DOI: https://doi.org/10.37075/EA.2022.1.03 Economic Alternatives, 2022, Issue 1, pp. 35-51

Regional Economic Development Divergence Research in Georgia by PCA Method

Received: 03.12.2021
Available online: 29.03.2022

Nikoloz Chikhladze*, Gocha Ugulava**

Abstract

The study of a mix of the negative effects of unequal economic development between regions is one of the challenges economic science is now facing. Numerous methods are provided in terms of assessing regional disparities and interpreting the assessment results. The article describes the methodology for studying economic divergence between regions using the method of principal component analysis. The introduction of an aggregate indicator describing divergence allows us to reflect on a variety of different data, to study and explore a complete, real, comprehensive, volumetric picture of divergence. It appears possible: to identify clusters of regions according to mathematical calculations, including distinguishing the growth poles, less developed and depressed regions of the middle zone; to determine the degree of disparity when classifying in terms of the phases of divergence: the cases of differentiation, asymmetry and polarization; to rank the regions according to their share in this divergence; to analyze how the level of divergence changes over time by observing the dynamics of indicators.

Following the description of this methodology, the existing problems of divergence of economic development between the regions of Georgia are analyzed.

Keywords: Regional Development Divergence, Principal Component Analysis Method, Economy of Georgia

JEL: C38, O18, R11

Introduction

For a number of objective economic, political and social reasons, economic space is characterized by a high degree of heterogeneity and imbalance. This causes a mix of the negative effects that create the stable preconditions for unequal economic development between regions. The existence of divergence in the process of creating a country's unified economic space is inevitable, however, excessive and exaggerated imbalances break the cohesion and unity of the economic space. International practice also confirms that rising divergence is correlated with reduced productive capacity and a slowdown in investment activity in

* Doctor of Economic Sciences, Professor, Akaki Tsereteli State University, Kutaisi University. Address: Tsereteli 13, Kutaisi, Georgia

** PhD student, Kutaisi University. Address: Tsereteli 13, Kutaisi, Georgia.

regions, weakening economic relations, and increasing social instability in society.

Under these circumstances, it is becoming particularly imperative to search for scientifically founded principles and approaches and effective mechanisms for overcoming spatial contrasts and imbalances in regional economic development. In order to develop and use the tools of divergence research, it is necessary to develop a methodology for its assessment.

Existing methods for assessing regional imbalances are characterized by considerable diversity in terms of approaches and interpretations of the assessment methods and results. However, the vast majority of them are geared to solving local tasks and serve to identify differences in certain parameters.

There are two approaches: analysis of indicators, comparison and rating estimations (by scores or intervals). The analysis of indicators is based on the analysis of the descriptive statistics of the test parameter (mean, minimum and maximum values, standard deviation, relative deviation from an average value, variance, etc.) or various indices or coefficients (Gini coefficient, asymmetry coefficient, Theil index, etc.) (Novotný, 2007; Wang, Y., et al., 2012; Panzera and Postiglione 2020).

The objective of studying divergence existing in the economic development of regions cannot be solved through the analysis of individual indicators, because the weak points and aspects evaluated by individual parameters cannot reflect the whole data set, and an overall, real and comprehensive picture of divergence. Their disadvantage is also that the individual indicators cannot represent a single aggregate parameter, the introduction of a single metric for different indicators, which makes it difficult to study divergence as a dynamic process, determine its stage of development and to fix relative differences.

Thus, the introduction of an aggregate indicator describing divergence is required to identify the factors affecting this process and their trends, in order to conduct a comprehensive study of how they evolve.

Methodology

The algorithm for calculating an aggregate indicator proposed by us for studying unequal economic development between regions involves the implementation of five successive phases.

In the first phase, the indicators are selected, which we will aggregate into a single indicator. In this phase, it is important to consider the following circumstances:

- The selected indicators should fully and adequately reflect various aspects of regional economic development and should also correspond to the goal and objectives of the research, should also be of a complex nature, that is, reflect both the trends and peculiarities of regional economic development;
- The indicators should be available, that is, they should be given an official status and should be compiled from official sources of statistics, ministries, regional and local territorial bodies.

For processing purposes, we distinguish two types of data: *absolute and relative* indicators. In order to expand the research area, it is necessary to take into account the size of the region, the number of population therein (it is also possible to conduct a similar study according to the land area). Thus, in parallel with the absolute indicators, we will also separately prepare and calculate the relative indicators per capita of region. All this will allow us to make an analysis of the region in terms of the absolute parameters, taking into consideration the relative specifics as well.

The second phase is feature scaling. The method of principal component analysis is sensitive to units of measurement of data, especially when dealing with variables with different dimensions (Grus, 2015). Feature scaling is used to equalize these disparities. In practice, it is often used in the form of normalization, min-max scaling, or standardization, the same is used as Z-normalization. In the case of normalization, all data are reduced to the range of [0, 1] (more rarely to [-1, 1] or other ranges). Acceptable values are calculated for each data value, minus minimum value of the whole row, and then by comparing with the difference between the maximum and minimum values (*minimax* scaling). A disadvantage of normalization is the accumulation of (normalized) values obtained in the presence of anomalous values in a narrow range. (Han, J., et al., 2011).

Thus, a standardizing method is more commonly used. In the case of standardization, each data value is subtracted from the sample mean and is correlated with the standard deviation, i.e. the data are transformed into a form with a standard normal distribution. As a result, we get data whose average is zero and the standard deviation is equal to one:

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x}$$

In practice, the scaling algorithms were developed in programming language, and they are given in the specialized Machine Learning (ML) Python library *scikit-learn*, from the *sklearn.preprocessing* package methods MinMaxScaler (), StandardScaler (), RobustScaler () or others. (Sklearn, 2011)

In the third phase, we implement the method of principal component analysis. Principal Component Analysis (PCA, Karhunen-Loeve Transformation, Hotelling's Transformation) is one of the main methods for reducing the spatial dimension of data characteristics by losing the lowest amount of useful information. (Jackson, 1991).

From a mathematical perspective, the method of principal component analysis is an orthogonal linear transformation that displays data from the original characteristic space into a new, smaller dimensional space. The first axis of the new coordinate system is constructed in such a way that the data variance is maximal. The second axis is drawn orthogonally to the first one so that the variance of the data is maximal from the remaining possibilities, and so on.

The first axis is called the first principal component, the second one - the second principal component, and so on. Thus, the significance of the method is that a certain part of the total variance of the initial data flow is associated with each principal component (it is called the loading). In turn, variance, which is a measure of data variability, reflects the extent of their informative value. The variability along some axes of the initial space of the features may be large, for some - small, and for some - it may even not exist at all. It is assumed that the smaller the data variance along the axis, the less important the contribution of the variable associated with this axis and, therefore, by excluding this axis from the space (variable – from a model), it is possible to reduce the data dimension without losing informative value. Accordingly, the purpose of the PCA method is to construct a new reduced dimension space of the characteristics, the variance between the axes of which is redistributed so as to maximize each of them. (UCLA, 2018)

Notwithstanding the fact that the method is very effective and is widely used in the data processing process, it also has some limitations that need to be considered in its implementation process (PCA, 2019).

In particular:

1. Components have no specific semantic load, they are the aggregated data and absorb the variance of different initial variables;
2. The method only works with respect to continuous data.

Consider a more rigorous mathematical description of this method *(Jolliffe, 2002).* Suppose there is given n-th numerical characteristic $f_j(x),\ j = \overline{1,n}$; the objects are identified by their properties with the descriptions $x_i \equiv \big(f_1(x_i), \cdots, f_n(x_i)\big),\ i = \overline{1,l}$

Let us consider the matrix *F:*

$$F_{l\times n} = \begin{pmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & & \vdots \\ f_1(x_l) & \cdots & f_n(x_l) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_l \end{pmatrix} \quad (2)$$

Let us introduce the notation $z_i = \big(g_1(x_i), \cdots, g_m(x_i)\big),\quad i = \overline{1,l}.$ The description of the same objects in a new lower-dimensional space, where $Z = \mathbb{R}^m$, $m < n$.

$$G_{l\times n} = \begin{pmatrix} g_1(x_1) & \cdots & g_m(x_1) \\ \vdots & & \vdots \\ g_1(x_l) & \cdots & g_m(x_l) \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_l \end{pmatrix} \quad (3)$$

Request here that the original descriptions be restored according to the new descriptions by means of any linear transformation defined by the matrix $U = (u_{js})_{n\times m}$:

$$\hat{f}_j(x) = \sum_{s=1}^{m} g_s(x) u_{js} \quad j = \overline{1,n} \quad x \in X \quad (4)$$

The problem, while minimizing the total errors of the restored descriptions, is to find the matrix (G) and at the same time the linear transformation matrix (U) of the new descriptions:

$$\Delta^2(G,U) = \sum_{i=1}^{l} \|\hat{x}_i - x_i\|^2 = \sum_{i=1}^{l} \|z_i U^T - x_i\|^2 = \|GU^T - F\|^2 \ \to min_{G,U} \quad (5)$$

Let us assume that the matrices are nonzero, $rank\ G = rank\ U = m$ and the norms are Euclidean.

To solve the formulated problem, the following theorem is used: if $m \le rank\ F$, then the minimum $\Delta^2(G,U)$ is reached when the eigenvectors $F^T F$ of the matrix U are its own columns corresponding to the m maximum eigenvalues. Here, $G = FU$, the matrices U, G are orthogonal. The eigenvectors that correspond to the maximum eigenvalues are the principal components.

The following algorithm of action is provided to use this method:

1. Calculating the total variance of the initial space of the characteristics, which is done by summing up the relative variances of the variables defined from the covariance matrix.

2. Calculating the eigenvectors and eigenvalues of the covariant matrix, which determine the direction of the principal components and the magnitude of the associated variance.
3. Performing dimension reduction. The diagonal elements of the covariant matrix represent the variance according to the initial coordinate system, while its eigenvalues represent the new one. By correlating the total variance with the variance associated with each principal component, we shall obtain the share of each component. We will leave the principal components whose total share will be 80-90%, or we will select the components through special criteria used in different practices (Statistics Handbook, 2021).

In our case, the algorithm for calculating the principal components through the singular decomposition of the data matrix was developed in the programming language and given in the specialized Machine Learning Python library *scikit-learn sklearn. decomposition.PCA* class, which will be used to perform calculations in the research (PCA, 2011).

In the fourth phase, we analyze the adopted lower-dimensional, aggregated indicators in the new coordinate system. For the analysis, we use the first principal component for each type of indicators (absolute, relative), which, with small losses of information, with acceptable accuracy, reflects a single, aggregate indicator of the economic situation in the region. Thus, if we give a geometric interpretation, for a two-dimensional space (for a plane), where the first principal components of the absolute values are counted on the abscissa axis, while the first principal components of the relative values are counted on the ordinate axis, each region is represented by a point with two coordinates. For further analysis, it becomes necessary to introduce a metric on the plane, which will allow us to count the distances and use the values obtained as a tool for analysis (Burago et al., 2004).

From the existing distance metrics (Euclid, Minkowskiy, Manhattan, Heming, etc.) we will focus on the Euclidean distance (Smith, 2013), which we will calculate by the following general formula:

$$dist_{euclidean} = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2} \tag{6}$$

Thus, these tools allow us to do data analysis:

- Additionally, using the clustering algorithms, it is easy to group regions by level of economic development (overall, absolute, relative indicators) (Lavrova, 2012);
- We can calculate both unified and clustering the so-called “average standard”, which will be given by the coordinates of centroid of values. (Baranov S., 2014) The centroid of the finite set of points minimizes the sum of the squares of Euclidean distances between it and these points, respectively for the point k of $\mathbb{R}^n$:

$$C = \frac{x_1 + x_2 + \cdots + x_k}{k} \tag{7}$$

- Calculation of the Euclidean distances in pairs for all points, which will give us a picture of the existing inequalities between the regions;
- Calculation of the distances from the centeroid (absolute and relative), which will allow us to redistribute the regions according to the phases of divergence.

We will use the existing classification of phases of developmental inequalities: differentiation - when there are slight deviations in economic development, asymmetry - if the deviations are substantial, and polarization - when the deviations are critical. According to the relative distance, which we will calculate by the following formula:

$$\delta = \frac{C - x_i}{C} \tag{8}$$

we will classify as follows: differentiation – up to 0,33; asymmetry – within 0,33 ÷ 0,67 interval and polarization – over 0,67 (Gubanova and Kleshch, 2018)

- Determining of the level of economic development impact of each region in the formation of divergence in overall economic development:

$$I_i = \frac{\frac{1}{n}\left(\frac{x_i - C}{C}\right)^2}{\sigma^2} \tag{9}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - C}{C}\right)^2 \tag{10}$$

- We will analyze these indicators in dynamics and will observe how the level of divergence changes over time, divergence of economic development between the regions converges or diverges over time.

In the final fifth phase, we visualize and interpret the data, and then, we draw conclusions on the research problem.

Let us use the presented methodology to analyze the divergence of economic development beween the regions of Georgia.

Data use and sources

To conduct research, let us identify and find the source data. We review the indicators according to the regional division given in the official economic statistics of Georgia (National Statistics Office of Georgia) - according to eleven regions (including Tbilisi). We matched the conditional notation code to these indicators (see Table 1).

Table 1. *Regions with reference to their code*

Code	Region
TBS	Tbilisi
ADJ	Adjara
GUR	Guria
IME	Imereti
KAH	Kakheti
MTS	Mtskheta-Mtianeti
RLQ	Racha-Lechkhumi and Qvemo Svaneti
SZS	Sameqrelo-Zemo Svaneti
SJV	Samtskhe-Javakheti
QQR	Kvemo Kartli
SHQ	Shida Kartli

We selected the data needed for the analysis according to the above criteria. Indicators for the aggregation are given for the period of 2010-2019, by regions for each year and type. We also matched the conditional notation code to all the indicators. (Geostat, Pc-Axis) (see Table 2).

Table 2. *Economic development indicators*

Code	Indicator (2010-2019)	Unit of measure
gdp	Regional gross domestic product	mln GEL
fdi	Foreign direct investment by regions	mln USD
lcs	Labor inputs by regions	mln GEL
emp	Number of employees in the business sector by regions	person
prd	The production of goods in the business sector by regions	mln GEL
gsp	Purchases of goods and services by regions	mln GEL
ind	Value added in industry by regions	mln GEL
trd	Value added volume of wholesale and retail, car and motorcycle repair enterprises by regions	mln GEL
acf	Value added of enterprises engaged in accommodation and food delivery activities by regions	mln GEL
cns	Value added in construction by regions	mln GEL
trs	Value added of enterprises engaged in transport and warehousing activities by regions	mln GEL

The data were processed by a program written in a Python programming language. The *NumPy* and *Pandas* libraries were used for data wrangling and calculations. Feature scaling was performed through the *sklearn.preprocessing.StandardScaler*. *Sklearn.decomposition.PCA* was used for the analysis of principal components. The results were visualized using the *Matplotlib* library. A software code in the form of *Jupiter Notebook* is available in the *GitHub* repository (see Python Developer's Guide; NumPy v1.20 Manual; Pandas documentation; Matplotlib Documentation; Scikit-learn 0.20.1 documentation; Pedregosa F., at al., (2011); Bressert (2012).

Description of the results

Based on the population data, we calculated relative rates per capita for each region by years and types. Using the principal component analysis, the first principal component was calculated for each region, per year for both absolute and relative data. We visualized the obtained indicators for the end of the year 2019 that is under review (see Figure 1, 2).

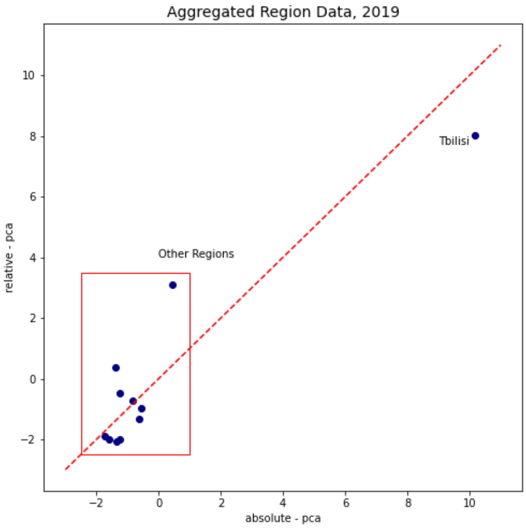


Figure 1. *Aggregated data for regions, 2019*

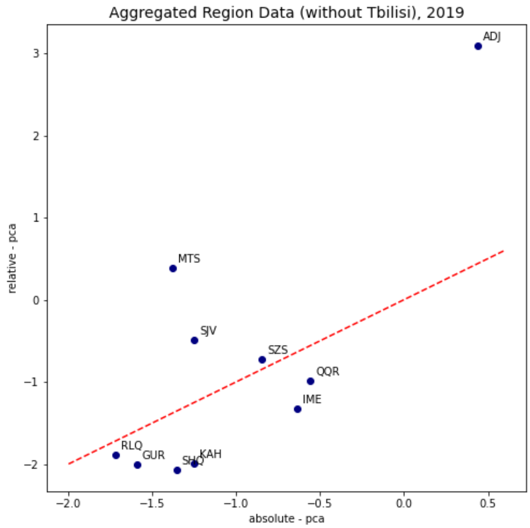


Figure 2. *Aggregated data for regions (excl. Tbilisi), 2019*

The first principal component for the absolute values was displayed as the abscissa coordinate, while the ordinate coordinate – displayed for the relative values. Even at first glance, there is a large disparity between Tbilisi and other regions. Thus, throughout Georgia, according to the state of the economy, we can make a distinction between Tbilisi and other regions. In addition to visual observations, the same is evidenced by the analysis of the metric between the regions in the space of principal components (see Table 3).

Table 3. *Distance (metric) between regions in the space of principal components*

Region	TBS	ADJ	GUR	IME	KAH	MTS	RLQ	SZS	SJV	QQR	SHQ
TBS	0	10.9	15.45	14.28	15.18	13.84	15.48	14.06	14.24	14	15.32
ADJ	10.9	0	5.49	4.55	5.35	3.27	5.43	4.03	3.96	4.2	5.47
GUR	15.45	5.49	0	1.17	0.34	2.39	0.17	1.48	1.55	1.45	0.25
IME	14.28	4.55	1.17	0	0.9	1.86	1.22	0.63	1.03	0.35	1.04
KAH	15.18	5.35	0.34	0.9	0	2.37	0.48	1.32	1.5	1.22	0.13
MTS	13.84	3.27	2.39	1.86	2.37	0	2.3	1.23	0.88	1.59	2.45
RLQ	15.48	5.43	0.17	1.22	0.48	2.3	0	1.45	1.48	1.48	0.41
SZS	14.06	4.03	1.48	0.63	1.32	1.23	1.45	0	0.47	0.39	1.44
SJV	14.24	3.96	1.55	1.03	1.5	0.88	1.48	0.47	0	0.85	1.59
QQR	14	4.2	1.45	0.35	1.22	1.59	1.48	0.39	0.85	0	1.35
SHQ	15.32	5.47	0.25	1.04	0.13	2.45	0.41	1.44	1.59	1.35	0

The distance between any two regions (average - 2.8, minimum - 0.13, maximum - 5.49) is 4-10 times shorter than their distance from Tbilisi (average - 14.3, minimum - 10.9, maximum - 15.45).

For a detailed analysis of other regions, it is necessary to bring them together visually and review them on a scale excluding Tbilisi. In this case too, there is a certain disparity, which, even without the use of clustering algorithms, is visually evident. The graphical analysis is supported by the metric of the centroid of the regions in the space of principal components (see Table 4).

The centroid is, of course, a conditional concept, though it can be identified with the weighted, intermediate place of the country's unified economy, the proximity to which equates the economies of regions. Consequently, the extent of their distance from this hypothetical place demonstrates the degree of divergence.

Visually, based on the graphical analysis, the rest of the regions can be divided into three groups, which is also confirmed by their metrics:

- The so-called "central" group, which is closest to the centroid (the distance to the centroid is less than 2.0), is characterized by small migration in the last decade and did not exceed 2.0, that is, their growth rate is close to the overall average growth rate of the country, so they cannot be the growth leaders, and they also contribute to the divergence. These groups include the following regions - Samegrelo-Zemo Svaneti (last year - 1.12, in dynamics 0.95 - 1.12, minimum - 0.82, maximum - 1.18), Kvemo Kartli (last year - 1.13, in dynamics 0.27 - 1.13, minimum - 0.27, maximum - 1.13), Mtskheta -Mtianeti (last year - 1.43, in dynamics 1.53 - 1.43, minimum - 1.33, maximum - 1.57) and Imereti (last year - 1.25, in dynamics 1.25 - 1.47, minimum - 1.25, maximum - 1.79).
- The so-called "lower" group (the distance to the centroid is over 2.0 - in reverse), they have stagnated in the dynamics in the last decade, maintaining the role of outsiders - Kakheti (last year - 2.35, in dynamics - 2.35 - 2.35, minimum - 2.03, maximum - 2.35), Shida Kartli (last Year - 2.48, in dynamics 2.02 - 2.48, minimum - 1.65, maximum - 2.63), Racha-Lechkhumi-Kvemo Svaneti (last year - 2.56, in dynamics 1.25 - 2.56, minimum - 2.3, maximum - 2.83) and Guria (last year - 2.56, in dynamics - 2.58 - 2.56, minimum - 2.31, maximum - 2.61). It should be noted that the economies of regions from the "lower" group are characterized by a relatively small volume (both in absolute and relative indicators). They can be classified as less developed (Kakheti, Shida Kartli) and depressed (Racha-Lechkhumi-Lower Svaneti and Guria) regions.
- Adjara is explicitly included in the next group (last year - 3.12, in dynamics - 1.29 - 3.12, minimum - 1.29, maximum - 3.12). Adjara is "a growth pole", it is characterized by stable growth in the last decade, the growth rate is higher than the overall average growth rate. However, the rapid growth of Adjara and Tbilisi contributes to the overall divergence. It is interesting to observe the migration of indicators of the regions in the space of principal components in dynamics (except for Tbilisi).

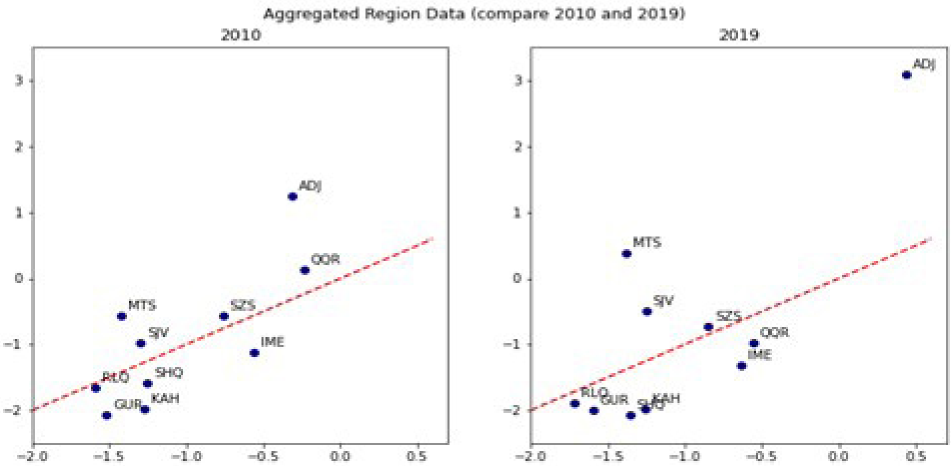


Figure 3. *Comparison of aggregated data by regions (excl. Tbilisi), 2010 and 2019*

Table 4. *Distance (metric) from the regions to the centroid in the space of the principal components*

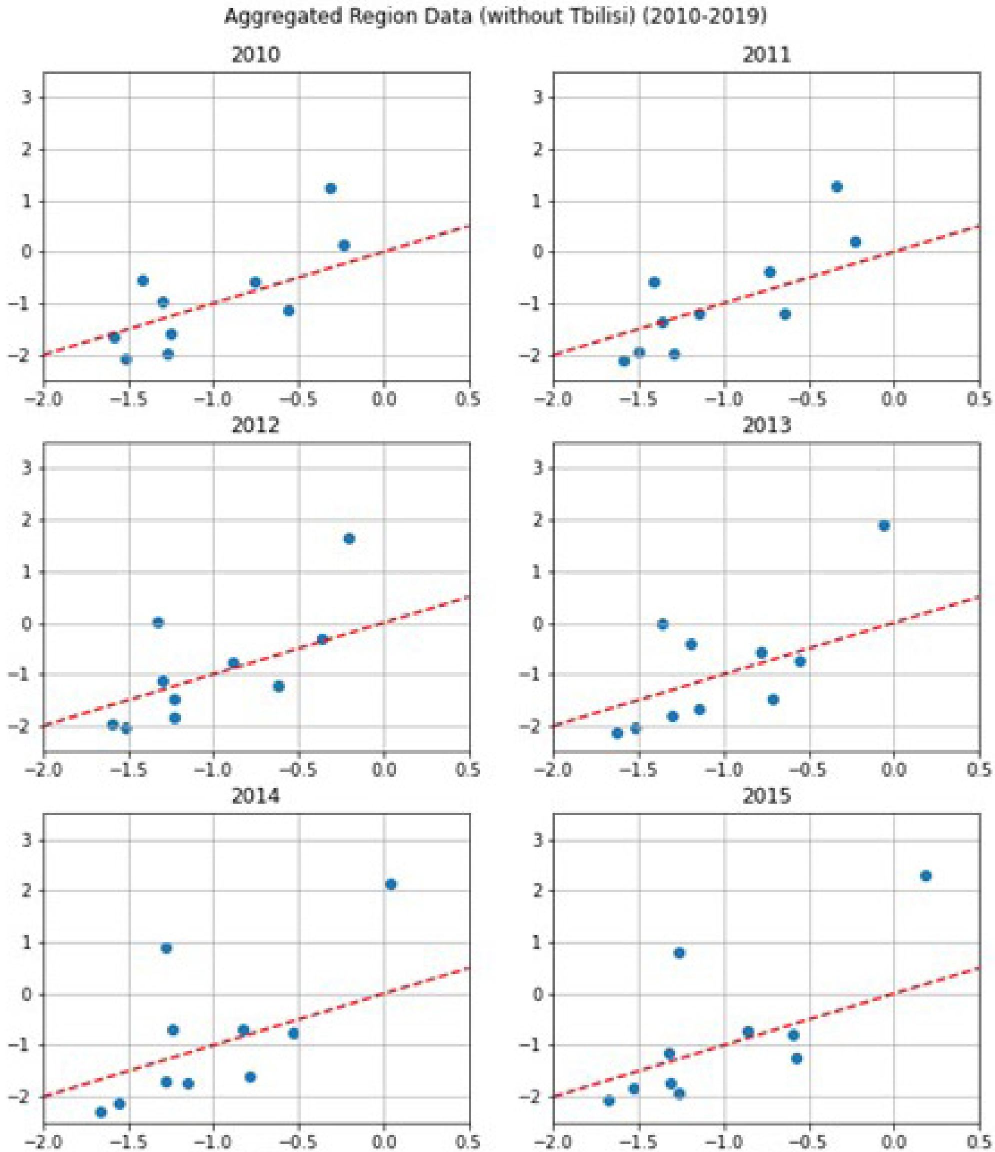
Region	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
TBS	13.71	13.77	13.71	13.6	13.38	13.21	13.48	13.19	13.33	12.95
ADJ	1.29	1.32	1.64	1.91	2.13	2.31	2.52	2.2	2.14	3.12
GUR	2.58	2.45	2.55	2.53	2.63	2.39	2.31	2.4	2.41	2.56
IME	1.25	1.35	1.36	1.63	1.79	1.38	1.64	1.31	1.38	1.47
KAH	2.35	2.35	2.21	2.03	2.08	2.31	2.24	2.18	2.34	2.35
MTS	1.53	1.52	1.33	1.36	1.57	1.51	1.34	1.41	1.51	1.43
RLQ	2.3	2.64	2.54	2.67	2.83	2.67	2.54	2.58	2.42	2.56
SZS	0.95	0.82	1.18	0.96	1.07	1.13	1.14	0.92	1.5	1.12
SJV	1.62	1.93	1.71	1.27	1.42	1.75	1.52	1.45	1.43	1.34
QQR	0.27	0.32	0.47	0.92	0.93	0.98	1.06	0.8	0.85	1.13
SHQ	2.02	1.65	1.93	2.23	2.13	2.18	2.28	2.63	2.4	2.48

Between 2010 and 2019, there was an internal separation of regional entities into the groups in terms of maintaining the total variance.

Even without using the clustering algorithms (the K-means algorithm can be used if necessary), we can visually identify three clusters (groups) of subjects. It can be concluded that the extent of divergence has not generally worsened, although it has not also decreased during this period. In terms of general economic growth, the subjects' own growth rates were differentiated, which led to a kind of regrouping them.

Subjects with relatively high growth characteristics (Adjara) were separated, promoted and further distanced away from the previous position. Subjects (Mtskheta-Mtianeti, Samtskhe-Javakheti, Samegrelo-Zemo Svaneti, Imereti, Kvemo Kartli) were placed in the middle positions that are closer to the position of the centroid, which largely determines its location.

Their growth rates in the past period are mostly close to the average growth rates. However, this group was formed in different ways. For example, due to high growth rate delays and negative dynamics, Kvemo Kartli appeared in the group, which moved down from the higher positions. It is noteworthy that in the past period some subjects migrated in a positive direction along the ordinate axis, while their position along the abscess axis did not change relatively (Mtskheta-Mtianeti, Samtskhe-Javakheti).



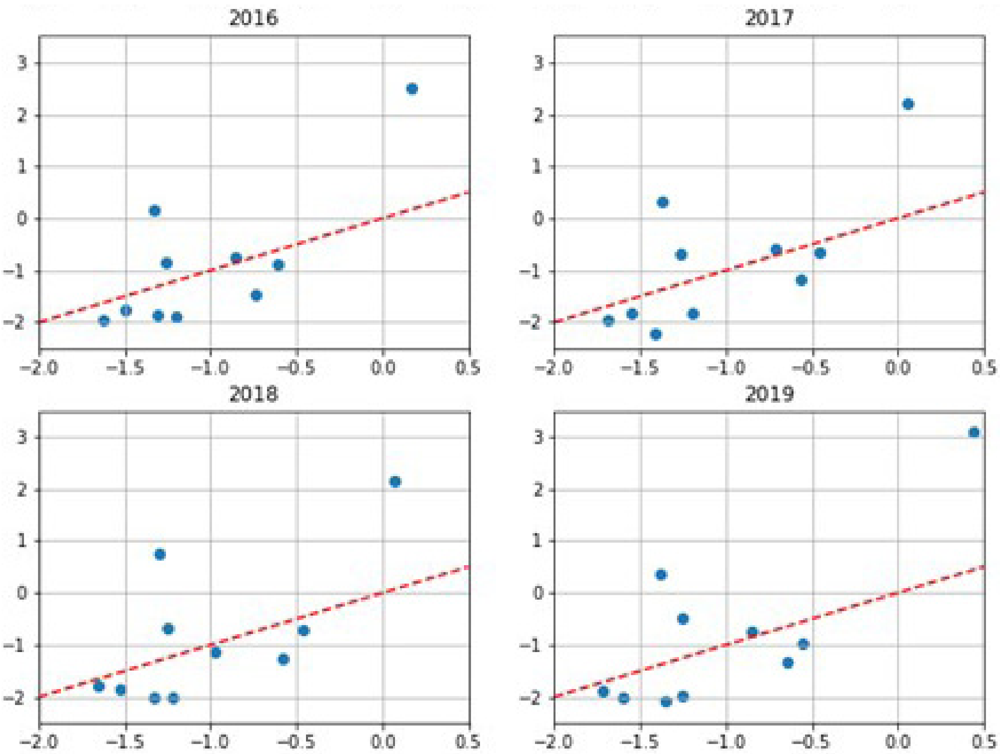


Figure 4. *Dynamics of aggregated data for regions (except Tbilisi), 2010 - 2019*

In the context of maintaining the average trend of change in absolute indicators, the sharp increase in relative rates may be mostly explained by the decline in population, that is, the regions are characterized by the outflow of population. These subjects occupy a kind of intermediate position between this group and the third group of outsiders, they differ from the subjects of the third group only in high relative indicators, and therefore their belonging to this group is conditional, although more appropriate given the dynamic processes.

The dynamics of the third group of subjects (Racha-Lechkhumi-Kvemo Svaneti, Shida Kartli, Guria, Kakheti), which specify the main determinant of the negative trend of the existing divergence, shows that in the past they maintained the previous growth rates and therefore the previous location remained unchanged. To reduce divergence, they need the advance, breakthrough development, including the planning and implementation of special programs.

The analysis of the coefficient of its relative distance from the centroid allows determining the phase of divergence between the regions (see Table 5).

Table 5. *Relative distance coefficients of regions*

Region	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
TBS	5.05	5.03	4.92	4.81	4.61	4.57	4.62	4.67	4.62	4.38
ADJ	0.48	0.48	0.59	0.68	0.73	0.8	0.86	0.78	0.74	1.06
GUR	0.95	0.89	0.92	0.89	0.91	0.83	0.79	0.85	0.84	0.87
IME	0.46	0.49	0.49	0.58	0.62	0.48	0.56	0.46	0.48	0.5
KAH	0.87	0.86	0.79	0.72	0.72	0.8	0.77	0.77	0.81	0.8
MTS	0.56	0.56	0.48	0.48	0.54	0.52	0.46	0.5	0.52	0.48
RLQ	0.85	0.96	0.91	0.94	0.97	0.92	0.87	0.91	0.84	0.87
SZS	0.35	0.3	0.42	0.34	0.37	0.39	0.39	0.33	0.52	0.38
SJV	0.6	0.7	0.61	0.45	0.49	0.6	0.52	0.51	0.5	0.45
QQR	0.1	0.12	0.17	0.33	0.32	0.34	0.36	0.28	0.29	0.38
SHQ	0.74	0.6	0.69	0.79	0.73	0.75	0.78	0.93	0.83	0.84

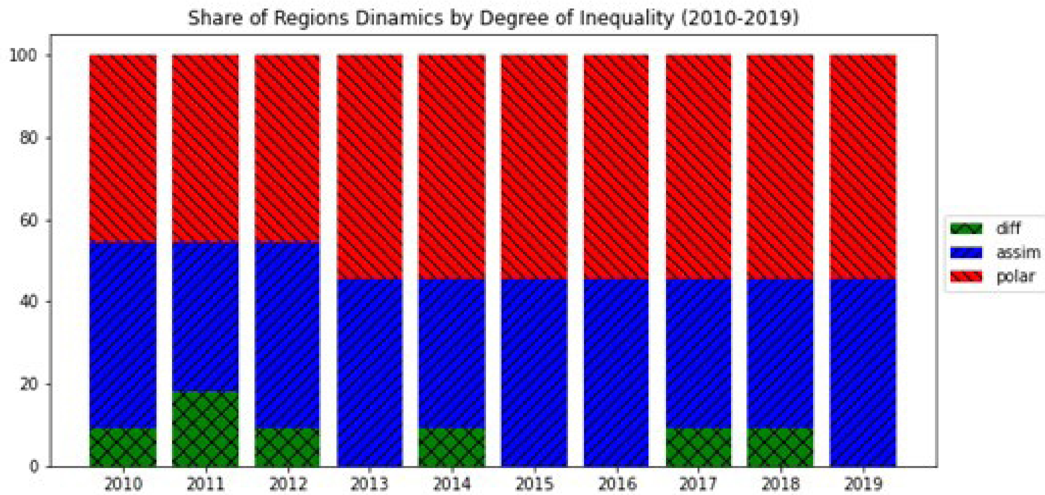


Figure 5. *Dynamics of the degree of divergence of regions in 2010-2019*

Classification of regions according to the phases of divergence will allow us to:

- Reasonably select and implement targeted development programs;
- Provide optimal regulation of regional development;
- Identify the growth poles;
- Identify depressive regions and accordingly plan governance measures to overcome disparities and ensure solid, stable, equitable growth.

According to the classification scale described above, there are significant disparities in terms of the economic situation of the regions. Imereti, Mtskheta-Mtianeti, Samegrelo-Zemo Svaneti, Samtskhe-Javakheti and Kvemo Kartli regions are characterized by asymmetry. The economies of Tbilisi, Adjara, Guria, Kakheti, Racha-

Lechkhumi-Kvemo Svaneti and Shida Kartli regions are polarized. This division indicates a serious divergence and the need to plan and undertake appropriate measures. If we look at the dynamics, the lowest degree of divergence in the past period, differentiation was observed in a number of years with insignificant volume (see Figure 5). The distribution of the degree of divergence in the regions is mainly characterized by a steadily high disparity in the past period. It can be said that the level of divergence has neither converged nor diverged in the last decade.

It is interesting to rank the regions according to their share in divergence (see Table 6).

Table 6. *Share in total divergence (%)*

Region	share %
TBS	79.5
ADJ	4.7
GUR	3.1
IME	1
KAH	2.7
MTS	1
RLQ	3.1
SZS	0.6
SJV	0.8
QQR	0.6
SHQ	2.9

Naturally, this divergence is mostly due to Tbilisi (79.5%). Among the other regions contributing to an increase in this divergence are: Adjara (4.7%), Guria (3.1%), Racha-Lechkhumi-Kvemo Svaneti (3.1%), Shida Kartli (2.9%) and Kakheti (2.7%), while Imereti (1%), Mtskheta-Mtianeti (1%), Samtskhe-Javakheti (0.8%), Samegrelo-Zemo Svaneti (0.6%) and Kvemo Kartli (0.6%) have a small share in the growth of divergence.

Conclusion

Thus, we may conclude that the analysis of economic development divergence that was done by the method of principal components provides quite interesting and important information not only about the divergence between regions, but also about the state and dynamics of the economic development of regions in general. Through this method, it is possible to aggregate different data (including different data types and dimensions) into a single feature.

Reducing the spatial dimension of the characteristics simplifies the analysis, allows visualizing them, and makes the entire research process more efficient. Loss of information during aggregation is negligible. In this case, the coefficient of variation explained for the absolute indicators was 97.1%, while for the relative indicators - 83.6%, which is quite high.

Despite the complex computational procedures, the available software tools greatly simplify performing calculations and the preparation of visual material. The study of the divergence of economic development between the regions of Georgia conducted by the mentioned methodology once again confirmed the existing large disparities in the economic development of regions.

In addition to Tbilisi, three clusters of regions were identified, including the distinctions between the growth poles, middle zone, less developed and depressed regions. Under conditions of maintaining the total variance in dynamics, the internal separation of subjects into clusters continued. Classification by the phases of divergence revealed significant disparities over the last decade. Regions are evenly distributed between asymmetric and polarized degrees, this distribution neither converges nor diverges in dynamics. The methodology allowed us to calculate the share of each region in the formation of overall divergence.

The research indicated that during the last period there was the internal separation of regional entities into groups in the context of maintaining the total variance. Even without using the clustering algorithms (the K-means algorithm can be used if necessary), we can visually identify three clusters (groups) of entities. By analyzing data, we conclude that during this period divergence did not generally increase, although it did not decline either. Under conditions of overall economic growth, the entities' own growth rates were differentiated, which resulted in a kind of regrouping of them. The entities with relatively high growth characteristics (Adjara) were separated, promoted and moved further away from the previous position. The entities that are closer to the position of the centroid (Mtskheta-Mtianeti, Samtskhe-Javakheti, Samegrelo-Zemo Svaneti, Imereti, Kvemo Kartli) were placed in the middle positions, which largely determines its position.

The practical significance of the research is dependent on the factor that the classification of regions by the phases of divergence identified in the research process will allow us to:

- properly choose and implement the targeted development programs;
- provide optimal regulation of regional development;
- identify the growth poles;
- identify depressed regions and, correspondingly, plan management response to disparities in order to ensure sustainable and equitable economic growth.

Studies have revealed significant disparities in terms of the economic situation of the regions. Imereti, Mtskheta-Mtianeti, Samegrelo-Zemo Svaneti, Samtskhe-Javakheti and Kvemo Kartli regions are characterized by asymmetry. The economies of Tbilisi, Adjara, Guria, Kakheti, Racha-Lechkhumi-Kvemo Svaneti and Shida Kartli regions are polarized. This divergence indicates a serious divergence and the need to plan and implement relevant measures.

This methodology can be used to analyze not only regions, but also other territorial units, subdivisions and complexes distinguished by different criteria. It is also possible to present the angle of vision in different ways (for example, in economic and social aspects, etc.). However, in itself, this methodology cannot solve the problem of a complete study of divergence. In terms of sectoral and type analysis, it is necessary to conduct research into the context of each characteristic using appropriate tools.

The method of principal component analysis through the synthesis of regional characteristics allows analyzing the general, aggregate indicators, presents the region as a single, whole economic entity, it also helps us to understand the overall picture, trends and dynamics. Thus, it would be desirable if this method could take its place in the toolkit of specialists working on regional economic issues, in the field of economic policy-making, strategic planning, regional economic analysis and diagnostics, and generating scenarios for the development.

References

Advantages of Principal Component Analysis "What are the Pros and cons of the PCA?". URL:https://www.i2tutorials.com/what-are-the-pros-and-cons-of-the-pca/ [Accessed 21 October 2021]

Baranov S., (2014) Comparative assessments of socio-economic dynamics of the subjects North of Russia // North and Market.№ 2 (39). p. 2–6. (in Russian)

Bressert E., (2012). SciPy and NumPy: an overview for developers. O'Reilly. p. 43.

Burago D., Burago G., Ivanov S., (2004) Course in metric geometry.ISBN 5-93972-300-4 (in Russian)

Engineering Statistics Handbook Section 6.5.5.2

Geostat - statistical information by regions of Georgia URL: https://www.geostat.ge/regions/ (in Georgian) [Accessed 21 October 2021]

Geostat - Statistical information in the Pc-Axis database. URL: http://pc-axis.geostat.ge/PXWeb/pxweb/ka/Database (in Georgian) [Accessed 21 October 2021]

Grus, J., (2015). Data Science from Scratch. Sebastopol, CA: O'Reilly. pp. 99, 100. ISBN 978-1-491-90142-7

Gubanova E., Klesh V., (2018) Methodology for assessing the unequality of socio-economic development of the region - *Problems of development of the territory*, vol. 6 (98), p 30-41. (in Russian)

Han, J., Kamber, M., Pei, J., (2011). Data Transformation and Data Discretization. Data Mining: Concepts and Techniques. Elsevier. pp. 111–118.

Jackson, J., (1991). A User's Guide to Principal Components (Wiley).

Jolliffe, I., (2002). Principal Component Analysis. Springer Series in Statistics. New York: Springer-Verlag. doi:10.1007/b98835, ISBN 978-0-387-95442-4.

Lavrova V., (2012) Modeling the growth of regional socio-economic systems, made by agreeing on differentiations in their development: author. dis. Vladimir,24 p. (in Russian)

Matplotlib Documentation. URL: https://matplotlib.org/stable/contents.html [Accessed 21 October 2021]

Novotný, J. (2007). On the measurement of regional inequality: does spatial dimension of income inequality matter?. *Ann Reg Sci* 41, p 563–580 https://doi.org/10.1007/s00168-007-0113-y [Accessed 21 October 2021]

NumPy v1.20 Manual. URL: https://numpy.org/doc/stable/ [Accessed 21 October 2021]

Pandas documentation. URL: https://pandas.pydata.org/docs/ [Accessed 21 October 2021]

Panzera D., Postiglione P., (2020) Measuring the Spatial Dimension of Regional Inequality: An Approach Based on the Gini Correlation Measure Social Indicators Research 148:379–394 https://doi.org/10.1007/s11205-019-02208-7 [Accessed 21 October 2021]

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Perrot M., Duchesnay E., (2011). "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research*. 12: 2825–2830.

Principal Components Analysis. Institute for Digital Research and Education. UCLA. Retrieved 29 May 2018.

Python Developer's Guide. URL: https://devguide.python.org/ [Accessed 21 October 2021]

Scikit-learn 0.20.1 documentation. URL:scikit-learn.org [Accessed 21 October 2021]

sklearn.decomposition.PCA - Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011. URL: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html [Accessed 21 October 2021]

Sklearn.preprocessing.StandardScaler - Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011. URL: https://scikit-learn.org/stable/

modules/generated/sklearn.preprocessing.StandardScaler.html [Accessed 21 October 2021]

Smith, K., (2013), Precalculus: A Functional Approach to Graphing and Problem Solving, Jones & Bartlett Publishers, p. 8, ISBN 978-0-7637-5177-7

Wang, Y., Fang, C., Xiu, C. et al. (2012). A new approach to measurement of regional inequality in particular directions. Chin. Geogr. Sci. 22, 705–717 https://doi.org/10.1007/s11769-012-0556-7 [Accessed 21 October 2021]