*Sustainable Social Development (2023) Volume 1 Issue 2*
doi: 10.54517/ssd.v1i2.2225

ORIGINAL RESEARCH ARTICLE

# Economic complexity of regions and its interrelation with indicators of socio-economic development

**Mikhail Yu. Afanasiev*, Aleksei A. Gusev**

*Central Economics and Mathematics, Institute of the Russian Academy of Sciences (CEMI RAS), 117418 Moscow, Russia*

*** Corresponding author:** Mikhail Yu. Afanasiev, mi.afan@yandex.ru

**ABSTRACT**

The aim of the research is to compare estimates of the economic complexity of the regions obtained on the basis of different data. An approach to assessing the economic complexity of Russian regions by types of economic activity (TEA) is proposed. The approach is based on the standard method of assessing economic complexity. The question of choosing the RCA threshold value is considered. 0-1 matrices are constructed for sectors and TEA at different thresholds. Their structures correctly reflect the idea embedded in the economic complexity index. As a result of the correlation analysis, it is shown that at threshold 1, the index of economic complexity by sector and the index of economic complexity by TEA have greater resistance to changes in the threshold than at other threshold values. A comparative analysis of economic complexity indices constructed for 79 regions by 82 sectors and 24 TEA on the data of 2019 was carried out. Their significant statistical relationship with a number of indicators of socio-economic development characterizing the quality of life has been established. The results of this research can be used to help with building situational models of the economic development of regions as well as to coordinate decisions made by regions when choosing priority areas of their development related to increasing diversification.

***Keywords:*** diversification; regional economy; strong sector; economic complexity; quality of life

***JEL Classification:*** G10; G15; G40

## 1. Introduction

An urgent task is to create a methodology that allows for determining priority areas of diversification in the regional economy, taking into account its economic complexity. Innovation activity and socio-economic indicators that characterize the quality of life. In accordance with the approach proposed in the works of Afanasiev and Ilyin[1] and Afanasiev et al.[2] priority areas of diversification are selected based on recommendations for the development of sectors. This approach creates opportunities for the development of the theory of diversification based on the construction and analysis of the structures of strong sectors of the regional economy. The theory of diversification and empirical estimates are presented in the works of Fuchs[3] and Illy et al.[4]. Diversification, defined as the expansion of the structure of the economy, is an important goal in all countries and has become one of the most important priorities of economic development. There is no universal solution to promote economic development and structural changes in the regions. It is necessary to

**ARTICLE INFO**
Received: 3 July 2023 | Accepted: 16 August 2023 | Available online: 7 September 2023

**CITATION**
Afanasiev MY, Gusev AA. Economic complexity of regions and its interrelation with indicators of socio-economic development. *Sustainable Social Development* 2023; 1(2): 2225. doi: 10.54517/ssd.v1i2.2225

take into account the peculiarities of each region when developing and designing industrial and regional economic policies. Therefore, it is important to consider how to expand the process of economic diversification and ensure the creation and development of new sectors and industries, using the advantages of existing knowledge and competencies. One of the modern ways of reflecting diversification is the concept of economic complexity. There are several approaches to assessing economic complexity. Relatively recently, Hausmann and Rodrik[5], Hausmann et al.[6], Hidalgo and Hausmann[7] have developed a standard approach to measure economic complexity using data on export patterns by country. In the work of Lyubimov et al.[8], this concept was applied to data on the structure of exports by regions of Russia. Later, Afanasiev and Kudrov[9] applied economic complexity to data on tax revenues in 82 sectors of the economy for Russian regions.

The paper[10] proposed an approach to the formation of recommendations for the development of sectors in order to diversify the regional economy, focused on increasing its economic complexity. Estimates of the economic complexity of the regions were obtained in this work on the basis of data on tax revenues by economic sectors, which allows us to characterize the structures of regional economies, including sectors focused on both external and domestic markets. However, alternative approaches can be used to assess the economic complexity of regions.

In the modification of the standard approach proposed below, industrial production indicators are used to assess the economic complexity of regions: the volume of shipped goods of their own production, performed works, and services on their own in four enlarged types of economic activity, as well as the structure of the volume of shipped products for each enlarged type of economic activity. As a result, estimates of the volume of shipped products are formed for 24 types of economic activity (TEA). Data on 24 TEA is important for understanding the economic state of the regions of Russia, as it is information about industrial production from the official Rosstat source.

On this basis, a matrix of identified comparative advantages was constructed, describing the structure of regional economies according to TEA, and estimates of the economic complexity of the regions were obtained in accordance with the standard approach. One of the topical issues of the application of the economic complexity index is considered, which is associated with the choice of the RCA threshold value at the stages of formation and the comparative analysis of indices based on various data.

The aim of the research is to compare estimates of the economic complexity of the regions, obtained on the basis of data on 24 TEA, with previously obtained estimates of economic complexity based on data on 82 sectors of the economy. The high level of correlation of these estimates indicates their stability in relation to the data used and to the level of detail in describing the structure of the regional economy. Recommendations on the choice of the RCA threshold value for assessing the relationship between economic complexity indices and indicators of socio-economic development are substantiated.

## 2. Methodology and data used

### 2.1. Description of the structure of the regional economy

To describe the structure of the regional economy, data on the volumes of shipped products for 24 TEA were used. First, we will determine the indicator $RCA_{cp}$ of the identified comparative advantages:

$$RCA_{cp} = (y_{cp}/\sum_{p} y_{cp})/(\sum_{c} y_{cp}/\sum_{cp} y_{cp}) \tag{1}$$

where $y_{cp}$ is the volume of production by TEA $p$ of the region $c$; $RCA_{cp}$—the ratio of the share of production by TEA $p$ in the total volume of production by all TEA of the region $c$ to the share of production by TEA $p$ of all regions in the volume of production by all types of economic activity of all regions. According to the work of Hausmann and Klinger[11], to identify comparative advantages in economies, an indicator $RCA_{cp}$ is used for which a condition of the type of restriction from below is checked. Namely, if the value $RCA_{cp}$ exceeds one, then it is considered that the economy of the region $c$ has identified comparative advantages in the production of TEA $p$; otherwise, there are no identified comparative advantages:

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

The matrix $A = (a_{c,p})$ contains data on TEA, which in different regions are developed at the level of the identified comparative advantages determined using the Equation (1). The rows of this matrix correspond to regions, the columns correspond to TEA. The vector $(a_{c,p_1}, \cdots, a_{c,pm})$ will be called the structure of strong TEA of the economy of region $c$.

## 2.2. Economic complexity

The concept of "economic complexity of a region" is considered as a characteristic reflecting the level of its technological development, which is determined by strong TEA in the structure of its economy. Similarly, the economic complexity of TEA depends on the level of technological development of those regions in the structure of which this TEA is present as a strong one. Economic complexity is a latent characteristic of a region $ECI_c$ or TEA $ECI_p$. Estimates of economic complexity have the following properties: the economic complexity of a region is proportional to the average level of economic complexity of strong TEA in the structure of its economy:

$$ECI_c = a_1 \sum\nolimits_p r_{c,p} ECI_p \tag{2}$$

where $r_{c,p} = \frac{a_{c,p}}{k_{c,0}}$, $k_{c,0} = \sum_p a_{c,p}$ , $a_1$ is a positive constant.

The economic complexity of TEA is proportional to the average level of economic complexity of the regions in the structure of the economies of which this TEA is strong:

$$ECI_p = a_2 \sum\nolimits_c r^*_{p,c} ECI_c \tag{3}$$

where $r^*_{p,c} = \frac{a_{c,p}}{k_{p,0}}$, $k_{p,0} = \sum_c a_{c,p}$, $a_2$ is a positive constant. An indicator $k_{c,0}$ equal to the number of strong TEA in the region $c$, we will call the diversification of the structure of the economy of region $c$ by TEA. Let $c = (ECI_{c_1}, ECI_{c_2}, \cdots)^T$ be a vector-column of values of economic complexity for regions; $p = (ECI_{p_1}, ECI_{p_2}, \cdots)^T$—vector-column of values of economic complexity for TEA; $R_1 = (r_{c,p})$, $R_2 = (r^*_{p,c})$—matrices of weights. From the Equations (2) and (3), it follows that $c = a_1 a_2 R_1 R_2 c$, $p = a_1 a_2 R_2 R_1 p$. Thus, the economic complexity of the region is defined as the eigenvector of the matrix $R_1 R_2$, and the economic complexity of TEA is the eigenvector of the matrix $R_2 R_1$. The matrices $R_1 R_2$ and $R_2 R_1$ are stochastic: their elements are non-negative, and their row sum is 1. Due to stochasticity, the matrix $R_1 R_2$ has an eigenvalue equal to 1 and an eigenvector corresponding to it, which consists of identical coordinates. In the works of Hausmann and Rodrik[5] and Sciarra et al.[12] , it is proposed to use the value of the eigenvector of matrices $R_1 R_2$ corresponding to the second maximum eigenvalue as the values of estimates of the economic complexity. Note that estimates of the economic complexity of the structures of regional economies, sectors and TEA are relative. They do not depend on the scale of the regional economy. They can take both positive and negative real values[9].

### 2.3. Description of the data

In the proposed modification of the assessment of the economic complexity of the regions by TEA, the indicators of industrial production for four enlarged types of economic activity, as well as the structure of the volumes of shipped products for 24 TEA, are used. As a result, according to the Federal State Statistics Service, the volumes of shipped goods, completed works and services for 2019 for 24 TEA were obtained. Speaking about this in more detail, using the volumes of products shipped, works performed and services of industrial production for the four main enlarged types of economic activity $V_c^{(r)}$ by region for 2019, as well as the structures of distribution of these volumes as a percentage $x_{cp}^{(r)}$ for each types of economic activity within the corresponding main enlarged type of economic activity in the $r$ = {1, 2, 3, 4} “mining industry”, “processing industries”, “provision of electric energy, gas and steam; air conditioning” and “water supply; wastewater disposal, waste collection and disposal, pollution elimination activities”, it is possible to estimate the volume of shipped products according to TEA for all regions. Multiplying the shares of TEA in the volume of the corresponding enlarged type of economic activity by the total volume of shipped products for this type gives the values of the volumes of shipped products for the specified year for each TEA in all regions:

$$y_{cp} = \frac{x_{cp}^{(r)}}{100} \cdot V_c^{(r)}$$

where $c$—region; $p$—type of economic activity. As a result, estimates of the volume of shipped products for 24 TEA for 79 regions of Russia for 2019 were obtained.

## 3. Results

### 3.1. Economic complexity by types of economic activity

To assess the economic complexity by TEA, a standard approach of Hartmann et al.[13], Hausmann and Rodrik[5], Hausmann et al.[6], Hidalgo and Hausmann[7] was applied to the data of shipped industrial products for 2019. As a result, estimates of the economic complexity of the regions for 24 TEA were obtained (column (3) of **Table A1** of the appendix) and the economic complexity of TEA (column (3) of **Table A2** of the appendix). **Figure 1** shows a nonlinear dependence of the economic complexity of regions on the number of strong TEA (column (2) of **Table A1** of the appendix). Regions with a small number of strong TEA have relatively low estimates of economic complexity. With the increase in the number of strong TEA, there is a tendency to increase the economic complexity of the regions.

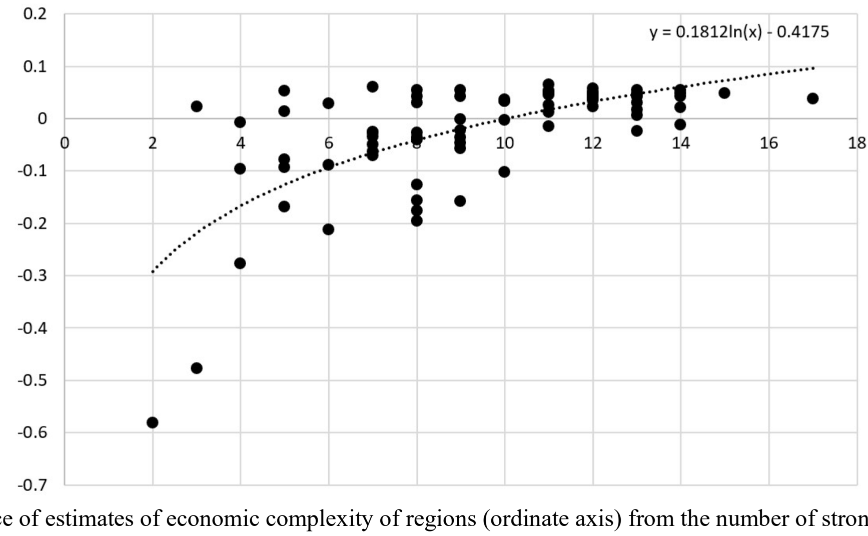



**Figure 1.** Dependence of estimates of economic complexity of regions (ordinate axis) from the number of strong TEA (abscissa axis).

Based on the results of calculations of estimates of the economic complexity of the regions (**Table A1** of the appendix, column (3)), the highest values are in the Tambov region (11 strong TEA; economic complexity of the region 0.06441), Kaluga region (7; 0.06103), Ulyanovsk region (12; 0.05662), the Republic of Mari El (9; 0.05488), the Republic of Mordovia (8; 0.05437), Penza region (13; 0.05418), the Chuvash Republic (14; 0.054), etc. These regions are characterized by a high diversification of the economy by types of economic activity. Taking into account the types of specialization of regional economies described in the work of Aivazian et al.[14], these regions can be classified as a mixed type. They specialize in manufacturing and agriculture. Relatively low estimates of economic complexity exist in the Tyumen region (2; −0.58096), Sakhalin region (3; −0.47772), Astrakhan region (4; −0.2764), Orenburg region (6; −0.21304), Republic of Sakha (Yakutia) (8; −0.19579), and Republic of Komi (8; −0.17566). These are mainly regions with a specialization in the mining industry. Thus, the developed "processing" and "agricultural regions" have relatively high estimates of economic complexity, while the "mining" regions have low estimates.

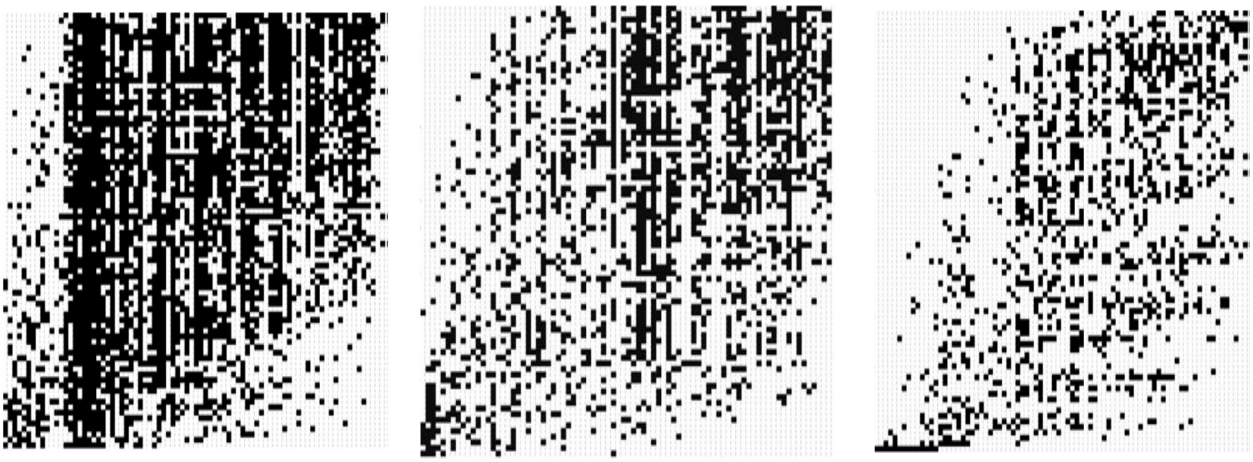
If we arrange the estimates of the economic complexity of types of economic activity in ascending order (**Table A1** of the appendix, column (3)), then first comes the TEA from the enlarged type of economic activity "mining". Including "provision of services in the field of mining" D5, which is a strong type of economic activity in 13 regions and has an economic complexity of −0.643959; "oil and natural gas production" D2 (15; −0.617883); "coal mining" D1 (14; −0.294573); "metal ore mining" D3 (22; −0.165385); and "extraction of other minerals" D4 (25; −0.073529). At the end with the highest estimates of economic complexity are manufacturing industries: "manufacture of furniture; manufacture of other finished products" OP11 (31; 0.08514); "manufacture of other non-metallic mineral products" OP7 (38; 0.08907); "manufacture of computers, electronic, and optical products; manufacture of electrical equipment" OP9 (35; 0.10366); "food production; beverage production; tobacco production products" OP1 (41; 0.10943); "manufacture of textiles; manufacture of clothing; manufacture of leather and leather products" OP2 (33; 0.12856). Thus, mining is among the least economically complex types of economic activity, and manufacturing is among the most economically complex.

# 4. Discussion

## 4.1. To the question of choosing the RCA threshold

Estimates of the economic complexity of regions by economic sectors obtained in the work of Afanasiev and Gusev[10] are presented in column (5) of **Table A1** applications. The economic complexity index (ECI) measures the complexity of a region's production structure by combining information about the diversity of the economy (the number of strong sectors, or TEA) and the prevalence or ubiquity of strong sectors (the number of regions in which the sector, or FEA, is strong, that is, produces products at the level of the identified comparative advantages). The idea behind ECI is that developed regional economies are diverse (diversified) and produce products of strong sectors, or TEAs, which on average have a low prevalence because only a few regional economies have developed a sector or TEA to the level of a strong one. The characteristics of the complexity of the production structures of the regions can be considered an indicator of the level of human and social capital of the regional economy, since the ability of the region to produce products from strong sectors with high complexity ratings depends on the accumulated knowledge and the ability of people to form social and professional networks in order to collect, accumulate, and use new knowledge in production[15]. One of the topical issues of practical application of the economic complexity index is related to the choice of the RCA threshold value at the stages of formation and comparative analysis of indices based on various data.

**Figure 2** shows 0-1 matrices describing the structure of regional economies constructed for different values of the RCA threshold. The rows of the matrices correspond to regions, and the columns correspond to sectors of the economy. The dark cell of the matrix means that the corresponding element of the matrix is equal to 1. That is, the sector is strong in the region's economy. Otherwise, the matrix element is zero, and the sector is not strong. The rows of each matrix are ordered from bottom to top in ascending ECI estimates of the economic complexity of the regions. The columns are ordered from left to right in an ascending order of estimates of the economic complexity of the sectors. In applied research, the threshold value of 1 is used by default, as it allows for a simple interpretation. In the context of regions and sectors, with the RCA threshold equal to 1, the production of a strong sector in the region's economy is higher than the share of this sector in the national economy.



RCA threshold value = 0.5 **(a)** RCA threshold value = 1 **(b)** RCA threshold value = 1.5 **(c)**

**Figure 2.** Matrix 0-1 region-sector, **(a)** RCA threshold value = 0.5; **(b)** RCA threshold value = 1; **(c)** RCA threshold value = 1.5.

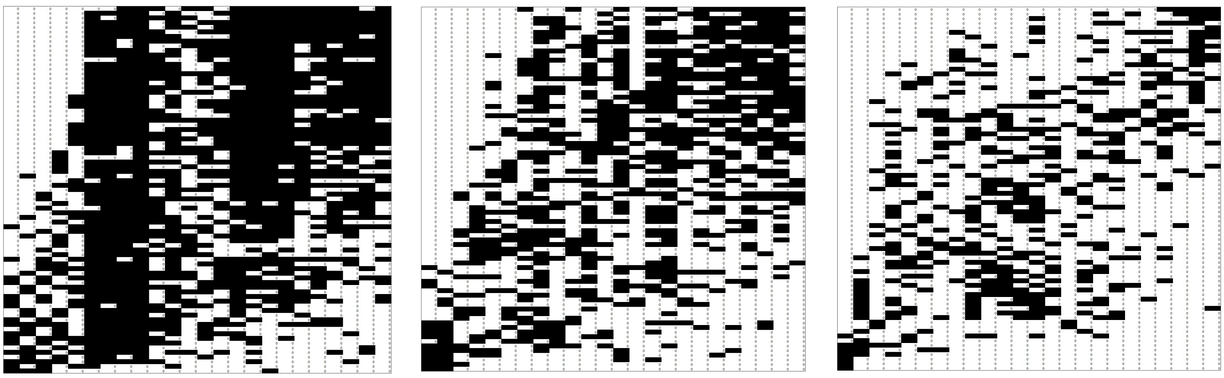
Let's consider the matrix structure at the RCA threshold equal to 1 (in **Figure 2b**). The upper rows of the matrix contain significantly more units than the lower rows. Accordingly, regions with higher estimates of economic complexity are more diversified than regions with lower estimates. Sectors with relatively high complexity scores are strong, mainly in regions with relatively high ECI economic complexity scores. Accordingly, the lower right corner of the matrix is poorly filled with units. Sectors with relatively low complexity scores are strong in regions with relatively low ECI economic complexity scores. Accordingly, the upper left corner of the matrix is poorly filled with units. In the matrix constructed for the RCA threshold equal to 0.5 (in **Figure 2a**), there are more elements equal to 1. It should be noted that, at a threshold of 0.5, some sectors are strong in almost all regions. This reduces the informativeness of the matrix. But this matrix has approximately the same structure as the matrix with an RCA threshold of 1. The matrix constructed for the RCA threshold equal to 1.5 (in **Figure 2c**) has relatively few elements equal to 1. Therefore, it may also not be informative enough to calculate the economic complexity index. However, the structure of each of the three matrices constructed for close RCA thresholds correctly reflects the idea embedded in the economic complexity index. Therefore, there is no serious reason to use a threshold other than 1 for calculating RCA. However, it is advisable to check that the ECI economic complexity index calculated at threshold 1 is stable. That is, it does not change much with a relatively small change in the RCA threshold. To do this, it is advisable to consider the correlation matrix of economic complexity indices constructed for different threshold values. Such a matrix is presented in **Table 1**.

**Table 1** reflects the correlation relationship of six indices of the economic complexity of regions, constructed for threshold values in the range of 0.5 to 1.5. Under the main diagonal of the table are the Pearson correlation coefficients. Above the main diagonal are Spearman's rank correlation coefficients. The use of smaller or larger thresholds is not advisable, since with such thresholds it is difficult to interpret the concept of a "strong sector" and the corresponding index of economic complexity. In addition, the evaluation of the relationship of such an index with the index constructed for threshold 1 becomes insignificant, and the possibility of their comparison is lost. There is high stability in the index constructed for threshold 1 when the threshold value changes in the range from 0.75 to 1.125. The stability of the index constructed for other values is lower. These conclusions do not contradict the recommendations to use in applied research an index of economic complexity constructed for the RCA threshold equal to 1.

**Table 1.** ECI by sector correlation for different RCA thresholds.

| RCA thresholds | 0.5 | 0.75 | 1 | 1.125 | 1.25 | 1.5 |
|---|---|---|---|---|---|---|
| 0.5 | 1 | 0.9136 | 0.8360 | 0.8114 | 0.7299 | 0.5659 |
| 0.75 | 0.8824 | 1 | 0.9361 | 0.9050 | 0.8070 | 0.5755 |
| 1 | 0.7776 | 0.9565 | 1 | 0.9508 | 0.8479 | 0.5919 |
| 1.125 | 0.7610 | 0.9362 | 0.9699 | 1 | 0.9506 | 0.7198 |
| 1.25 | 0.6854 | 0.8377 | 0.8632 | 0.9434 | 1 | 0.8418 |
| 1.5 | 0.1678 | 0.1919 | 0.2199 | 0.5810 | 0.5810 | 1 |

**Figure 3** shows 0-1 matrices describing the structure of regional economies based on TEA for different values of the RCA threshold. The dark cell of the matrix means that the corresponding element of the matrix is equal to 1. That is, the products of this TEA are produced by the region at the level of the identified comparative advantages. The rows of each matrix are ordered from bottom to top in ascending ECI estimates of the economic complexity of the regions. The columns are ordered from left to right in ascending order of estimates of the economic complexity of TEA. The structure of the matrices in **Figure 3** reflects the same features that we observe in **Figure 2**. Therefore, for a comparative analysis, we can use the RCA threshold 1, making sure that the corresponding index of economic complexity is stable.



RCA threshold value = 0.5 RCA threshold value = 1 RCA threshold value = 1.5

**Figure 3.** Matrix 0-1 region-TEA.

**Table 2** below the main diagonal shows the Pearson correlation coefficients of the indices of economic complexity of regions constructed according to TEA. Above the main diagonal are Spearman's correlation coefficients. The stability of the index based on TEA at a threshold of 0.75 is slightly higher than at a threshold of 1. But the high correlation coefficients of Pearson 0.95 and Spearman 0.94 for these indices allow us to use any of them. As can be seen from **Table 3**, the Pearson correlation coefficient of the index by sector at a threshold value of 1 is higher with the index by TEA at a threshold value of 1 than with the index by TEA at a threshold value of 0.75. The difference in Spearman correlation indices is insignificant. For these reasons, it makes sense to use the economic complexity index by sector with a threshold of 1 and the economic complexity index by TEA with a threshold of 1 in the applications.

**Table 2.** ECI by TEA correlation for different RCA thresholds.

| RCA thresholds | 0.5 | 0.75 | 1 | 1.125 | 1.25 | 1.5 |
|---|---|---|---|---|---|---|
| 0.5 | 1 | 0.9157 | 0.8957 | 0.7361 | 0.7224 | 0.2669 |
| 0.75 | 0.9815 | 1 | 0.9400 | 0.7967 | 0.8027 | 0.3246 |
| 1 | 0.9385 | 0.9539 | 1 | 0.8222 | 0.8254 | 0.3021 |
| 1.125 | 0.8357 | 0.8513 | 0.8977 | 1 | 0.7560 | 0.2165 |
| 1.25 | 0.8362 | 0.8596 | 0.9351 | 0.8704 | 1 | 0.4746 |
| 1.5 | 0.0689 | 0.0499 | 0.0842 | 0.0147 | 0.1215 | 1 |

**Table 3.** Pearson and Spearman correlation of ECI by sectors with ECI by TEA.

| ECI | ECI by TEA with a threshold 0.75 | ECI by TEA with a threshold 1 |
|---|---|---|
| ECI by sectors with a threshold 1. Pearson correlation | 0.696 | 0.771 |
| ECI by sectors with a threshold 1. Spearman correlation | 0.769 | 0.758 |

## 4.2. Comparative analysis of estimates of economic complexity

Let's compare the estimates of the economic complexity of 79 regions obtained with RCA thresholds equal to 1, based on data on tax revenues for 82 sectors (column (5) of **Table A1** of the appendix) and based on data on shipped products for 24 TEA for 2019 (column (3) of this table). The Pearson correlation coefficient of the index of economic complexity of regions for 82 sectors and the index of economic complexity for 24 TEA is 0.771. Spearman's rank correlation coefficient is 0.758.

There is also a relative similarity of estimates for the most economically complex TEA and sectors. This is the manufacturing industry of OP2, which corresponds to sector 1125 "production of leather and leather products". TEA and sectors with low estimates of economic complexity relate to mining. For types of economic activity it is D5 "provision of services in the field of mining" and for sectors, it is 1060 ("production of natural gas and gas condensate"). When visualizing in **Figure 4** the dependence of estimates of the economic complexity of regions on the number of strong sectors (column (4) of **Table A1** of the appendix), one can observe the same trends that were mentioned earlier in **Figure 1**.

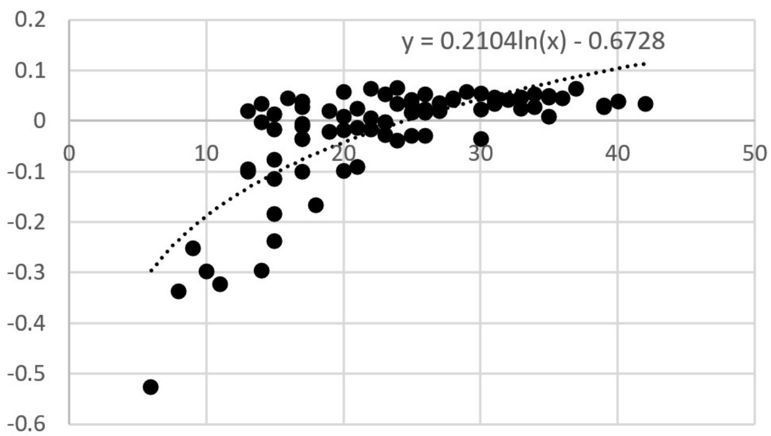



**Figure 4.** Dependence of estimates of economic complexity of regions (ordinate axis) from the number of strong sectors (the abscissa axis).

In **Figure 5**, the dot characterizes the region. Its coordinate along the abscissa axis is an assessment of the economic complexity of regions by sector. The coordinate on the ordinate axis is an assessment of the economic complexity of regions by TEA. In the lower part of the figure, there are points corresponding to three mining regions (Orenburg, Tyumen, and Sakhalin regions). These regions have the largest production volumes in the 1055 sector (crude oil production). In the upper right corner are regions with developed sectors of manufacturing and agricultural products. Regions located close to the diagonal have similar estimates of economic complexity by sector and TEA.

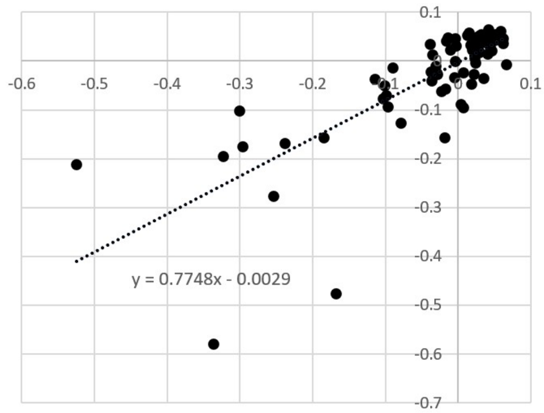



**Figure 5.** Estimates of economic complexity of regions by sectors (abscissa axis) and by TEA (ordinate axis) for 79 regions.

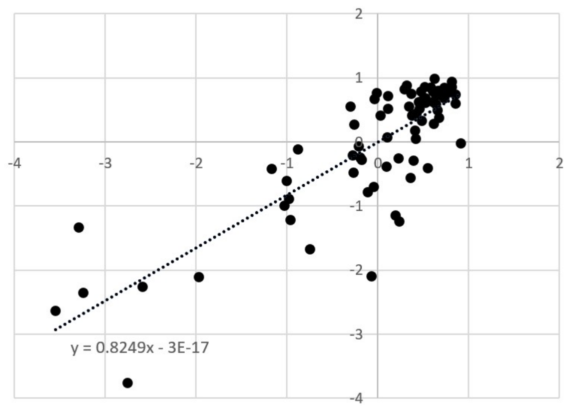



**Figure 6.** Estimates of the economic complexity of regions by sectors (abscissa axis) and by TEA (ordinate axis) without three mining regions

Of particular interest is the consideration of the situation without the three mining regions mentioned above (**Figure 6**). In the absence of three mining regions (Orenburg, Tyumen, and Sakhalin regions), the correlation of estimates increases to 0.825. This suggests that, in some cases, estimates of the economic complexity of regions by 24 TEA can be used. Estimates of the economic complexity of regions have high stability in the transition from data on tax revenues to data on production volumes and from data on sectors to data on TEA.

## 4.3. The relationship of economic complexity indices and indicators of socio-economic development

In the work of Afanasiev[16], to assess the relationship between economic complexity and material well-being, the regions are divided into two groups. The first group includes regions with relatively low estimates of economic complexity, according to data on economic sectors. The structures of the economies of most regions in the first group include strong sectors of the mining industry. The regions of the first group are marked with a sign (*) in column (1) of **Table A1** of the appendix. In the work of Aivazian et al.[17], these regions are classified as clusters of "mining". But there are also regions with a high level of diversification in this group. The second group includes all the other regions. They have relatively high estimates of economic complexity. For each of the two groups, the interrelationship of the ECIs economic complexity index, built by sectors and indicators of socio-economic development of regions for 2019, was verified. For comparison, the interrelationship between the ECIc economic complexity index for TEA and indicators of socio-economic development was also checked. To do this, we used Rosstat data on 36 indicators of socio-economic development given in the work of Aivazian et al.[18] and used in these works when constructing indices of the main directions of socio-economic development to assess the quality of life. The index of economic complexity by sector and the index of economic complexity by types of economic activity are constructed at six different thresholds for calculating RCA. Accordingly, the correlation coefficients of 12 different indices of economic complexity with each of the 36 indicators of socio-economic development were calculated. It is shown that for each of the two groups of regions, there is a significant statistical relationship between the indices of economic complexity by sector and by types of economic activity and indicators of socio-economic development: GRP per capita; average per capita income; unemployment rate; agricultural products per capita; morbidity from neoplasms; diseases of the digestive system; natural population growth rate.

As shown in **Table 5**, for the regions of group 1, the index of economic complexity, constructed by sectors at threshold 1, is statistically significantly correlated with three indicators: average per capita income, unemployment rate, and the coefficient of natural population growth. Thus, the growth of the economic complexity of regions with developed sectors of the extractive industry is accompanied by trends in the growth of average per capita income, a decrease in the unemployment rate, and the coefficient of natural population growth. The correlation level of −0.628, achievable at a threshold of 0.5, allows us to use the economic complexity index to build a model for predicting the unemployment rate of the regions in this group. A significant correlation with the ECIc economic complexity index, estimated by TEA, is observed only for the natural population growth rate. Moreover, the use of a threshold value other than 1 does not significantly increase the level of interrelation between the ECIc index and the four indices of socio-economic development of group 1 regions indicated in **Table 5**. Thus, to assess the economic complexity of regions with a developed extractive industry, the ECIs index is more preferable than the ECIc index.

**Table 5.** Correlation of indicators with ECIs and ECIc indices for 25 regions of group 1; correlation coefficients are significant at the 5% level.

| Indicator | Correlation with ECIs at threshold 1 | Maximum modulo correlation with ECIs | The threshold at which the maximum correlation is achieved | Correlation with ECIc at threshold 1 | Maximum modulo correlation with ECIc | The threshold at which the maximum correlation is achieved |
|---|---|---|---|---|---|---|
| Per capita income | 0.341 | 0.341 | 1 | Insignificant | 0.335 | 1.75 |
| Unemployment rate | −0.357 | −0.628 | 0.5 | Insignificant | −0.461 | 1.25 |
| Natural population growth rate | −0.325 | −0.434 | 1.25 | −0.447 | −0.471 | 0.5 |
| Diseases of the digestive system | Insignificant | −0.337 | 0.25 | Insignificant | −0.454 | 0.5 |

For the regions of group 2, the ECIs economic complexity index, constructed by sectors at threshold 1, is statistically significantly correlated with six indicators: average per capita income, unemployment rate, natural population growth rate, GRP per capita, agricultural products per capita, and morbidity from neoplasms (**Table 6**). Thus, the growth of the economic complexity of regions with relatively high estimates of complexity according to the ECIs index is accompanied by trends in the growth of average per capita income, GRP per capita, agricultural products per capita, and morbidity from neoplasms. As well as trends in reducing the unemployment rate and the natural growth rate of the population. The level of correlation achievable at the threshold of 0.25 allows using the economic complexity index (ECI) to build a model for predicting the unemployment rate of the regions in group 2. The ECIs index, at a threshold of 0.75, can be used to build a model for predicting natural population growth. A significant correlation of the ECIc economic complexity index, estimated by types of economic activity, is observed for the unemployment rate, the coefficient of natural population growth, and agricultural products per capita. Moreover, changing the threshold value does not significantly increase the level of interrelation of the ECIc index with the indicators of socio-economic development of the regions of group 2 given in **Table 6**. Thus, to assess the economic complexity of the regions of the second group, the ECIs index is also more preferable than the ECIc index.

**Table 6.** Correlation of indicators with ECIs and ECIc indices for 54 regions of group 2; correlation coefficients are significant at the 5% level.

| Indicator | Correlation with ECIs at threshold 1 | Maximum modulo correlation with ECIs | The threshold at which the maximum correlation is achieved | Correlation with ECIc at threshold 1 | Maximum modulo correlation with ECIc | The threshold at which the maximum correlation is achieved |
|---|---|---|---|---|---|---|
| Per capita income | 0.297 | 0.351 | 0.5 | Insignificant | 0.331 | 0.5 |
| Unemployment rate | −0.559 | −0.745 | 0.25 | −0.431 | −0.551 | 0.5 |
| Natural population growth rate | −0.522 | −0.781 | 0.75 | −0.451 | −0.569 | 0.5 |
| GRP per capita | 0.297 | 0.397 | 0.5 | Insignificant | Insignificant | |
| Agricultural products per capita | 0.469 | 0.481 | 0.5 | 0.352 | 0.351 | 1 |
| Morbidity from neoplasms | 0.311 | 0.492 | 0.25 | Insignificant | 0.333 | 0.25 |

# 5. Conclusion

An approach to assessing the economic complexity of regions based on data on production volumes for 24 TEA is presented. The calculation of economic complexity estimates is based on the standard approach of Hartmann et al.[13], Hausmann and Rodrik[5], Hausmann et al.[6], Hidalgo and Hausmann[7]. The question of choosing the RCA threshold value is considered. The structure of 0-1 matrices for sectors and types of economic activity, constructed for close RCA thresholds, correctly reflects the idea embedded in the economic complexity index. It is shown that at threshold 1, the index of economic complexity by sector and the index of economic complexity by TEA have greater resistance to changes in the threshold than at other threshold values. A comparative analysis of the estimates of economic complexity of 79 regions for 24 TEA and the estimates of economic complexity presented in the work of Afanasiev and Gusev[10] for 82 sectors based on data for 2019 was carried out. Their correlation was 0.771; without three mining regions (Orenburg, Tyumen, and Sakhalin regions), the correlation of estimates of economic complexity based on data on tax revenues by sector and the volume of shipped products for TEA increases to 0.825. Thus, estimates of economic complexity remain highly stable during the transition from data on tax revenues to data on production volumes and from data on sectors to data on TEA. Assessment of the economic complexity of the regions according to 24 TEA can be useful in solving management tasks aimed at increasing the economic complexity of the region. It is shown that the economic complexity index, based on sector data at the RCA threshold of 1 for a group of regions with strong sectors of mining industries and relatively low estimates of economic complexity, is statistically significantly correlated with three indicators: average per capita income, unemployment rate, and natural population growth rate. And for a group of regions with relatively high estimates of economic complexity with six indicators: average per capita income, unemployment rate, natural population growth rate, GRP per capita, agricultural products per capita, and morbidity from neoplasms, At the same time, to assess the economic complexity of the regions of each group, the index of economic complexity by sector is preferable to the index for types of economic activity.

Estimates of the economic complexity of the structures of regional economies, sectors, and types of economic activity can be used to develop a methodology for selecting priority areas for the diversification of the region's economy, as described in the work of Afanasiev and Ilyin[1]. This methodology allows us to justify the choice of a sector for development to the level of a strong one in the region using a number of criteria based on estimates of the economic complexity of the structures of strong sectors in the regions. The number of criteria can be expanded by using estimates of the economic complexity of the structures of strong economic activities, considering them as characteristics of the level of human capital development that affect the level of material well-being of the population. The application of the methodology for choosing priority areas of diversification using digital technologies in regional situation centers can ensure coordination of decisions taken by regions when choosing priority areas of diversification.

## Author contributions

Conceptualization, MYA; methodology, MYA; software, AAG; validation, MYA and AAG; formal analysis, MYA and AAG; investigation, MYA and AAG; resources, MYA and AAG; data curation, AAG; writing—original draft preparation, MYA and AAG; writing—review and editing, MYA and AAG; visualization, MYA; supervision, MYA and AAG; project administration, MYA; funding acquisition, MYA. All authors have read and agreed to the published version of the manuscript.

## Conflict of interest

The authors declare no conflict of interest.

## References

1. Afanasiev MY, Ilyin N. New guidelines for choosing priority areas of economic diversification based on the system of situational centers (Russian). *Ekonomika I Matematicheskie Metody* 2022; 58(4): 29–44. doi: 10.31857/S042473880023017-7
2. Afanasiev MY, Kudrov AV, Lysenkova MA. The approach to the diversification of the regional economy taking into account evolutionary conditionality, resource availability and innovation activity. *Montenegrin Journal of Economics* 2022; 18(2): 7–17. doi: 10.14254/1800-5845/2022.18-2.1
3. Fuchs M. The determinants of local employment dynamics in Western Germany. *Empirical Economics* 2011; 40(1): 177–203. doi: 10.1007/s00181-010-0420-5
4. Illy A, Schwartz M, Hornych C, Rosenfeld MTW. Local economic structure and sectoral employment growth in German cities. *Journal of Economic and Social Geography* 2011; 102(5): 582–593. doi: 10.1111/j.1467-9663.2011.00659.x
5. Hausmann R, Rodrik D. Economic development as self-discovery. *Journal of Development Economics* 2003; 72(2): 603–633. doi: 10.1016/S0304-3878(03)00124-X
6. Hausmann R, Hwang J, Rodrik D. What you export matters. *Journal of Economic Growth* 2007; 12(1): 1–25. doi: 10.1007/s10887-006-9009-4
7. Hidalgo CA, Hausmann R. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences* 2009; 106(26): 10570–10575. doi: 10.1073/pnas.0900943106
8. Lubimov IL, Gvosdeva MA, Kazakova MV, Nesterova KV. Complexity of the economy and possibility of export diversification in Russian regions (Russian). *Zhurnal Novoy Ekonomicheskoy Assotsiatsii* 2017; 34(2): 94–122. doi: 10.31737/2221-2264-2017-34-2-4

9. Afanasiev MY, Kudrov AV. Economic complexity and embedding of regional economies' structures (Russian). *Ekonomika I Matematicheskie Metody* 2021; 57(3): 67–78. doi: 10.31857/S042473880016410-0
10. Afanasiev MY, Gusev AA. Approximation of estimates of economic complexity when choosing priority areas of diversification (Russian). *Tsifrovaya Ekonomika* 2022; 17(1): 52–59. doi: 10.34706/DE-2022-01-05
11. Hausmann R, Klinger B. Structural transformation and patterns of comparative advantage in the product space. Available online: http://www.tinyurl.com/2bdm7aml (accessed on 5 September 2023).
12. Sciarra C, Chiarotti G, Ridolfi L, Laio F. Reconciling contrasting views on economic complexity. *Nature Communications* 2020; 11(1): 3352. doi: 10.1038/s41467-020-16992-1
13. Hartmann D, Guevara MR, Jara-Figueroa C, et al. Linking economic complexity, institutions, and income inequality. *World Development* 2017; 93: 75–93. doi: 10.1016/j.worlddev.2016.12.020
14. Aivazian SA, Afanasiev MY, Kudrov AV. Method of clustering of regions of the Russian Federation taking into account the sectoral structure of GRP (Russian). *Prikladnaya Ekonometrika* 2016; 41(1): 24–46.
15. Hidalgo C. Why information grows: The evolution of order, from atoms to economies. *The Review of Austrian Economics* 2015; 30: 147–151. doi: 10.7551/mitpress/9647.001.0001
16. Afanasiev MY. New guidelines for the digital economy: On the relationship between economic complexity and material well-being (Russian). *Bulletin of the CEMI RAS* 2022; 5(1): 1–10. doi: 10.33276/S265838870019868-7
17. Aivazian SA, Afanasiev MY, Kudrov AV. Indicators of economic development in the basis of characteristics of regional differentiation (Russian). *Prikladnaya Ekonometrika* 2018; 50(2): 4–22.
18. Aivazian SA, Afanasiev MY, Kudrov AV. Methodology of socio-economic development assessment given the characteristics of regional differentiation. *Model Assisted Statistics and Applications* 2020; 15(4): 311–314. doi: 10.3233/MAS-200502

# Appendix

**Table A1.** Estimates of the economic complexity of regions by TEA and sectors.

| (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|
| Belgorod region | 4 | −0.00795 | 24 | 0.0670 |
| Bryansk region | 11 | 0.04989 | 31 | 0.0402 |
| Vladimir region | 13 | 0.04643 | 37 | 0.0617 |
| Voronezh region | 12 | 0.05004 | 34 | 0.0285 |
| Ivanovo region | 12 | 0.03532 | 28 | 0.0430 |
| Kaluga region | 7 | 0.06103 | 29 | 0.0584 |
| Kostroma region | 11 | 0.0253 | 33 | 0.0246 |
| Kursk region | 10 | 0.0366 | 22 | 0.0618 |
| Lipetsk region | 6 | 0.02831 | 36 | 0.0443 |
| Moscow region* | 17 | 0.03826 | 39 | 0.0266 |
| Oryol region | 12 | 0.04852 | 30 | 0.0561 |
| Ryazan region | 12 | 0.04535 | 16 | 0.0435 |
| Smolensk region | 11 | 0.0508 | 31 | 0.0455 |
| Tambov region | 11 | 0.06441 | 28 | 0.0419 |
| Tver region | 13 | 0.04177 | 42 | 0.0335 |
| Tula region | 9 | 0.04189 | 34 | 0.0514 |
| Yaroslavl region | 14 | 0.04206 | 25 | 0.0318 |
| Moscow* | 10 | 0.03332 | 24 | −0.0386 |
| Republic of Karelia | 8 | −0.0277 | 26 | 0.0215 |
| Komi Republic* | 8 | −0.17566 | 14 | −0.2957 |
| Arkhangelsk region* | 8 | −0.15681 | 20 | −0.0186 |
| Volgograd region | 5 | 0.01381 | 25 | 0.0413 |
| Kaliningrad region | 5 | 0.05247 | 15 | 0.0126 |
| Leningrad region | 12 | 0.03752 | 14 | 0.0332 |
| Murmansk region* | 9 | −0.02193 | 17 | −0.0367 |
| Novgorod region | 12 | 0.03768 | 32 | 0.0404 |
| Pskov region | 13 | 0.04384 | 35 | 0.0503 |
| Saint petersburg* | 13 | 0.04498 | 23 | −0.0031 |
| Republic of Adygea | 8 | 0.04157 | 22 | −0.0160 |
| Krasnodar territory | 9 | −0.03585 | 27 | 0.0354 |
| Astrakhan region* | 4 | −0.2764 | 9 | −0.2532 |
| Volgograd region | 8 | 0.03034 | 17 | 0.0274 |
| Rostov region | 14 | 0.02063 | 33 | 0.0461 |
| Republic of dagestan | 11 | 0.04681 | 19 | 0.0198 |
| Republic of Ingushetia | 9 | −0.05716 | 15 | −0.0168 |
| Kabardino-Balkarian Republic | 11 | 0.05357 | 17 | 0.0381 |
| Karachay-Cherkess Republic | 12 | 0.02282 | 27 | 0.0202 |
| Republic of North Ossetia-Alania | 13 | 0.03022 | 14 | −0.0029 |
| Chechen Republic | 9 | −0.04686 | 13 | 0.0186 |

**Table A1.** (*Continued*).

| (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|
| Stavropol territory | 12 | 0.04797 | 23 | 0.0527 |
| Republic of Bashkortostan* | 7 | −0.04997 | 17 | −0.1003 |
| Republic of Mari El | 9 | 0.05488 | 31 | 0.0325 |
| Republic of Mordovia | 8 | 0.05437 | 20 | 0.0582 |
| Republic of Tatarstan* | 5 | −0.07789 | 13 | −0.1025 |
| Udmurt Republic* | 8 | −0.12676 | 15 | −0.0780 |
| Chuvash Republic | 14 | 0.054 | 40 | 0.0383 |
| Perm territory* | 7 | −0.07007 | 20 | −0.0983 |
| Kirov region | 14 | 0.0477 | 35 | 0.0473 |
| Nizhny novgorod region | 12 | 0.04005 | 24 | 0.0330 |
| Orenburg region* | 6 | −0.21304 | 6 | −0.5249 |
| Penza region | 13 | 0.05418 | 26 | 0.0513 |
| Samara region* | 8 | −0.0367 | 15 | −0.1144 |
| Saratov region | 15 | 0.04786 | 21 | −0.0135 |
| Ulyanovsk region | 12 | 0.05662 | 25 | 0.0152 |
| Kurgan region | 10 | 0.03317 | 26 | 0.0175 |
| Sverdlovsk region | 13 | 0.00622 | 30 | 0.0229 |
| Tyumen region* | 2 | −0.58096 | 8 | −0.3363 |
| Chelyabinsk region | 7 | −0.02482 | 35 | 0.0071 |
| Republic Altai | 11 | 0.01246 | 30 | −0.0352 |
| Republic of Tyva | 7 | −0.03444 | 25 | −0.0047 |
| Republic of Khakassia | 6 | −0.08919 | 17 | 0.0044 |
| Altai territory | 11 | 0.04551 | 22 | 0.0473 |
| Krasnoyarsk territory* | 5 | −0.16893 | 33 | −0.2383 |
| Irkutsk region* | 9 | −0.15761 | 19 | −0.1845 |
| Kemerovo region | 4 | −0.09598 | 15 | 0.0080 |
| Novosibirsk region | 13 | 0.0172 | 15 | 0.0295 |
| Omsk region | 3 | 0.02255 | 20 | −0.0104 |
| Tomsk region* | 10 | −0.10242 | 39 | −0.3000 |
| Republic of Buryatia | 13 | −0.0243 | 17 | −0.0295 |
| Republic of Sakha (Yakutia)* | 8 | −0.19579 | 10 | −0.3225 |
| Trans-Baikal territory | 7 | −0.06277 | 11 | −0.0221 |
| Kamchatka krai* | 9 | −0.00123 | 23 | −0.0039 |
| Primorsky territory* | 14 | −0.01156 | 26 | −0.0307 |
| Khabarovsk territory* | 11 | −0.01456 | 21 | −0.0896 |
| Amur region* | 8 | −0.0408 | 17 | −0.0358 |
| Magadan region* | 7 | −0.02697 | 23 | −0.0278 |
| Sakhalin region* | 3 | −0.47772 | 18 | −0.1673 |
| Jewish autonomous region | 10 | −0.00344 | 21 | 0.0240 |
| Chukotka autonomous okrug* | 5 | −0.09384 | 13 | −0.0964 |

Table structure by columns: (1) name of the region; (2) the number of strong TEA in the region; (3) estimates of the economic complexity of the regions by TEA; (4) the number of strong sectors in the region (from the work of Afanasiev and Gusev[10]); (5) estimates of the economic complexity of regions by sector (from the work of Afanasiev and Gusev[10]).

**Table A2.** Estimates of the economic complexity of TEA.

| (1) | (2) | (3) |
| --- | --- | --- |
| D5 | 13 | −0.643959 |
| D2 | 15 | −0.617883 |
| D1 | 14 | −0.294573 |
| D3 | 22 | −0.165385 |
| D4 | 25 | −0.073529 |
| OP8 | 21 | −0.067646 |
| OP12 | 33 | −0.053073 |
| OE3 | 52 | −0.019888 |
| OP3 | 31 | −0.013296 |
| B4 | 30 | −0.001626 |
| OE1 | 43 | 0.01941 |
| OP4 | 27 | 0.02714 |
| OE2 | 48 | 0.02902 |
| OP5 | 14 | 0.02962 |
| B2 | 43 | 0.03215 |
| B1 | 52 | 0.03792 |
| OP6 | 24 | 0.06375 |
| OP10 | 30 | 0.07254 |
| B3 | 37 | 0.08353 |
| OP11 | 31 | 0.08514 |
| OP7 | 38 | 0.08907 |
| OP9 | 35 | 0.10366 |
| OP1 | 41 | 0.10943 |
| OP2 | 33 | 0.12856 |

(1) The name of the TEA (according to the increasing economic complexity); (2) the number of regions in which TEA is strong; (3) assessment of the complexity of TEA.