Method for evaluating the robustness of rankings generated by composite indices

T V Zhgun
Yaroslav-the-Wise Novgorod State University, ul. B. St. Petersburgskaya, 41
173003 Veliky Novgorod, Russia
E-mail: Tatyana.Zhgun@novsu.ru

Abstract. Characteristics of the quality of many systems in the economics, politics, society is in the form of a composite index, take the form of a linear composite index that linearly aggregates over several dimensions using a weight vector. The construction of the complex indexes of a system can be considered as the task of extracting the useful signal from the background noise. The signal in this case is the weights of the linear convolution of the indicators, which should reflect the structure of the system being evaluated. Principal component analysis determines the structure of the principal components for successive observations differently. The reason for this may be the presence of fatal errors in the data used. A modification of the principal component method that takes into account the presence of errors in the data used determines the structure of the system unambiguously. If the nature of the aggregation method of variables and the choice of weights adequately reflect the system quality model, then to study the reliability of the composite index, it is necessary to investigate the stability of ratings over time. The consequence of stability is on average a slight change (increment) in the rating of objects for different measurements. This increment can be a posteriori estimated by a set of observations on the proposed dispersion criterion. The results of assessing the stability of different integral characteristics by this criterion are given. Complex indexes, calculated by the author’s method, show good stability.

1. Introduction
Composite indicators of efficacy are regularly used as relevant information tools in different fields, such as politics, economics, health, education. “Composite indices might not rule our world, but they are highly influential in it” [1]. It is especially related to composite international indices which assess achievements of countries in different fields. Rapid increase of the number of composite indices is a marked sign of their significance in politics and economics on the whole [2].

The main field of applying composite indices is the ranking of objects. It is the position of an object relative to other objects that is the basis both for attraction of public attention and making political decisions. If small changes of input data, when calculating a composite index, cardinally influence the ranking of objects made on the basis of computed integral indicators, then such an integral indicator cannot be considered reliable. The necessary sign of reliability of a composite index is the stability of ratings defined by integral characteristics relative to perturbations of the source data.

There are two approaches of investigating the quality of integral indicators. Proponents of the first approach consider that the ratings defined by complex indicators are the functions of weights, and if there is any doubt about the right choice of weights, then there is uncertainty about the reliability of
these ratings. So, the stability analysis of integral indices reduces to the stability analysis of ranks generated by composite indices with weights defined by experts.

The researches [3–9] are directed to the stability analysis of ranks defined by composite indicators with equal weights. In the works [3, 4] the methods for assessing the degree of the rank stability to changing weights reflecting the “degree of confidence” to the used vector of weights are presented.

In the article [1] an approach is suggested to select a reference set of weights (alternative weight schemes) for rank stability assessment. The suggested approach is based on the assumption that there is an agreement about minimal and maximally allowable weights of every variable. This agreement defines the specific set of alternative weights for rank stability assessment.

When studying the ambiguity of the calculation result of composite indices, proponents of the second approach do not make a semantic difference between the variables of linear convolution of indicators defining the composite index [10–13]. The weights and values of indicators are considered as inputs of the model, and the value of a complex indicator is considered as an output variable. Effect of input variables on results is assessed by examining the value of the complex indicator in the entire space of input data. In [12–13] the composite index stability is examined by studying the rating impact of a random change in indicators using Monte-Carlo simulation.

Both approaches consider the complex system quality indicator as a value with a measure of great ambiguity (practically as a random value) and study the impact of input ambiguities of different type on the finite result.

Complex integral characteristics describe a system in general, this allows distracting from side effects associated with the randomness of indicators. All conservation laws in the mechanics and electrodynamics of continuous media are written for integral values. When describing stochastic dynamical systems in the problems of hydrodynamics, magnetic hydrodynamics, astrophysics, plasma physics, radiophysics, the integral values characterizing such systems are their main characteristics and the reliability of such indicators is beyond doubt.

If the nature of the method of aggregating variables and the choice of weight indicators adequately reflect the quality model of the system, then the reliability of the composite index should be evaluated relative to changes in indicators. Therefore, to study the reliability of the composite index, it is necessary to study the stability of ratings over time.

2. Problem statement

Let us consider the making of integral evaluation of the system from \( n \) of objects for which the tables of object description for a number of observations are known – matrices with dimension \( m \times n \)

\[
A^t = \{a_{ij}^t\}_{i=1}^m_{j=1}^n, \quad t = 1, \ldots, T.
\]

The matrix element \( a_{ij}^t \) is the value of \( j \) indicator of \( i \) object, the vector

\[
a_i^t = (a_{i1}^t, \ldots, a_{in}^t)
\]

is the description of \( i \) object at time \( t \). For every time \( t \) the vector of integral indicators looks like

\[
q^t = A^t \cdot w^t,
\]

or, for \( i \) object at time \( t \)

\[
q_i^t = \sum_{j=1}^n w^t_j \cdot a^t_{ij}
\]

where \( q^t = (q_1^t, q_2^t, \ldots, q_m^t)^T \) is the vector of integral indicators of the moment \( t \), \( w^t = (w_1^t, w_2^t, \ldots, w_m^t)^T \) is the vector of weights of indicators for the moment \( t \), \( A^t \) is the matrix of previously processed data for the moment \( t \).
The numeric characteristics of the system have been preliminary unified, - the values of variables have been brought onto the segment \([0, 1]\). If the initial indicator is associated with the analyzed integral quality property by a monotonous dependence, then the initial variables \(x^t_{ij}\) for every observation moment are transformed by the rule:

\[
a^t_{ij} = \begin{cases} 
  s_j + (-1)^{s_j} \cdot \frac{x^t_{ij} - m_j}{M_j - m_j}, & \text{if } j \text{ is maximal;} \\
  s_j, & \text{if } j \text{ is minimal;}
\end{cases}
\]

where \(s_j = 0\), if an optimal value of \(j\) indicator is maximal, and \(s_j = 1\), if an optimal value of \(j\) indicator is minimal; \(m_j\) is the lowest value of \(j\) indicator across the sample (a global minimum), \(M_j\) is the highest value of \(j\) indicator across the sample (a global maximum).

If the initial indicator is associated with the analyzed integral quality property by a non-monotonic dependence (i.e., within the range of change of the given indicator there is a value \(x^w\), at which the highest quality is achieved), then the value of the corresponding unified indicator is calculated by the formula:

\[
a^t_{ij} = 1 - \frac{|x^t_{ij} - x_j^{opt}|}{\max((M_j - x_j^{opt}), (x_j^{opt} - m_j))}.
\]

In further considerations the data to which necessary transformations have been applied will be input data. Let us consider an evaluated object as a complex (which cannot be satisfactorily formalized) large (with the number of states above modern computing capabilities) system. Such systems may be both biological objects and social/socio-economic systems. The system is accessible for observation, and the finite, enough large, number of numeric characteristics of this system registered with some accuracy at different points in time is known. Significance of the registered indicators for the system functioning is unknown in general case. To solve the management problem, it is required to give a motivated estimate for every observed object over the entire observation interval, i.e., to calculate the integral characteristic of the system quality in dynamics. To construct the sought-for integral indicator of the system quality it is required to find weights of indicators \(w^t\) for every moment in time which reflect adequately properties of the system under consideration.

**3. Stability criterion of composite indices**

If the weights are not determined arbitrarily, and there is no doubt about their reliability, we should consider the stability of composite indices relative to changes in the characteristics of objects over time. The observed parameters of changes in the characteristics of the system will determine the set of admissible values of the variables and one can study the behavior of the integral characteristics when the change in the input data obviously does not leave the range of admissible values. A consequence of the stability of the integral characteristic, in particular, is a slight (average) change in the rating of objects for different points in time.

For example, the ranks of 25 countries of the European Union for 2009-2011, based on the values of the Human Development Index (HDI), which calculates the value of the HDI using a linear convolution of variables with equal weights, give an average rating change for the year of 7.7% [14]. A rating change of more than 15% is 14% of cases, more than 30% is 2%. Such rating behavior gives grounds to call a stable integral indicator of the HDI.

In a study determining the quality of life of the population of the Russian Federation according to the method proposed by S.A. Aivazyan, the author of the method [15, 16] ranked the constituent entities of the Russian Federation using two ways: in one case, the weights of the indicators in a linear convolution for each year were determined by experts, in the other, they were determined according to the method proposed by the author. There are cases when the rating change was more than 15%: in the case of expert weights 18, in the case of calculated weights - one. The average rating change for the first
method is 3.6%, for the second - 1.6%, i.e., the proposed method showed the brilliant quality of the constructed composite index.

However, with careful application of this methodology by other authors, the results are completely different. For example, in the work [17], the integral characteristics of the life quality of the municipalities of the Tyumen region for 2005-2008 were calculated. In this case, a rating change of more than 15% of the maximum possible is 48.9% of the total number of cases. In 17.9% of cases, this value exceeds 30%. The average rating change is 16.9%. In the work [18], the values of the ratings of the municipalities of the Samara region which were calculated using the same methodology are given. A simultaneous change in rating of more than 15% of the maximum possible is 45% of the total number of cases. In 21.6% of cases, this value exceeds 30%. The average rating change is 16.9%. In this case, it is worth talking about the instability of the integrated indicator.

If small changes in the input data when calculating the composite index cardinaly change the ranking of objects, then such an integral indicator cannot be considered reliable. A necessary sign of the reliability of a composite index is stability with respect to disturbances in the initial data. In particular, the consequence of this is a slight (on average) change in the rating of objects for different dimensions. Schemes for determining weights using the factor analysis or the principal component analysis do not possess such a property. Next, we define a measure to assess the smallness of changes in the rating of objects.

Let \( R_t = (r_{t1}, r_{t2}, \ldots, r_{tm}) \) be the ratings of \( m \) objects for the moment \( t \). The values of sets of ratings \( R_0 \) are known for the moments \( t = 1, \ldots, T \), which are random values uniformly distributed on the interval \([1, m]\) with numeric characteristics corresponding to a uniform distribution:

\[
M(R_t) = \frac{(m + 1)}{2}, \quad D(R_t) = \frac{(m - 1)^2}{12}, \quad t = 1, \ldots, T. \tag{3}
\]

Expected mathematic values and dispersion of the values \( R_t \) defined by (3) are constant and do not depend on \( t \). Let us note that the values of the ratings for the object \( i \) at successive points in time \( r_{ti}, r_{t+1,i}, r_{t+2,i} \) represent the numeric realization of a complex functional dependence reflecting the properties of the examined system. It is not possible to find a formal description of this dependence. The degree of linear relationship between the sets of ratings \( R_t, R_{t+1} \) is high, the Pearson and Spearman correlation coefficients are close to unity and do not allow to draw conclusions about the quality of the ratings.

Incremental sizes of ratings are more informative to describe the rating quality. Rating changes over time are dictated mainly by the system properties and – to a lesser extent- by random factors. The measure of this randomness can be evaluated. The smaller is the share of randomness for the difference value of the ratings, the better the system structure when selecting weight coefficients is described, the higher is the quality of the integral indicator. Let us suppose a method for evaluating the composite index stability based on the rating unbalance analysis.

If the ratings of objects for consistent observations do not reflect the system properties and are defined absolutely randomly, then the random values of ratings \( R_0, R_{t+1} \) for the successive points in time are independent, and the unbalance dispersion of independent random values \( D(R_{t+1} - R_t) \) is maximum and equal to the sum of dispersions of values \( R_0, R_{t+1} \), uniformly distributed on the interval \([1, m]\).

\[
D_{\text{max}} = D(R_{t+1} - R_t) = D(R_{t+1}) + D(R_t) = \frac{(m - 1)^2}{6}, \tag{4}
\]

where \( m \) is the number of objects.

One can assess the integral characteristic stability, having assessed the measure of randomness of difference in ratings \( \Delta_t \). Such a measure is the dispersion share of value realization \( \Delta_t = R_{t+1} - R_t \) relative to the dispersion (4) which achieves the maximum, if ratings \( R_0, R_{t+1} \) for two consecutive moments of time do not depend absolutely on each other and are completely random.
\[ k = \frac{D(R_f + 1 - R_f)}{D_{\text{max}}} \cdot 100\% \] (5)

Since the randomness is not the chief cause of changing the rating, this share must not be great. Table 1 gives the results of the quality assessment of different integral characteristics by comparing the dispersion share of the difference in the ratings relative to the maximum possible value of such dispersion. As a conventional unit of stability, we can consider the estimate of the integral indicator of the HDI, which is about 7%. The value of the estimate, comparable with this value, will indicate a good stability of the integrated indicator according to the input data. Values that significantly exceed this value characterize the instability of the integrated indicator and, therefore, its low quality. In the given table these are examinations [17] and [18]. Integrated indicators calculated by the author’s method [21, 22] show good stability.

**Table 1.** Comparison of quality indicators of integral characteristics.

| Complex index                                | Source                                      | Period             | Weights source | Number Variables, n x objects, m | Evaluation |
|----------------------------------------------|---------------------------------------------|--------------------|----------------|----------------------------------|------------|
| HDI of the EU countries                      | Human Development Reports, 1990-2014         | 2009-2011          | expert         | 5 x 25                           | 6.9        |
| Quality of life of the Russian Federation    | Ajvazjan, 2003a                             | 1997, 1999         | expert         | 9 x 79                           | 6.3        |
| Quality of life of the Russian Federation    | Ajvazjan, 2003b                             | 1997, 2000         | expert         | 9 x 79                           | 1.2        |
| Quality of life in the Samara Region         | Ajvazjan et al., 2009                       | 2002-2004          | PCA            | 11 x 37                          | 31.5       |
| Quality of Life of the Tyumen Region         | Gajdamak and Hohlov, 2009                   | 2005-2008          | PCA            | 17 x 26                          | 22.7       |
| Quality of life of the Russian Federation    | Zhgun, 2017                                 | 2007-2014          | PCA            | 37 x 83                          | 1.7        |
| Quality of life of the Russian Federation    | (*)                                         | 2007-2016          | PCA            | 37 x 85                          | 1.6        |
| SSI. Human Wellbeing                         | http://www.ssfindex.com/data-all-countries/| 2006-2016          | expert         | 9 x 154                          | 1          |
| SSI. Environmental Wellbeing                 | http://www.ssfindex.com/data-all-countries/| 2006-2016          | expert         | 8 x 154                          | 3.5        |
| SSI. Economic Wellbeing                      | http://www.ssfindex.com/data-all-countries/| 2006-2016          | expert         | 5 x 154                          | 4.5        |

Note: Integral indicators (*) are calculated by the author’s method.

4. Conclusion
The stability of solving a computational problem with respect to changes in input data is a prerequisite for a high-quality solution. In particular, the calculation of composite system quality indices requires a detailed understanding of the effect of the errors in the data used on the calculated output characteristics. These characteristics are the value of the composite index and the rank of the objects defined by this index. The value of the composite indicator is calculated on a discrete set of data and cannot be estimated
analytically. However, it is possible to evaluate the stability of the integral indicator by the change in the rank of objects in time. A consequence of the stability of the composite index is, on average, a slight change (increment) in the rating of objects for successive observations. This increment can be a posteriori estimated by the proposed dispersion criterion. Estimates of the stability of various integral characteristics by the dispersion criterion are given. The criterion compares the dispersion of the rank difference of objects of two consecutive observations with respect to the maximum dispersion of the rank difference, which is observed with absolutely random ranking of objects. If the calculated ratio significantly exceeds the reference value, the complex indicator should be considered unstable and, therefore, unreliable. Composite indices calculated using a modification of the principal component analysis, which takes into account errors in the data used, show good resistance to changes in input data.

References

[1] Seth S and McGillivray M 2016 Composite indices, alternative weights, and comparison robustness OPHI Working Paper 106 (University of Oxford)
[2] Bandura R 2011 Composite Indicators and Rankings: Inventory 2011 ( Working Paper) p 257
[3] Foster J E, McGillivray M and Seth S 2012 Rank Robustness of Composite Indices: Dominance and Ambiguity OPHI Working Papers 26b (University of Oxford)
[4] Foster J E, McGillivray M and Seth S 2013 Composite Indices: Rank Robustness, Statistical Association, and Redundancy Econometric Reviews 32 35–56
[5] Cahill M B 2005 Is the Human Development Index Redundant? Eastern Economic Journal 31 1–5
[6] Chercyhe L, Ooghe E and Puyenbroeck T V 2008 Robust human development rankings Journal of Economic Inequality 6 287–321
[7] Foster J E, McGillivray M and Seth S 2009 Rank robustness of composite indices (Working Paper 26, Oxford Poverty and Human Development Initiative. Oxford: University of Oxford)
[8] Permanyer I 2011 Assessing the Robustness of Composite Indices Rankings Review of Income and Wealth 57 306–326
[9] Zheng B and Zheng C 2015 Fuzzy ranking of human development: A proposal Mathematical Social Sciences 78 39–47
[10] Saltelli A, Ratto M, Andres T and Campolongo F 2008 Global Sensitivity Analysis: The Primer (Wiley-Interscience) p 304
[11] Huergo L, Münnich R and Saisana M Robustness assessment for composite indicators with R https://www.researchgate.net/publication/237308183_Global_Sensitivity_Analysis_The_Primer (accessed 07.09. 2018)
[12] Saltelli A and Sobol I M 1995 Sensitivity analysis for nonlinear mathematical models: numerical experience Matem. Modelirovanie 7 (11) 16–28
[13] Jacobs R, Smith P and Goddard M 2004 Measuring Performance: An Examination of Composite Performance Indicators (York, UK: Centre for Health Economics, University of York) p 124
[14] Zhgun T V 2017 Building an integral measure of the quality of life of constituent entities of the Russian Federation using the principal component analysis Economic and Social Changes: Facts, Trends, Forecast 10 2 214–235 doi: 10.15838/esc/2017.2.50.12
[15] Zhgun T 2017 An algorithm for constructing integral quality indicator of complex systems for a sequence of observations Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering Vol 6 1 p 5–25 (in Russian) doi: 10.14529/cmse170101
[16] Human Development Reports 1990-2014 United Nations Development Programme 1990-2014 URL: http://hdr.undp.org/en/reports/
[17] Ajvazjan S A 2003 Empirical analysis of synthetic categories of life quality of the population Economics and mathematical methods 3 39 19–53
[18] Ajvazjan S A 2003 The methodology for the measurement of synthetic categories of life quality of the population Economics and mathematical methods Vol 39 2 33–53

[19] Gajdamak I V and Hohlov A G 2009 Modeling of integral indicators of quality of life of the population of the South of the Tyumen region Vestnik TjumenSU 6 176–186

[20] Ajvazjan S A, Stepanov V S and Kozlova M I 2009 Applied Econometrics Measurement of synthetic categories of quality of life of the population of the region and identify key areas of improvement in socio-economic policy (in the Samara region and its municipalities an example) 3 (19) 18–84

[21] Zhgun T V 2017 Building an integral measure of the quality of life of constituent entities of the Russian Federation using the principal component analysis Economic and Social Changes: Facts, Trends, Forecast 10 2 214–235 doi: 10.15838/esc/2017.2.50.12

[22] Zhgun T 2017 An algorithm for constructing integral quality indicator of complex systems for a sequence of observations Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering Vol 6 1 5–25 (in Russian) doi: 10.14529/cmse170101