Economic complexity and inclusion of regional economies

Mikhail Afanasiev[*], and *Aleksander* Kudrov

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

Abstract. The paper presents a probabilistic interpretation of the elements of the matrix, which is used to assess the economic complexity in accordance with the traditional approach. Their properties are given, on the basis of which aggregate indicators are introduced that characterize the nesting of the structures of strong sectors of regional economies. It is shown that aggregate nesting indicators are statistically significant explanatory variables for economic complexity. It is proved that the used procedure for calculating the economic complexity is correct in the sense of the existence and uniqueness of the solution. It is shown that the data that are used to assess the economic complexity in accordance with the author's approach allow to reflect the formation of value chains and groups of related sectors focused on both the external and internal markets. For this economic complexity, calculated on regional data, its high values correspond to large values of the aggregated nesting indicators. Low values of economic complexity correspond to low values of nesting indicators.

1 Introduction

Modern understanding of economic complexity is associated with the diversification of production. Countries exporting "complex" goods tend to have higher levels of per capita wealth than countries exporting "simple" goods. A relatively recent procedure has been developed that allows to measure the economic complexity of both sectors and the structure of the economy as a whole. Further, in contrast to the traditional approach, for which the concept of economic complexity is applied in relation to countries and the products they export, the emphasis is on the study of the sectors of the economy of the regions of the Russian Federation.

2 Methods

To describe the structure of strong sectors of the economy, regional data on production volumes in a fairly wide range of sectors are used. At first we will define an indicator RCA_{cp} of the revealed comparative advantages:

---

[*] Corresponding author: mi.afan@yandex.ru

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

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 work (Hausmann, Klinger, 2006), the indicator RCA_{cp} is employed to identify comparative advantage in economies, for which the restriction type condition from below is checked [4]. Namely, if value RCA_{cp} is above one, then it is considered that the economy of the region c has an identified comparative advantage in the output of the sector p. Otherwise, it is considered that the revealed comparative advantages do not exist. More formally:

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

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.

2.1 Economic complexity

The concept of the economic complexity of a region is viewed 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 level of technological development of those regions in which this sector is present in the structure as a strong one.

Definition (*economic complexity*).

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

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

where a_1 - a positive constant;

Let us note that $k_{c,0}$ may not be equal to zero, since for any c there is p, for which $a_{c,p} = 1$

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

$$ECI_p = a_2 \sum_c r_{p,c}^* ECI_c, \text{ where } r_{p,c}^* = \frac{a_{c,p}}{k_{p,0}}, k_{p,0} = \sum_c a_{c,p}, \tag{2}$$

where a_2- a positive constant. Further, we will call the indicator $k_{c,0}$ equal to number of strong sectors in the region c, *diversification of structure of economy* of the region c.

Let us employ some additional designations:

$c = \left(ECI_{c_1}, ECI_{c_2}, \dots\right)^{\mathrm{T}}$- a column vector of economic complexity values for regions; $p = \left(ECI_{p_1}, ECI_{p_2}, \dots\right)^{\mathrm{T}}$- a column vector of values of economic complexity for sectors; $R_1 = (r_{c,p})$ $R_2 = (r_{p,c}^*)$- matrixes of weights.

Let us write down ratios (2) in a matrix look: $c = a_1 R_1 p$, $p = a_2 R_2 c$. From where iit 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 own vector of a matrix, and economic complexity of the sector - own vector of a matrix. $R_1 R_2 R_2 R_1$

Let us notice that an element on crossing of the i-th row and j-th column of the matrix $R_1 R_2$, i.e. $(R_1 R_2)_{ij}$, is given by the formula:

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

The works (Hausmann, Rodrik, 2003), (Sciarra, Chiarotti et al., 2020) employ the standardized eigenvector value of matrices $R_1 R_2$ and $R_2 R_1$, respectively, which correspond to the second maximum eigenvalue, as the values of estimates of the economic complexity of regions and sectors. Herewith, it is supposed that own value and coordinates of own vector are real numbers [5, 9].

It should be noted that authors of this article did not find any research demonstrating the validity property of the coordinates of such vectors. Below we will correct this defect and provide the proof together with other properties of a matrix $R_1 R_2$.

2.2 Properties of matrix $R_1 R_2$ and their interpretation

Let us list some properties of a matrix which will allow to consider the procedure of calculation of economic complexity correct:
1. The matrix $R_1 R_2$ is stochastic.

The stochasticity property of a matrix $R_1 R_2$ is fair as elements of a matrix are not negative, and their sum on lines is equal to 1. Indeed, it is:

$$\sum_j (R_1 R_2)_{ij} = \sum_j \frac{1}{k_{c_i,0}} \sum_t \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}} = \sum_t \sum_j \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{c_i,0} k_{p_t,0}} = 1$$

Owing to stochasticity of a matrix $R_1 R_2$, it has own value equal to unit and own vector corresponding to it, which consists of identical coordinates.
2. All own values of a matrix $R_1 R_2$ except for single on the module it is less than one.

According to Perron–Frobenius theorem, the following statements are fair:

Let P be the stochastic matrix, all elements of which are positive. Then:
a.　　　P has exactly one eigenvector x_0, wherein all coordinates are positive. If λ_0- own value of this vector, then $\lambda_0 = 1$;
b.　　　$\lambda_0 = 1$ is a simple root of the characteristic equationв $\det(P - \lambda I) = 0$.
c.　　　All own values (possibly, the complex ones) of matrix P satisfy the inequality $|\lambda| \leq \lambda_0$. Equality is reached only if all coordinates of the corresponding own vector are positive.

The evidence is provided in (Seneta, 2006).
3. All own values of a matrix $R_1 R_2$ are valid. As
$$R_1 R_2 = diag\left(k_{c_1,0}^{-1}, \dots, k_{c_n,0}^{-1}\right) \cdot A \cdot diag\left(k_{p_1,0}^{-1}, \dots, k_{p_m,0}^{-1}\right) \cdot A^T$$

we have:

$$\det(R_1 R_2 - \lambda I) = \det\left(diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) \right.$$

$$\cdot \left[diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) \cdot A \cdot diag(k_{p_1,0}^{-1}, ..., k_{p_m,0}^{-1}) \right.$$

$$\left. \cdot A^T diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) - \lambda I \right] \cdot diag\left(k_{c_1,0}^{\frac{1}{2}}, ..., k_{c_n,0}^{\frac{1}{2}} \right) \right)$$

Further, using the fact that the determinant of the product of two square matrices is equal to the product of the determinants of these matrices, we obtain

$$\det(R_1 R_2 - \lambda I) = \det\left(diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) \cdot A \cdot diag(k_{p_1,0}^{-1}, ..., k_{p_m,0}^{-1}) \cdot A^T \cdot \right.$$

$$\left. diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) - \lambda I \right).$$

Matrix

$$B = diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) \cdot A \cdot diag(k_{p_1,0}^{-1}, ..., k_{p_m,0}^{-1}) \cdot A^T \cdot diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right)$$

is symmetric. It is known that for symmetric matrixes, all own values are valid.
4. The eigenvectors of the matrix $R_1 R_2$ corresponding to different eigenvalues are orthogonal.

Let λ_1, λ_2 two non-matching eigenvalues of the matrix $R_1 R_2$, and v_1, v_2 – their corresponding eigenvectors. Eigenvalues, eigenvectors matrix B and $R_1 R_2$ coincide. Let us consider the quadratic form $v_1^T B v_2$, for which:

$$v_1^T B v_2 = \lambda_2 v_1^T v_2 = (B^T v_1)^T v_2 = (B v_1)^T v_2 = \lambda_1 v_1^T v_2$$

From where, since $\lambda_1 \neq \lambda_2$, we get that $v_1^T v_2 = 0$.
5. All eigenvalues of the matrix $R_1 R_2$ are non-negative.

Let us take vector $y \neq 0$ of dimensions n and consider the following quadratic form:

$$y^T B y = y^T diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) \cdot A \cdot diag(k_{p_1,0}^{-1}, ..., k_{p_m,0}^{-1}) \cdot A^T \cdot diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) y$$

$$= \left[diag(k_{p_1,0}^{-1/2}, ..., k_{p_m,0}^{-1/2}) A^T diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) y \right]^T$$

$$\cdot \left[diag(k_{p_1,0}^{-1/2}, ..., k_{p_m,0}^{-1/2}) A^T diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) y \right]$$

Having made the replacement:

$$z = diag(k_{p_1,0}^{-1/2}, ..., k_{p_m,0}^{-1/2}) A^T diag\left(k_{c_1,0}^{-\frac{1}{2}}, ..., k_{c_n,0}^{-\frac{1}{2}} \right) y$$

we receive that:

$$y^T B y = z^T z \geq 0.$$

From the fact that the eigennumbers of the matrix B are real, and the quadratic form $y^T B y$ for any $y \neq 0$ is non-negative it follows that the eigennumbers of the matrix are B, and hence the eigennumbers $R_1 R_2$ are non-negative as well.

2.3 Aggregated inclusion

Let us designate

$$w_{i,j} = (R_1 R_2)_{ij} = \sum_{t=1}^{m} \frac{a_{c_i,p_t}}{k_{c_i,0}} \frac{a_{c_j,p_t}}{k_{p_t,0}} = \frac{1}{k_{c_i,0}} \sum_{t=1}^{m} \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}}.$$

Let us note some properties of $w_{i,j}$:
1. Since the matrix $R_1 R_2$ is stochastic, then for each $i, j \in \{1, \dots, n\}$:

$$w_{i,j} \geq 0$$

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

Therefore, $w_{i,1}, \dots, w_{i,n}$ may be interpreted as a probability distribution.
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. The value $w_{i,j}$ increases monotonically with increasing number of elements of the set:

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

6. 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 (2-3) 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 aggregated indicator

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

characterizes degree of inclusion of structure of strong sectors of the region c_i in structures of strong sectors of other regions. Let us call this indicator *the aggregated nesting of the structure* of economy of region. Minimum value I_i^1, which is equal to one, occurs either when the structure of the strong sectors of the region c_i consists of unique sectors for which $w_{i,j} = 0$ for all $i \neq j$. A value I_i^1 close to one occurs when the distribution $w_{i,j}$ differs little from the uniform distribution.

Indicator

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

characterizes the degree of inclusion in structure of difficult sectors of the region c_j. Let us call this indicator *the aggregated nesting in the structure* of region. The value $I_j^{(2)}$ does not decrease with an increase in the number of complex sectors in the region c_j. The minimum value I_j^2 equal to one arises if the structure of complex sectors for the region c_j consists of unique sectors, i.e. when $w_{i,j} = 0$ for all $i \neq j$.

3 Results

3.1 Relationship between economic complexity and indicators of aggregate inclusion

The work (Afanasyev, Kudrov, 2020) employed the data on tax revenues for 82 sectors of the economy for the regions of the Russian Federation. Table 1 presents the estimates of the parameters of the regression of the economic complexity of the RF regions, obtained in (Afanasyev, Kudrov, 2020), on the logarithm of the aggregate inclusion of the structure of strong sectors of the economy and the logarithm of the aggregate nesting in the structure [1].

Table 1. Regression for economic complexity of regions. Note: estimate data on economic complexity are taken from research (Afanasyev, Kudrov, 2020) [1].

	Assessment	Standard deviation	t-statistics	p-value
Constant	-4.18	0.45	-9.22	0
Log (Aggregated inclusion of structure)	0.96	0.19	5.14	0
Log (Aggregated inclusion in to the structure)	0.97	0.15	6.65	0
R-square	0.55			
Vel. R-square	0.54			

Estimates of economic complexity were obtained for Russian regions, for which other regional data were also used. Namely, the data on the export of the RF regions. In the work (Lyubimov, Gvozdeva, Kazakova, Nesterova, 2017), using these data, two estimates of the economic complexity of the regions of the Russian Federation were calculated: an estimate of the economic complexity obtained from the data on the export structure of the regions of the Russian Federation; an estimate of the economic complexity obtained from the data on the structure of exports of the regions of the Russian Federation and the countries of the world [8]. Table 2 shows the results of constructing a regression of these estimates of economic complexity on the logarithms of the indicators of the aggregated nesting of the structure of the strong sectors of the region and the aggregate nesting in the structure of the strong sectors of the region.

Table 2. Left - regression for the economic complexity of regions according to data with countries; on the right - regression for the economic complexity of regions according to data without countries.

economic complexity (by data with the countries)	Assess ment	Standa rd deviati on	t-statis tics	p-valu e	economic complexity (by data without the countries)	Assess ment	Standa rd deviati on	t-statis tics	p-valu e
Constant	0.15	0.25	0.61	0.54	Constant	-1.03	0.67	-1.55	0.13
Logarith m of the aggregate d enclosure of structure	-0.19	0.10	-1.87	0.06	Logarithm of the aggregated enclosure of structure	0.06	0.27	0.22	0.82
Logarith m of the aggregate d enclosure into the structure	0.31	0.08	3.92	0.00	Logarithm of the aggregated enclosure in to the structure)	0.42	0.21	1.95	0.06

R-square	0.17
Vel. R-square	0.15

R-square	0.05
Vel. R-square	0.03

The regression for economic complexity on data with countries corresponds to a significantly higher R-squared value than regression for economic complexity on data without countries.

4 Discussions

Large values of economic complexity from the work (Afanasyev, Kudrov, 2020) correspond to a high level of the index of nesting of the structure of the economy and a high level of the index of nesting in the structure [1]. Low-level economic complexity is characterized by correspondence to the low level of the index of nesting of the structure of the economy and the index of nesting in the structure. Thus, the indicator of economic complexity can be interpreted as a measure of the level of development of adjacent sectors and technological chains or value chains.

Two indicators of economic complexity from the work (Lyubimov, Gvozdeva, Kazakova, Nesterova, 2017) are characterized by low compliance with the nesting indicators [8]. The explanation seems to be that exported goods are embedded in international value chains. Therefore, if we restrict ourselves only to export data by region, then, due to the absence of long value chains for export goods at the local Russian level, the assessment of economic complexity does not reflect the links between Russian export goods. On the contrary, if we connect to the available data on the export structure of the regions of the Russian Federation information on the structure of exports of the countries of the world and estimate the economic complexity based on them, then such an assessment already reflects the technological chains of the formation of added value at the international level and allows us to quantitatively assess the place of the regions of the Russian Federation in them. As a result,

as can be seen from Table 2 (left), for this indicator there is a significant relationship with the indicator of nesting in the structure.

5 Conclusion

It has been proved that the procedure for calculating economic complexity used in (Hausmann, Hidalgo, Bustos, Coscia, Simoes, Yildirim, 2011) is correct in the sense of existence and uniqueness of the solution. Properties of a matrix on which the economic complexity is estimated allow to interpret any (i, j) −element of this matrix as characteristic of degree of enclosure of structure of strong sectors -th region in structure of j-th region. Based on the properties of the elements of this matrix, aggregate indicators are introduced that characterize the nesting of the structure of the strong sectors of each region in the structures of the economies of all other regions [3].

For the economic complexity from the work (Afanasyev, Kudrov, 2020), data on tax revenues by sectors of the economy are used, which allow reflecting the formation of value chains and groups of related sectors focused on both the external and internal markets [1]. For this economic complexity, calculated on regional data, its high values correspond to large values of the aggregated nesting indicators; low values of economic complexity correspond to low values of nesting indicators.

References

1. M.Yu.Afanasiev, A.V. Kudrov, Montenegrin Journal of Economics **16 (4)**, 43 (2020)

2. D. Hartmann, Linking Economic Complexity, Institutions, and Income Inequality. World Development **93**, 75 (2017)

3. R. Hausmann, C. Hidalgo, S. Bustos, M. Coscia, A. Simoes, M.A. Yildirim, The Atlas of Economic Complexity: Mapping Paths to Prosperity. Cambridge: Center for International Development, Harvard University, MIT, (2011)

4. R. Hausmann, B. Klinger Structural Transformation and Patterns of Comparative Advantage in the Product Space. CID Working Paper, 128 (2006)

5. R. Hausmann, D. Rodri, Journal of Development Economics **72(2)**, 603 (2003)

6. R. Hausmann, Journal of Economic Growth **12(1)**, 1 (2006)

7. C. A. Hidalgo, R. Hausmann The building blocks of economic complexity. Proceedings of the National Academy of Sciences **106(26),** 10570 (2009)

8. I.L. Lyubimov, M.A. Gvozdeva, M.V. Kazakova, K.V. Nesterova Economic Complexity of Russian Regions and their Potential to Diversify, NEA Journal, 2 (34), 94 (2017)

9. C. Sciarra, G.Chiarotti, L. Ridolfi, et al. Reconciling contrasting views on economic complexity. Nat Commun **11,** 3352 (2020)

10. E. Seneta, Non-negative Matrices and Markov Chains, Springer Science (2006)