with elements $r_{c.p} = \frac{a_{c.p}}{k_{c.0}}$, where $k_{c.0} = \sum_p a_{c.p}$.

Let $R_2 = (r^*_{p,c})$ be a matrix with elements $r^*_{p,c} = \frac{a_{c,p}}{k_{p,0}}$, where $k_{p,0} = \sum_c a_{c,p}$.

According to (Afanasiev & Kudrov, 2020; Hausmann et al., 2011), the economic complexity of a region is defined as an eigenvector of the matrix R_1R_2 [1, 6]. Note that the element $w_{i,j}$ at the intersection of the *i*-th row and the *j*-th column of the matrix R_1R_2, that is $(R_1R_2)_{ij}$, is given by the formula:

$$\frac{1}{k_{c_i,0}} \sum_t \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}}$$

Let's note some properties of matrix $w_{i,j}$ elements:

1. Values $w_{i,1}$, ..., $w_{i,n}$ may be interpreted as a probability distribution. The matrix R_1R_2 is stochastic, since the elements of the matrix are non-negative, and their row sum is 1. Therefore, for each $i, j \in \{1, \dots, n\}$: $w_{i,j} \geq 0$, $\sum_{j=1}^{m} w_{i,j} = 1$.
2. If a region c_i holds one strong region, then $w_{i,i} > 0$. Otherwise, $w_{i,i} = 0$.

It is easy to show justice of this statement 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 the 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.

3. Elements $w_{i,j}$ are equal to zero if and only if the condition is satisfied:

$$\{t: a_{c_i,p_t} = 1\} \cap \{t: a_{c_j,p_t} = 1\} = \emptyset$$

Performance of this condition means lack of the general strong sectors in structures of economy of regions c_i and c_j.

4. In each row of the matrix $(w_{i,j})$ the maximum element corresponds to the diagonal element, that is $w_{i,i} = \max_{j \in \{1,\dots,n\}} (w_{i,j})$. Let us show it.

Because of justice:

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

we receive that $w_{i,j} \leq w_{i,i}$. Moreover, equality in the last 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 of the region c_i are strong sectors in the structure of the region c_j as well. If this condition is not met, then we have a strict inequality:

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

5. Asymmetry of a matrix $(w_{i,j})$.

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

From properties (1-5) for $w_{i,j}$ it follows that the ratio $\frac{w_{i,j}}{w_{i,i}}$ can be interpreted as a characteristic of *nesting level* the region's set of strong sectors c_i into the set of strong sectors for the region c_j. The lesser the ratio is, the lesser is the number of strong sectors of the region c_i in the set $\{t: a_{c_j,p_t} = 1\}$ of strong sectors of the region c_j.

The matrices $(a_{i,j})$ and $(w_{i,j})$, used below, as well as estimates of the economic complexity of the regions of the Russian Federation, calculated on the basis of data on tax revenues for 82 sectors of the economy, are presented in (Afanasyev, Kudrov, 2020) [1].

2 Methods

At present, algorithms for generating recommendations for selection shall be widely used [5]. A very general formulation of this task consists in recommending those choices that correspond to the greatest extent to the characteristics of the object for which the recommendations are being developed. For the structure of the strong sectors of the region's economy, the recommendations give a list of potentially achievable strong sectors given the characteristics of the region, as well as using information about strong sectors in similar regions.

Let the random variables

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

where $f^1_{c_i,p_k}, f^2_{c_i,p_k}, \dots -$ are the factors influencing the occurrence of events $A_{c_i,p_k} = 1$ and $A_{c_i,p_k} = 0$.. In what follows, we will assume that the dependence $\pi\left(f^1_{c_i,p_k}, f^2_{c_i,p_k}, \dots\right)$ has the form:

$$\pi\left(f^1_{c_i,p_k}, f^2_{c_i,p_k}, \dots\right) = \frac{e^{\beta_0 + \sum \beta_h f^h_{c_i,p_k}}}{1 + e^{\beta_0 + \sum \beta_h f^h_{c_i,p_k}}}, \text{где } \beta_0 - constant \quad (3)$$

Suppose that the 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 a realization of a random sequence $A_{c_1,p_1},\dots,A_{c_1,p_m},\dots,A_{c_n,p_1},\dots,A_{c_n,p_m}$, the elements of which are assumed to be independent when $f^1_{c_1,p_1}, f^1_{c_1,p_2},\dots$ are given. In addition, suppose that distribution $P(A_{c_i,p_k}|f^1_{c_1,p_1}, f^1_{c_1,p_2},\dots) = P(A_{c_i,p_k}|f^1_{c_i,p_k},\dots,f^h_{c_i,p_k},\dots)$, that is, the factors influencing distribution A_{c_i,p_k}, are limited to the set $f^1_{c_i,p_k},\dots,f^h_{c_i,p_k},\dots$, which have the lower pair of indices (c_i,p_k). Then the likelihood function has the form:

$$\begin{aligned}P\big(A_{c_1,p_1}=a_{c_1,p_1},\dots,A_{c_1,p_m}=a_{c_1,p_m},\dots,A_{c_n,p_1}=a_{c_n,p_1},\dots,A_{c_n,p_m}\\ =a_{c_n,p_m}|f^1_{c_1,p_1},f^1_{c_1,p_2},\dots\big)\\ =P\big(A_{c_1,p_1}=a_{c_1,p_1}|f^1_{c_1,p_1},\dots,\ f^h_{c_1,p_1},\dots\big)\cdot\dots\\ \cdot P\big(A_{c_n,p_m}=a_{c_n,p_m}|f^1_{c_n,p_m},\dots,\ f^h_{c_n,p_m},\dots\big)\end{aligned}$$

$$=\frac{e^{a_{c_1,p_1}(\beta_0+\sum\beta_h f^h_{c_1,p_1})}}{(1+e^{\beta_0+\sum\beta_h f^h_{c_1,p_1}})}\cdot\dots\cdot\frac{e^{a_{c_n,p_m}(\beta_0+\sum\beta_h f^h_{c_n,p_m})}}{(1+e^{\beta_0+\sum\beta_h f^h_{c_n,p_m}})}$$

By maximizing this likelihood function, we obtain estimates of the parameters $\beta_1,\dots,\beta_h,\dots$. To assess the quality of the model, we will use the Akaike criterion, as well as the value:

$$G=\frac{1}{mn}\sum_{i,k}I\big(I\big(\pi\big(f^1_{c_i,p_k},f^2_{c_i,p_k},\dots\big)>0.5\big)=a_{c_i,p_k}\big)$$

where $I(A)-$ is event indicator A. The value G characterizes the proportion of cases correctly identified by the model.

Consider the ways of generating explanatory variables $f^1_{c_i,p_k},\dots,f^h_{c_i,p_k},\dots$

Method 1: for this method $h=2$. Let us define factors $f^1_{c_i,p_k}, f^2_{c_i,p_k}$ for all $i=1,\dots,n; k=1,\dots,m$:

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

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

Estimates of the parameters of the model (2)-(3) for this method of forming the explanatory factors are presented in Table 1.

Table 1. Estimation of the parameters of the strong sector identification model for Method 1.

	Estimate	Standarddeviation	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			

Source: (Compiled by the authors).

As can be seen from Table 1, the coefficients at $f^1_{c_i,p_k}$ and $f^2_{c_i,p_k}$ are significant and positive. For this method of choosing explanatory variables, the proportion of cases correctly identified by the model is 80%.

The simulation method was used to test the hypothesis that the index of nesting of the structure of strong sectors of the region $c_i, i = 1, \dots, n,$ in the structure of strong sectors of the region $c_j, j = 1, \dots, n, j \neq i$ is equal to zero, namely:

$$H_0: \frac{w_{i,j}}{w_{i,i}} = 0 \text{ contrary } H_1: \frac{w_{i,j}}{w_{i,i}} \neq 0$$

It is shown that the 95%-quantile for the distribution function $\frac{w_{i,j}}{w_{i,i}}$ under the conditions of validity of the null hypothesis is equal to 0.47 at a given level of significance $\alpha = 0{,}95$.

Method 2: for this method $h = 2$. Let us define factors $f^1_{c_i,p_k}, f^2_{c_i,p_k}$ for all $i = 1, \dots, n; k = 1, \dots, m$:

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

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

Estimates of the parameters of the model (2)-(3) for this method of forming the explanatory factors are presented in Table 2.

Table 2. Estimation of the parameters of the strong sector identification model for Method 2.

	Assessment	Standard deviation	t-statistics	p-value
Constant	-3.03	0.07	-43.33	0.00
$f^1_{c_i,p_k}$	4.16	0.25	16.62	0.00
$f^2_{c_i,p_k}$	6.70	0.28	24.26	0.00
G	0.83			
AIC	4950			

Source: (Compiled by the authors).

As can be seen from Table 2, the coefficients at $f^1_{c_i,p_k}$ and $f^2_{c_i,p_k}$ are significant and positive. For this method of choosing the explanatory variables, the proportion of cases correctly identified by the model does not differ significantly from the proportion for case 2 and is equal to 83% of the total number of cases, but the value of the Akaike criterion is less than for the first method of choosing the explanatory variables. Method 2 is preferred.

3 Results

As a result of using the model presented above, lists of potentially strong sectors in the regions were formed. For example, for a number of regions specializing in agriculture, the development of the Transport and Storage sector is required, which makes it possible to have a bigger territorial coverage in the supply of manufactured products. Due to infrastructural problems, recommendations for the development of agriculture are also in the development of the sector 'Water intake, purification and distribution', which is necessary for the

functioning of irrigation systems, as well as the sector 'Power supply, gas and steam; air conditioning', which is also required for the development of greenhouse complexes that require high energy consumption when creating a microclimate. It should also be noted that the development of agriculture contributes to the formation of demand for the development of the agricultural machinery sector, as well as many other sectors. As an example, Table 3 shows estimates of the occurrence probability of a sector 'Transportation and storage' in the regions as a strong sector. The range of identified potential strong sectors in the structures of regional economies based on the results of using the model is wider.

Table 3. Estimates of the occurrence probability of of the "transport and storage" sector as a strong.

Region	Estimate of probability
Belgorod region	0.89
Voronezh region	0.63
Kursk region	0.57
Moscow region	0.67
Saint Petersburg	0.92
Republic of Adygea	0.91

4 Discussions

There are two main theories describing the mechanism of knowledge creation and distribution: localized specialization and economic diversification. Localized specialization theory was first presented in detail in the work (Marshall, 1890) and argues that companies surrounded by other representatives of the same industry will grow faster due to the circulation of knowledge within the industry [9]. This theory was further developed in the works (Arrow, 1962; Romer, 1986) [2, 11]. The opposite theory and the resulting empirical estimates are presented in (Blien & Wolf, 2006; Fuchs, 2011; Illy, Schwartz, Hornych & Rosenfeld, 2011). According to this theory, companies benefit from the fact that they are faced with a heterogeneous environment, consisting of different industries, as new ideas come from the external environment [3, 4, 8]. Variety leads to economic growth through mechanisms called diversification. Researches on theories of localization or diversification suggest that a choice shall be made in favor of one or another.

The results obtained do not contradict the assumption that the effects of localization and diversification can complement each other and are not mutually exclusive. Estimates of the occurrence probability of a specific sector in the structure of the region as a strong one indicate that the appearence and development of sectors is largely due to the evolution of past economic activity. These results are consistent with the conclusions in the work (Neffke, Henning & Boschma, 2011), which shown that it is easier for regions to develop new industries if they are associated with already existing ones in the region [10].

5 Conclusion

A probabilistic interpretation of the elements of the matrix by which the economic complexity shall bes estimated from the work (Hausmann et al., 2011) is shown. Their properties are given, on the basis of which indicators are introduced that characterize the nesting of the structures of strong sectors of regional economies [6].

A model has been built that explains the appearance and absence of a strong sector in the structure of the region's economy. On this basis, the condition for the appearance of a certain strong sector in the structure of the economy of a particular region with a probability

exceeding 0.5 shall be obtained. This condition used to form a list of sectors recommended for priority development in the region. On the basis of ordering the sectors by the value of these probabilities and assessing their potential contribution to socio-economic development, an expert assessment of the feasibility of developing a new strong sector in the region can be given.

References

1. M. Yu. Afanasiev, A.V. Kudrov, Montenegrin Journal of Economics **16(4)**, 43 (2020)
2. K. J. Arrow, The Review of Economic Studies **29(3)**, 155 (1962)
3. U. Blien, K. Wolf, Labour Economics **13(4)**, 445 (2006)
4. M. Fuchs, Empirical Economics **40(1)**, 177 (2011)
5. K. Glodberg, T. Roeder, D. Gupta, C. Perkins, Information Retrieval **4**, 133 (2001)
6. R. Hausmann, C. Hidalgo, S. Bustos, M. Coscia, A. Simoes, M. A. Yildirim, The Atlas of Economic Complexity. *Mapping Paths to Prosperity*. Cambridge, MA: Center for International Development, Harvard University, Harvard Kennedy School, Macro Connections, Massachusetts Institute of Technology (2011)
7. R. Hausmann, B. Klinger, SSRN Electronic Journal (2006)
8. A. Illy, M. Schwartz, C. Hornych, M. Rosenfeld, Journal of Economic and Social Geography **102(5)**, 582 (2011)
9. A. Marshall, *Principles of Economics* (London, UK: MacMillan, 1890)
10. F. Neffke, M. Henning, R. Boschma, Economic Geography. **87(3),** 237 (2011)
11. P. M. Romer, Journal of Political Economy **94(5**), 1002 (1986)