Synthetic Indicators of Quality of Life: Construction and Utilization for Social-Economic Management and Comparative Analysis

Serguei Aivazian

Central Economics and Mathematics Institute of RAS, Moscow

Abstract: Proposed basic concepts and methodology, and described information sources are aimed to construct and use in practice the aggregative quality of life indicators at the level of a region or a country. The methodology exploits factor analysis (with "no training samples" available) and "expert and statistical regression model" (with "training samples" available), and includes a multicriterion scheme.

1 Introduction

The problem of defining and measuring the quality of life of population and its synthetic categories might be looking like too ambitious, as pretending to discover a sort of "philosophers stone", if the problem were not commented properly.

First of all, it is important to put clearly that an aggregated indicator of the quality of life (which we construct as a function of partial features of the synthetic categories analyzed) have a conditional meaning within spatial and temporal limits only. For instance, the anthropogenic factor in ecology was not significant earlier as compared with the other features of quality of life. In particular, the population was apparently not able to cause or to prevent irreversible changes of the natural environment just few decades ago. So, the number and contents of the partial features being synthesized within an integral feature of quality of life are dependent on a series of objective and subjective time-and-space varying factors at any given period of history.

Secondly, the methodology of constructing and interpreting of the synthetic indicators of quality of life are to be defined in accordance with the goals of their intended application.

A wide spectrum of studies related to constructing and analysis of synthetic (latent usually) features of quality of life and life style of population has been reflected in special publications. Here follow some examples:

(i) measuring of the population's health and quality: human development index, index of health, etc. (HDR);

(ii) measuring of the institutional development level of society (Rodrik, 2003);

(iii) measuring of the environment quality (Shakin, 1998).

(iv) annual monitoring of the latent synthetic indicators of quality of life and mode of life (for a set of countries) such as "quality of life", "quality of education system", "personal and property's protection", "level of social cohesion", and al. (WCY);

(v) system of the social indicators (see, for example, SIR).

Thus, one can see that experts on macroeconomic analysis and sustainable development have agreed that is it worthwhile to consider and to assess the latent synthetic categories of quality of life and sustainable development.

Of course, destination, interpretation and methodology of constructing the synthetic indicators depend on a type of social paradigm, that is a frame of our analysis. Our approach is correspondent to *a structural-functional type*. That means that social structure has a leading role in development of a society, so correspondingly **the society as a whole** is a direct object of our analysis.

There exists a parallel approach based on *paradigm of interactionistic type.* According to the approach a direct object of analysis is a system of functional needs of an individual and extent to what these needs are met (see, for example, Bowling, 1994). The approach can be implemented by means of the survey studies and analysis of a number of special questionnaires. The approach is not within our study.

A series of the socio-economic problems have been solved using the synthetic indicators of *Quality of Life* (QoL, for short) and *Sustainable Development* (SD, for short). Main customers and users of *the Quality of Life and Sustainability Indicators* (QoL-SI, for short) were the Russian Federal Ministry of Economical Development and Trade and the State Committee on Statistics of the Russian Federation (GOSKOMSTAT) during the last 2-3 years.

2 A Hierarchical System of Statistical Indices, Partial Criteria and Synthetic Indicators of QoL

The following Figure 1 presents a hierarchical system of statistical indices, partial criteria and synthetic indicators of QoL that were used in our study. One can see five synthetic categories of quality of life and sustainable development. Each of them is characterized with a set of partial criteria that are structured, as presented here. In its turn, each of the parts enlisted in the left column is characterized with its own **a priori set of initial statistical indices**. For example, a priori set of statistical features of "reproduction and physical health of population" contains such statistical features as "infant mortality", "natural increment of population", "life expectancy at birth", "tuberculosis-caused mortality", "infectious and parasite-type diseases caused mortality", "a number invalids and people with the congenital defects per 1000 people", etc.

Each of the synthetic categories is characterized with one or more latent integral indicators. Each of the integral indicators is constructed as a convolution (function) of a definite set of statistically recorded features.

Now, let us come shortly to the methodology of constructing the statistical indicators of QoL and SD.

3 Methodology of Measuring the QoL Latent Synthetic Categories based on the Available Statistical Data

Let the variables $x^{(1)}, x^{(2)}, \ldots, x^{(p)}$ be the statistical indices (partial criteria) of the latent synthetic category under consideration. The problem is to construct some aggregated index (synthetic indicator) y of the synthetic category analyzed (QoL-SI) in form of a special convolution (function) of the variables $x^{(1)}, \ldots, x^{(p)}$, i.e.

$$y = f(x^{(1)}, x^{(2)}, \ldots, x^{(p)}) \tag{1}$$

The problem might not have a satisfactory solution (i.e. multi-criteria scheme might not be reduced to a single-criteria scheme), and then one should consider construction of *several* latent synthetic indicators

$$\begin{aligned} y^{(1)} &= f_1(x^{(1)}, \ldots, x^{(p)}), \\ &\cdots\cdots\cdots\cdots\cdots \\ y^{(k)} &= f_k(x^{(1)}, \ldots, x^{(p)}), \end{aligned} \tag{2}$$

where $k \ll p$ (k is significantly less than p). Definition of possibly minimal k is one of the questions arising within the problem.

3.1 Unified Scales of Measurements

Prior to constructing the statistical indicators y we transform the explanatory variables (partial criteria) $x^{(1)}, x^{(2)}, \ldots, x^{(p)}$ to unify their measuring scales. Particularly:

$$\tilde{x}^{(j)} = \left(1 - \frac{\left|x^{(j)} - x_{opt}^{(j)}\right|}{\max\left\{\left(x_{max}^{(j)} - x_{opt}^{(j)}\right), \left(x_{opt}^{(j)} - x_{min}^{(j)}\right)\right\}}\right) \cdot N. \tag{3}$$

Here $x_{min}^{(j)}$ and $x_{max}^{(j)}$ are correspondingly minimal and maximal values of $x^{(j)}$, and $x_{opt}^{(j)}$ is an optimal value of the variable, i.e. the value corresponding to the best quality (of life or of sustainable development).

That way all the variables $\tilde{x}^{(j)}$ ($j = 1, 2, \ldots, p$) will be measured *in the same N-marks scale*, and values O and N will define the worst and the best quality correspondingly. It is evident that if the analyzed synthetic category of quality *is dependent* on variable $x^{(j)}$ *monotonically and increasing with it* (and then $x_{opt}^{(j)} = x_{max}^{(j)}$), then transform (3) can be presented in the following form:

$$\tilde{x}^{(j)} = \frac{x^{(j)} - x_{min}^{(j)}}{x_{max}^{(j)} - x_{min}^{(j)}} \cdot N. \tag{4}$$

Similarly if the analyzed synthetic category of quality *is decreasing with it* variable $x^{(j)}$ (and then $x_{opt}^{(j)} = x_{min}^{(j)}$) then transform (3) is reduced to the following form:

$$\tilde{x}^{(j)} = \frac{x_{max}^{(j)} - x^{(j)}}{x_{max}^{(j)} - x_{min}^{(j)}} \cdot N. \tag{5}$$

3.2 Construction of Aggregated (Synthetic) Indicators of QoL in the Presence of "the Training"

In this case an initial information consists of two parts: *statistical* and that of the *expert*.

(a) **Statistical part** contains the observed values $\tilde{x}_i^{(1)}, \ldots, \tilde{x}_i^{(p)}$ of statistical indices (partial criteria) $x^{(1)}, \ldots, x^{(p)}$ ($i = 1, 2, \ldots, n$; n is total number of observations) of QoL category analyzed.

(b) **Part of the expert** contains some information of expert about the values of synthetic indicator y. The experts can to submit this information in the following form:

(b.1) as expert mark estimates $\hat{y}_i$ ($i = 1, 2, \ldots, n$) in a given scale (for example in ten-marks scale) of the extent to what the synthetic category is revealed;

(b.2) as a partition of the analyzed objects (countries, regions) on a number of groups homogeneous over y and ordered in according to the extent to what the synthetic category is revealed;

(b.3) as a matrix

$$\Delta = (\delta_{ij}), \quad i, j = 1, 2, \ldots, n$$

of pair-wise comparisons filled in at least partly where row $(\delta_{i1}, \delta_{i2}, \ldots, \delta_{in})$ defines comparison of 'quality' of i-th object to the other objects, as follows:

$$\delta_{ij} = \begin{cases} 1, & \text{if } i\text{-th object is better than } j\text{-th object;} \\ 0 & \text{otherwise,} \end{cases}$$

or

$$\delta_{ij} = \begin{cases} 1, & \text{if } i\text{-th object and is } j\text{-th object are of the same qualty;} \\ 0 & \text{otherwise.} \end{cases}$$

A part of initial information coming from the experts is referred to as "learning" in the problem of constructing integral indicators of quality.

It is worthwhile to stress that practical implementation of learning in version (b.1) is the most difficult for experts and therefore it is very rare in applications. It is much easier to obtain learning information in form (b.2) and even more in form (b.3) that are less demanding towards the experts, so they can be used in practical applications. For instance, due to our own experience the (b.2) form was useful to make partition of analyzed regions of Russia (according to a given synthetic category) on three ranked parts, such as: *"leading regions"*, *"undistinguished regions"*, and *"regions outsiders"*.

To construct latent integral indicator of analyzed synthetic category using initial data (a)-(b), let us consider *linear* regression, as follows:

$$\hat{y}_i = \beta^T \tilde{X}_i + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{6}$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ – is vector of unknown weighting coefficients, $\tilde{X}_i = (\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}, \ldots, \tilde{x}_i^{(p)})^T$ – values of unified explanatory variables (partial criteria) recorded on the i-th object, and ε_i – random error terms of the model. Then the problem is reduced to estimation of unknown coefficients β using initial data (a)-(b) of various types.

3.2.1 Case of Initial Data of (a)-(b.1) Type.

In that case coefficients are defined using conditional least squares method, i.e.

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^{n} w_i (\hat{y}_i - \beta^T \tilde{X}_i)^2, \tag{7}$$

$$\text{under condition: } \beta_j \geq 0, \sum_{j=1}^{p} \beta_j = 1, \tag{8}$$

where the weights w_i are defined in each particular case that depending on a priori information concerning stochastic nature of the random disturbance components ε.

3.2.2 Case of Initial Data of (a)-(b.2) Type.

In that case problem of estimating parameters β is solved within model of discrete (either binary or multiple) choices. The model is well-known in econometrics (see, for example, Greene, 2001).

3.2.3 Case of Initial Data of (a)-(b.3) Type.

Let $\delta_{ij}(\beta)$ – be elements of matrix $\Delta(\beta)$ of pair-wise comparisons of values of latent integral indicator $y_i = \beta^T X_i$. Namely, let us define value of $\delta_{ij}(\beta)$ as follows:

$$\delta_{ij}(\beta) = \begin{cases} 1, \text{ if } \sum_{l=1}^{p} \beta_l \left(\tilde{x}_i^{(l)} - \tilde{x}_j^{(l)} \right) > 0, \\ 0 \quad \text{otherwise.} \end{cases} \tag{9}$$

Let Hamming distance

$$\rho\left(\Delta(1), \Delta(2)\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} |\delta_{ij}(1) - \delta_{ij}(2)| \tag{10}$$

be defined as measure of distance between two matrices of pair-wise comparisons.

Then estimators $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)^T$ of the parameters $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ are defined as the solution of the following optimization problem:

$$\hat{\beta} = \arg \min_{\beta} \rho(\Delta, \Delta(\beta)), \tag{11}$$

$$\text{under condition: } \beta_j \geq 0, \quad \sum_{j=1}^{p} \beta_j = 1, \tag{12}$$

where Δ is matrix of expert pair-wise comparisons, $\Delta(\beta)$ is matrix of pair-wise comparisons defined according to (9), and distance $\rho(\Delta, \Delta(\beta))$ is defined according to the relationship (10).

Comment concerning restrictions (8) and (12). These restrictions are of technical nature and they are introduced for the sake of convenience to use and to interpret latent integral indicator constructed and parameters β.

3.3 Construction of Aggregated (Synthetic) Indicators of QoL with no "Training"

In that case construction of integral indicator (measure of analyzed synthetic category of quality of life) is based on the following idea. Let $\mathbf{Y}$ be a class of all possible linear functions of unified partial criteria $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(p)}$. Then, as well known, the 1st principal component $y^{(1)}$ of the variables $\tilde{x}^{(1)}, x^{(2)}, \ldots, \tilde{x}^{(p)}$ is a solution of the following optimization problem:

$$y^{(1)} = \arg\min_{y \in \mathbf{Y}} \sum_{j=1}^{p} \mathbf{E}\left(\tilde{x}^{(j)} - a_0^{(j)}(y) - a_1^{(j)}(y) \cdot y\right)^2, \tag{13}$$

where $a_0^{(j)}(y)$ and $a_1^{(j)}(y)$ are the least squares estimators of regression $\tilde{x}^{(j)}$ on y. In other words, the 1st principal component of the variables $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(p)}$ is such a linear combination of these variables that supplies the best fit (in sense of (13)) to partial criteria $\tilde{x}_i^{(1)}, \tilde{x}_i^{(2)}, \ldots, \tilde{x}_i^{(p)}$.

However, if the percentage of total variation of partial criteria (i.e. value of $\sum_{j=1}^{p} Var(\tilde{x}^{(j)})$), explained by variation of the 1st principal component is not big enough (we used the 55% threshold in our study) then precision of approximation of values $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(p)}$ happens to be not sufficient. In that case it is recommended to construct **several** latent integral indicators $y^{(1)}, \ldots, y^{(m)}$ characterizing the analyzed synthetic category. Then the number m of such integral indicators is defined as follows:

$$m = \min_{1 \le k < p-1} \left\{ k : \frac{\lambda_1 + \cdots + \lambda_k}{\lambda_1 + \cdots + \lambda_k + \cdots + \lambda_p} \ge 0.55 \right\}. \tag{14}$$

Here, in relationship (14) $\lambda_1, \lambda_2, \ldots, \lambda_p$ ($\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_p \ge 0$) are eigenvalues of covariance matrix of the variables $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(p)}$.

In case of $m = 1$ we use only one *unique* integral indicator $y^{(1)}$. Correspondingly, its values obtained according to relationship (13) are used for classification and ranking of regions (countries).

In case of $m > 1$ we construct m integral indicators $y^{(1)}, \ldots, y^{(m)}$ measured in unified N-marks scale. Then classification and ranking of regions (countries) according to the analyzed synthetic category is to be done using values of distances d_i between 'etalon' and i-th region (country), i.e. according to values

$$d_i = \sqrt{\sum_{j=1}^{m} \lambda_j (\tilde{y}_i^{(j)} - N)^2}. \tag{15}$$

Let us explain how integral indicators $\tilde{y}^{(1)}, \ldots, \tilde{y}^{(m)}$ are constructed in that case. First of all, the set $\tilde{x}^{(1)}, \tilde{x}^{(2)}, \ldots, \tilde{x}^{(p)}$ of partial criteria is partitioned onto m groups in such a way that correlation between variables within each group would be as maximal as possible, whereas correlation between variables belonging to different groups would be as weak as possible. That problem is solved by means of the "*procedure of extremal grouping of variables*", as explained in, for instance, in Aivazian and Mkhitarian (2001). After

that, the 1st principal components $y^{(1)}, \ldots, y^{(m)}$ are constructed for each group *separately*. Finally, *unified* values $\tilde{y}^{(1)}, \ldots, \tilde{y}^{(m)}$ of these 1st principal components are computed using formula (3).

4   Some Results of Empirical Analysis of Synthetic Categories of Quality of Life of the Population

The methods described here have been used to fulfill inter-country and interregional analysis of various synthetic categories of quality of life of the population. *Inter-country analysis* has been conducted on synthetic category of the top level – on *"quality of life"* using data of type (a) ("without learning") and data of type (a)-(b.1) ("with learning") obtained from WCY.

Comparative analysis of the Russian regions has been done on synthetic categories "quality of population", "material welfare" and "quality of social sphere" using data of type (a) ("without learning") and data of type (a)-(b.2) ("with learning") obtained from the official publications of the State Committee on Statistics of the Russian Federation (for years 1997-1999). It is possible to reveal main factors of tendencies (both positive and negative) in socio-economical development of each region of Russia by means of observing dynamics of integral indicators (and related rankings) and analyzing "weights" β.

Results of the analysis are presented in details by Aivazian (2003).

References

S.A. Aivazian. Empirical analysis of the synthetic categories of the population quality of life. *Economics and Mathematical Methods*, 39(3):18–52, 2003. In Russian.

S.A. Aivazian and V.S. Mkhitarian. volume 1 of *Probability Theory and Applied Statistics*. Unity, Moscow, 2001. In Russian.

Ann Bowling. *Health: Review of Quality of Life Measuring Scales*. Open University Press, Buckingham and Bristol, 1994.

W.H. Greene. *Econometric Analysis*. Macmillan Publishing Co., New York, 2001.

HDR. *Human Development Report*. UND-Program, New York, 2002.

D. Rodrik. Institution, integration, and geography. In *Search of Prosperity – Analytic Country Studies on Growth*, number 7. Princeton University Press, 2003.

V.V. Shakin. Global environmental assessment, sustainable development criteria and mathematical modeling of living systems under extreme conditions. In *Proceedings of the International Conference on Environmental Indices "System Analysis Approach (INDEX-97)"*, pages 202–220, London, U.K., 1998. Encyclopedia of Life Support Systems Publisher, Ltd.

SIR.   *Social Indicators Research. An International and Interdisciplinary Journal for Quality-of-Life Measurement*, volume 55.   Kluwer Academic Publishers, Dodrecht/Boston/London, 2001.

WCY.   *The World Competitiveness Yearbook*.   Edition IMD-International, Lausanne, Switzerland, 1996-2002.

Author's address:

Prof. Dr. Serguei Aivazian
Central Economics and Mathematics Institute of the Academy of Sciences
47 Nakhimovsky prospect
117418 Moscow
Russia

Tel. +7(095)1291300
Fax +7(095)7189615
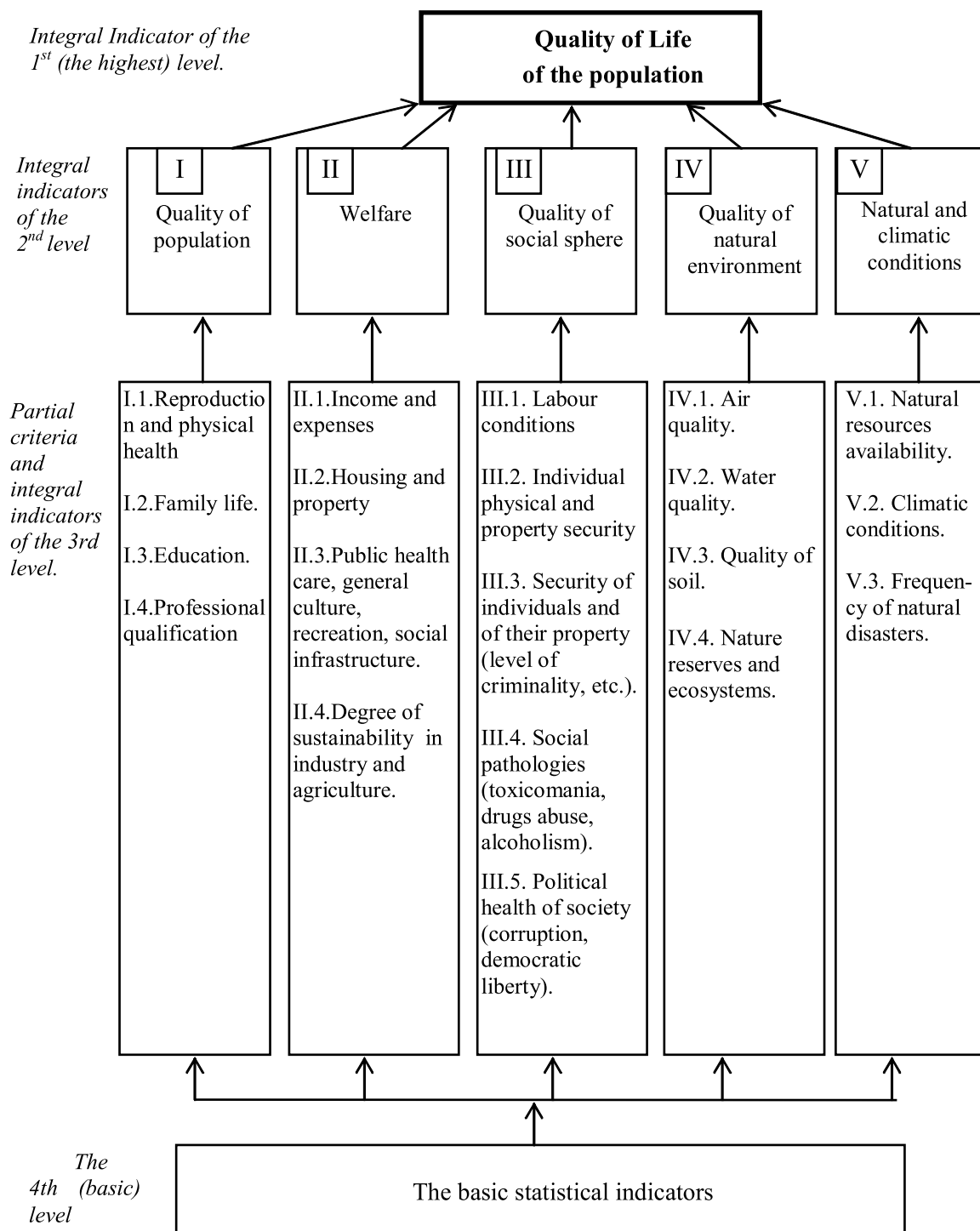E-mail: aivazian@cemi.rssi.ru
http: //www.cemi.rssi.ru

Figure 1: A hierarchical system of statistical indices, partial Criteria and synthetic indicators of QoL that were used in the study