Computation Algorithm for Integral Indicator of Socio-Economic Development

Oleg Pursky[0000−0002−1230−0305], Tetiana Dubovyk[0000−0001−9223−4629], Iryna Gamova[0000−0002−2032−8578], and Iryna Buchatska[0000−0003−2413−7370]

Kyiv National University of Trade and Economics, 02156, Kyiv, Ukraine
o.pursky@gmail.com

Abstract. The computation algorithm for determination of the socio-economic development integral indicator based on the methods of factor analysis and expert evaluations has been described in the paper. By taking into account the knowledge and experience of experts, the factor model for evaluation of the level of socio-economic development has been improved. Based on the joint use of the methods of factor analysis and expert evaluation, the algorithm of automated computation for integral indicators has been developed. The approach has increased the reliability of the results of calculations and made it possible to analyze the correlations between indicators in terms of their influence on the overall socio-economic situation. The developed computation algorithm is used in the educational process within the framework of teaching the discipline "Prognostics of socio-economic processes".

Keywords: socio-economic development, integral indicator, factor analysis, expert evaluation.

1 Introduction

The main problem that arises when using methods of factor analysis in socio-economic studies is reliable conclusions. In statistical calculations, the importance of specific indicators for the socio-economic system is not taken into account. In this case, only the weighted average of such indicators is considered. This problem is solved by expert evaluation. Knowledge and experience of experts make it possible to rank the indicators in terms of their importance for ensuring the effective functioning of the socio-economic system. At the same time, however, expert evaluation fails to establish the correlation between socio-economic indicators. This task is successfully managed by the factor analysis.

Thus, in order to increase the reliability of the procedures for evaluation of the level of socio-economic development, there was a need to improve the mechanism for determination of the integral indicators on the basis of factor analysis by taking into account knowledge and experience of experts in the calculation procedures. It is an expert statistical option that is most suitable, since taking into account knowledge and experience of experts in calculations significantly increases the reliability of the conclusions obtained in the study. At the same time, it gives a chance to perform an extensive socio-economic analysis by establishing correlations between indicators and to determine the influence of the change of a particular indicator (indicators) on the state of the system. Accordingly, within the framework of this study, the algorithm of automated evaluation of the level of socio-economic development has been developed on the basis of the joint use of the methods of factor analysis and expert evaluation.

2 Results and discussion

2.1 Mathematical model for determination of the socio-economic development integral indicator

In the framework of the presented studies, it is proposed to use a two-stage approach in order to calculate the socio-economic development integral indicator. At the first stage, the dimension of the initial feature space is reduced. Reducing the dimension of the feature space is based on the use of factor analysis [1, 2]. The approach is based on the transition from the description of a certain set of objects under study, given by a large set of indirect, directly measured features, to the description of a smaller number of maximally informative substantive variables (factors) that reflect the most important properties of the socio-economic phenomenon. In order to obtain such a reduced set of factors, one of the methods of factor analysis is used, namely the principal component analysis (PCA) [3].

The next stage is to obtain the one integral indicator based on the reduced set of independent factors, which would combine all these factors in the best way [4]. The determination of the most important factors makes it possible to optimize the process of making managerial decisions, and, as a result, improve the overall efficiency of the governance system.

Let us directly consider the model in which the factor is the estimated value, in other words, it represents a certain new characteristic of the studied set of objects. The description of the factor in terms of its connection with the set of initial indicators is in the form of an $n \times m$ matrix of factors A, where n is the number of features, m is the number of factors. The basis for constructing the matrix of factors A is the $n \times m$ matrix of pairwise correlations R. It reflects the degree of correlation between each pair of initial indicators, while the factor matrix characterizes the correlation between each of the n indicators and the m factors determined during the progress of analysis. In this case, the number of factors m should be significantly less than n, and the level of loss of informativeness is negligible.

We assume that there is a set $G(i = 1, 2, ...)$ of observations of a particular studied socio-economic phenomenon. In this case, the phenomenon is described by a set of $n(j = 1, 2, ...)$ features. That is, the information presented in the socio-economic study can be described as a $G \times n$ matrix Θ:

$$\Theta = \begin{vmatrix} \theta_{11} & ... & \theta_{1j} & ... & \theta_{1n} \\ ... & ... & ... & ... & ... \\ \theta_{i1} & ... & \theta_{ij} & ... & \theta_{in} \\ ... & ... & ... & ... & ... \\ \theta_{G1} & ... & \theta_{Gj} & ... & \theta_{Gn} \end{vmatrix}, \tag{1}$$

As a rule, the features selected to describe the socio-economic phenomenon have different dimensions, and therefore different scalability. In order to ensure the possibility of comparing the features of an object and avoiding the influence of their dimension, the initial data matrix Θ is usually transformed (normalized), introducing a single scale for all features. The most common ways of obtaining a normalized data matrix Z_{ij} are standardization [4]:

$$Z_{ij} = \left(\theta_{ij} - \overline{\theta}_j\right)/s_j, \tag{2}$$

where θ_{ij} is a value of j-th feature of i-th object; $\overline{\theta}_j$ is an arithmetic mean value of j-th feature; s_j is a mean-square deviation of j-th feature (dispersion of j-th feature).

As a result of the standardization of the indicators, we obtain a $G \times n$ matrix of the normalized values of the observations. Thus, a normalized matrix is obtained. Now it consists of vectors whose coordinates are indicators of socio-economic development.

According to the factor model [4], each of the features Z_j included in the study set can be represented as a function of a small number of common factors $F1, F2, \ldots, F_m$ and the characteristic factor U_j:

$$Z_j = f\left(F_1,\ F_2,\ ...,\ F_m,\ U_j\right), \tag{3}$$

The application of factor analysis to the matrix of pairwise correlations between the initial indicators, on the basis of which the statistical weight of the factor is determined, makes it possible to represent the initial indicators by factors using the principal component analysis [3]:

$$Z_j = \sum_{p=1}^{n} a_{jp} F_p, \tag{4}$$

The coefficients a_{jm} are called factor loadings and characterize the significance of each of the factors for describing the j-th feature. Factor loadings are correlation coefficients between the initial indicators and factors. Let us write the expression (4) in vector form:

$$\mathbf{Z} = \mathbf{AF}, \tag{5}$$

where $\mathbf{F} = (F_{1,}\ F_2,\ ...,\ F_n)^T$ is a centered random column-vector of uncorrelated principal components; $\mathbf{Z} = (Z_{1,}\ Z_2,\ ...,\ Z_n)^T$ is a centered random column-vector of initial features; $\mathbf{A} = (a_{ij})$ is a nonrandom matrix of factor loadings of random values Z_i on the components $F_j (i = 1, 2, \ldots, n; j = 1, 2, \ldots, n)$.

We consider that $\mathbf{\Omega}=\mathbf{M}(\mathbf{Z}\mathbf{Z}^{\mathbf{T}})$ is a covariant matrix of vector $\mathbf{Z}$. Being symmetric and positive definite, it has n positive eigenvalues λ_1, λ_2, ..., λ_n. Let us assume that $\lambda_1 > \lambda_2 > \ldots > \lambda_n$. We denote:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ddots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}, \tag{6}$$

If $\mathbf{x}_j = (x_{1j}, x_{2j}, ..., x_{nj})^T$ are normalized eigenvectors-columns of matrix $\mathbf{\Omega}$, which correspond to their eigenvalues $\lambda_j (j = 1, 2, ..., n)$, then for all $j = 1, 2, ..., n$ the following equalities are valid:

$$\det |\mathbf{\Omega} - \lambda_{\mathrm{j}} \mathbf{I}| = 0, \tag{7}$$

where $\mathbf{I}$ is an n-th order unit matrix. From this follows:

$$\mathbf{\Omega} \mathbf{x}_j = \lambda_j \mathbf{x}_j, \tag{8}$$

$$\mathbf{x}_p^T \mathbf{x}_j = \sum_{i=1}^{n} x_{ip} x_{ij} = \delta_{pj} = \begin{cases} 1, \, p = j \\ 0, \, p \neq j \end{cases}, \tag{9}$$

Let us introduce the matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$. Since, taking into account (8) and (9)

$$\mathbf{x}_j^T \mathbf{\Omega} \mathbf{x}_p = \lambda_j \mathbf{x}_j^T \mathbf{x}_p = \begin{cases} \lambda_j, \, p = j \\ 0 \quad p \neq j \end{cases}, \tag{10}$$

then

$$\mathbf{X}^T \mathbf{\Omega} \mathbf{X} = \Lambda, \tag{11}$$

Let us assume that

$$\mathbf{F}' = \mathbf{X}^T \mathbf{Z}, \tag{12}$$

and since

$$\mathbf{M}\mathbf{F}' = \mathbf{M}(\mathbf{X}^T \mathbf{Z}) = \mathbf{X}^T \mathbf{M} \mathbf{Z}, \tag{13}$$

then $\mathbf{F}'$ is a centered vector, and since

$$\mathbf{M}(\mathbf{F}'\mathbf{F}'^T) = \mathbf{M}(\mathbf{X}^T \mathbf{Z} \mathbf{Z}^T \mathbf{X}) = \mathbf{X}^T \mathbf{M}(\mathbf{Z}\mathbf{Z}^T) \mathbf{X} = \mathbf{X}^T \mathbf{\Omega} \mathbf{X}, \tag{14}$$

then by virtue of (11) the components of the vector $\mathbf{F}'$ are uncorrelated, and the dispersion of factors D is equal to

$$\mathbf{D} F_j' = \lambda_j (j = 1, 2, ..., n), \tag{15}$$

that is, $\mathbf{F}'$ is a vector of the principal components $\mathbf{F}$, which is calculated in accordance with (14) as follows:

$$F_j = \sum_{i=1}^{n} x_{ij} Z_i \ (j = 1, 2, ..., n), \tag{16}$$

Let us find the matrix of factor loadings $\mathbf{A}$. Using the orthogonality of the matrix $\mathbf{X}$ and the equation (12) we obtain:

$$\mathbf{XF} = \mathbf{XX}^T\mathbf{Z} = \mathbf{XX}^{-1}\mathbf{Z} = \mathbf{Z}, \tag{17}$$

taking into account (5) the expression (17) can be written as

$$\mathbf{A} = \mathbf{X}, \tag{18}$$

in other words, the factor loadings a_{ij} are the components of the eigenvectors x_{ij} of the matrix of pairwise correlation indicators.

In fact, for analysis, $n' < n$ first principal components are used, which exhaust at least $60-70\%$ of the initial random variables [3]. Within the framework of this model, one can use the mean-square deviation of factors as statistical weighting coefficients [4]:

$$\nu_i = \sqrt{D_i} \ \ (i = 1,\ 2,\ ...\ ,\ n), \tag{19}$$

or if we take into account (15) we obtain

$$\nu_i = \sqrt{\lambda_i} \ \ (i = 1,\ 2,\ ...\ ,\ n), \tag{20}$$

That is, for statistical determination of weighting coefficients it is possible to use calculated earlier eigenvalues of the correlation matrix of initial indicators. The larger the difference in the values of objects by factor, the greater the statistical weight of this factor.

The main disadvantage of statistical methods is the reliability of the conclusions, in particular, in this statistical mechanism of determining the integral indicators, the weight of the factor is determined by the dispersion of initial indicators, which is not always reliable in socio-economic studies, since in this case the importance of indicators for the socio-economic system is not taken into account. Therefore, in the framework of this study, in order to increase the reliability of the algorithm for evaluating the level of socio-economic development based on factor analysis, the implementation of an expert evaluation procedure in the mechanism of determining the weight of factors is proposed. The conclusions that change from method to method depend on the subjectivity of choosing the method of processing expert evaluations. In connection with this circumstance, it seems expedient in the automation procedures of expert evaluation of weighting coefficients of factors to make a simultaneous use of the method of median M_i and the method of scoring of indicators y_i. An important feature of this mechanism of ranking socio-economic indicators by various experts is the possibility of minimizing the factor of subjectivity of expert evaluations by virtue of the following procedures: 1) finding the density of the correlation between an arbitrary number of ranked features; 2) finding the density between the results of ranking of the two experts; 3) evaluating the consistency of expert conclusions in a group of more than two experts.

In order to solve the first problem, as a rule, the Spearman's rank correlation coefficient is used [5, 6]. In order to estimate the proximity of the conclusions of two experts it is advisable to use the Kendall rank correlation coefficient [6]. For said purpose, evaluations of all possible pairs of any indicators are considered and their consistency is determined. In order to evaluate the consistency of expert opinions in a group of more than two experts, which is typical of our case, Kendall's coefficient of concordance [6] is most often used, which is calculated using the following formulas:

$$K_K = 12P_C/(L^2 m\left(m^2-1\right)), \tag{21}$$

$$P_C = \sum_{i=1}^{m}\left(\sum_{j=1}^{L} r_{ij} - \frac{L\left(m+1\right)}{2}\right)^2, \tag{22}$$

where L is the number of experts, m is the number of evaluated parameters, r_{ij} is the rank of the i-th element, assigned by the j-th expert.

The evaluation of competence is carried out with the help of a control examination, on the assumption that the correct answers to the questions are not known in advance. The mechanism is based on the processing of normalized scoring. The essence of the calculation is as follows:

1. The number of experts L_j is determined, who take part in the examination and must rank the indicators by means of their evaluation y_{ij}, for example, using a 10-point scale.

2. The amount of scores is calculated, determined by each expert on all indicators:

$$S_j = \sum_{i=1}^{n} y_{ij}, \tag{23}$$

3. A table of normalized scores for each expert is calculated, by dividing the points of each indicator by the expert's score:

$$\overline{y}_{ij} = y_{ij}/S_j, \tag{24}$$

4. The weighted sums of relative scores for each expert are calculated:

$$\overline{S}_j = \sum_{i=1}^{n} \overline{y}_{ij}\left(\sum_{i=1}^{n} \overline{y}_i \middle/ n\right), \tag{25}$$

5. The sum of the obtained weighted evaluations is calculated:

$$S = \sum_{j=1}^{L} \overline{S}_j, \tag{26}$$

6. The coefficients of expert competence are determined by dividing the weighted sum of the relative points of the expert into the total sum of the weighted evaluations:

$$K_j = \overline{S}_j / S, \tag{27}$$

7. The average group competence of experts is calculated:

$$K_{avg} = \sum_{j=1}^{L} K_j \Bigg/ L, \tag{28}$$

Experts, whose significance of their competencies is closest to the average group competence, are considered to be the most competent, and then the evaluations of only the most competent experts are taken into account. Thus, expert scoring evaluation of the i-th indicator on the basis of the joint use of the medians clustering and the method of scoring evaluations will be defined as:

$$Y_i = (M_i + y_i)/2, \tag{29}$$

Then the weight of the i-th indicator evaluated by N most competent experts will be equal to:

$$c_i = Y_i \Bigg/ \sum_{i=1}^{L} Y_i, \tag{30}$$

and the weight of the i-th factor, according to expert evaluations, will be determined as the sum of the weight of each indicator included in this factor:

$$q_i = \sum_{j=1}^{m} c_{ij}, \tag{31}$$

where m is the number of indicators included in the i-th factor. As a result, we obtain a set of dimensionless coefficients q_i, $i = 1, 2, \ldots, n$ (n is a number of factors).

It should also be noted that the proposed mechanism of expert evaluation makes it possible to implement program realization of the expert-statistical procedure for determining the weighting coefficients of factors. The generalized weight of factors that takes into account both the weight of the factor, determined on the basis of expert evaluations, and the weight of the factor determined statistically, can be obtained as the weighted average of these two evaluations [7]:

$$w_i = (\overline{q_i} + \overline{\nu_i}) \Bigg/ \sum_{i=1}^{n} (\overline{q_i} + \overline{\nu_i}), \tag{32}$$

where $\overline{q_i} = q_i \Big/ \sum\limits_i^n q_i$, $\overline{\nu_i} = \nu_i \Big/ \sum\limits_i^n \nu_i$ are expert and statistical (factor analysis) weighted coefficients of the factor, respectively. Thus, the integral indicator is calculated as the sum of factors with the corresponding weighted average weighting coefficients w_i:

$$I_j = \sum_{i=1}^{n} w_i F_{ij} \ (j = 1, 2, ..., n), \tag{33}$$

where n is a number of factors; F_{ij} is the value of the i-th factor for the j-th object. The best is an object with a larger value of the integral indicator.

2.2 Computation algorithm of the socio-economic development level

One of the main aspects of developed and existing models is ensuring the possibility of the processing automatization of socio-economic information on the basis of modern computer facilities. The presented model of determination of socio-economic development integral indicators formalizes the settlement procedures and makes it possible to develop an algorithm for automated data processing of socio-economic research, based on joint use of methods of expert evaluation and factor analysis. Figure 1 illustrates a general scheme of the developed algorithm for determining the socio-economic development integral indicator.

The initial stage of the algorithm is characterized by the entering of values of indicators of socio-economic development and expert evaluations. Thus, the initial data base is formed in the form of the matrix of indicators Θ (1) and the matrix of expert evaluations of indicators. Data from statistical directories can be used as indicators. Further actions within the framework of the presented algorithm are related to realization of the principal component analysis and mechanisms of expert evaluation (Fig. 1). According to the principal component analysis, a matrix of indicators of socio-economic development is initially formed, followed by its reduction to a single scale of measurements. Then, bringing the indicators to the normal distribution law and calculating the matrix of pairwise correlations are carried out. For this matrix, its eigenvalues and eigenvectors are calculated. The following actions are associated with the multiplication of the normalized matrix of indicators and the matrix of eigenvectors, which results in a matrix of factors. Factors are normalized, the dispersion is determined for them. Further it can be used in the analysis of integral indicators.

The next stage of the developed algorithm for automated determination of socio-economic development integral indicators is the procedure for determining the number of N factors included in the integral indicator (Fig. 1) on the basis of a series of eigenvalues of the matrix of pairwise correlations of socio-economic indicators and the given boundary value L due to dispersion factors of normalized parameters. The contribution of factors in the description of the total dispersion of the entire set of n socio-economic indicators is compared with the given limit value L of the dispersion of the normalized parameters, with achievement of

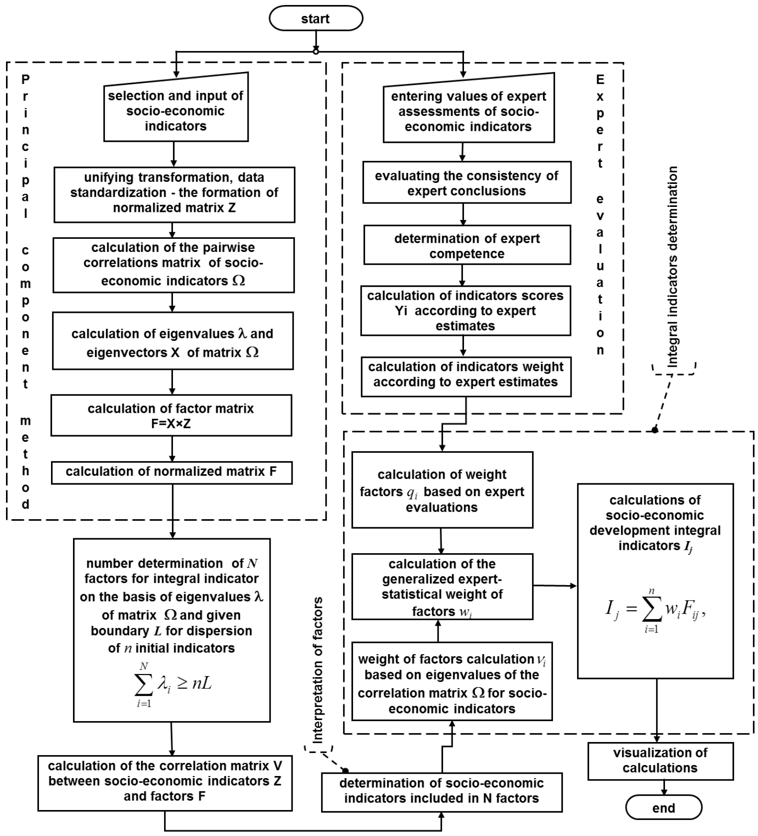

Fig. 1. Scheme of the automated algorithm for determination of socio-economic development integral indicator.

which the factorization is stopped by the determination of N factors, or in other words - a sampling of the minimum number of factors with maximal eigenvalues λ_i is made, the sum values of which are not less than nL:

$$\sum_{i=1}^{N} \lambda_i \geq nL \tag{34}$$

It is also worth noting that for analysis, one should use such number of factors that exhaust at least $60-70\%$ of the dispersion of the initial random variables [4], therefore the procedure described for determining the factors, that are included in the integral indicator by specification of the boundary value of L due to the dispersion factors of normalized socio-economic indicators, provides implementation of the mechanism of reducing the space of features without significant loss of informativeness, because N factors include the most important socio-economic indicators. The relative contribution $\%(F_i)$ of each of the N factors in the description of the total dispersion of all n indicators is determined as the ratio of the eigenvalue λ_i of the factor F_i to the total dispersion of the features, which is also equal to n:

$$\%(F_i) = \lambda_i \Big/ \sum \lambda_i = \lambda_i / n \tag{35}$$

The parallel branch of the algorithm (see Fig. 1) is associated with the program implementation of the expert evaluation mechanisms. For said purpose, initially, using the above procedure, the determination of the competence of experts is carried out, and then, by calculating Kendall's coefficient of concordance, one gets an evaluation of the consistency of their conclusions. Thus, as a result of the calculations, only the agreed conclusions of the competent experts remain. On the basis of these data, a group of experts is determined, the conclusions of which will participate in the evaluation. Using experts' performance of the ranking of indicators on the basis of the medians clustering and the method of arithmetic mean value, taking into account the competence of experts, the scoring evaluation of the indicators is calculated. Then, the weight of the indicators is determined according to expert evaluations [2].

The final stage of the algorithm is the determination of weighting coefficients, the calculation of integral indicators and the visualization of the results of data processing. Weighting coefficients for each factor are calculated by a combination of expert and statistical weight. The statistical weighting factors of the factors included in the integral indicator are determined by the formula (20), based on the eigenvalues of the pairwise correlations matrix of normalized socio-economic indicators. The expert weight of the factors included in the integral indicator is calculated by the formula (31). The generalized weight of the factors w_i, which takes into account both the weight of the factor, determined on the basis of expert estimates, and the weight of the factor determined statistically, we obtain by the formula (32). In order to directly determine the integral indicators, it is necessary to combine the calculated factors into a single indicator. Since all factors are independent, the combination is carried out using a simple linear convolution [4]. Thus, the integral indicator is calculated as the sum of factors with the corresponding weighted average weighting coefficients w_i, by the formula (33).

2.3 Modeling the process of regional socio-economic development assessing

Consider the process of assessing the level of regional socio-economic development in accordance with the developed computation algorithm for integral indicators of socio-economic development (Fig. 1) on the example of the Vinnytsia region districts.

According to the National State Statistics Service of Ukraine [8], one of the main socio-economic indicators that characterize the level regional development are (Table 1): Number of cars per 1000 people (P1); Services rendered per unit

Table 1. Normalized values of socio-economic indicators.

Districts	Socio-economic indicators										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Barsky	-0,190	0,028	0,225	-0,130	0,086	-0,235	0,044	-0,038	-0,154	0,052	-0,310
Bershadsky	-0,055	0,087	-0,070	0,182	0,006	0,294	0,003	0,034	-0,175	-0,246	-0,302
Vinnitsky	0,054	0,063	0,441	0,294	0,259	-0,083	0,951	-0,396	-0,325	-0,466	-0,236
Haysinsky	-0,063	0,430	0,067	0,169	0,230	0,013	0,014	-0,135	-0,188	-0,292	-0,328
Zhmerinsky	0,464	-0,136	-0,110	0,194	0,014	0,061	-0,018	-0,092	0,074	-0,227	0,126
Illinetsky	0,047	-0,144	0,185	-0,130	0,261	-0,195	-0,044	0,224	0,117	0,084	-0,065
Kalinovsky	-0,094	0,135	0,392	0,169	-0,159	-0,291	0,037	-0,271	-0,182	-0,174	-0,313
Koziatynsky	0,058	0,508	-0,041	0,157	0,036	0,254	-0,024	-0,207	-0,017	-0,179	0,001
Kryzhopol-sky	-0,022	0,283	0,067	-0,193	0,363	-0,115	-0,078	0,212	-0,041	0,097	-0,092
Lipovetsky	-0,152	0,016	0,205	0,119	-0,204	0,069	-0,021	-0,089	-0,080	0,063	-0,177
Litinsky	-0,109	-0,005	0,067	0,032	-0,243	0,053	-0,024	-0,142	0,158	0,054	0,084
Mohyliv-Podilsky	0,470	-0,110	-0,110	0,107	-0,046	-0,251	-0,070	0,046	0,197	-0,080	0,205
Murovanna-rurilotsky	-0,174	-0,217	-0,375	-0,205	-0,116	0,102	-0,087	0,312	0,223	0,185	0,246
Nemirovsky	-0,112	0,249	-0,129	0,182	0,335	0,053	0,002	-0,123	0,048	-0,155	-0,073
Orativsky	0,094	-0,223	0,038	-0,380	-0,204	0,190	-0,089	0,297	0,510	0,288	0,177
Pischancky	-0,152	-0,094	-0,119	-0,230	0,076	-0,147	-0,063	0,106	-0,018	0,237	0,308
Pogrebish-chensky	0,009	-0,015	-0,149	-0,230	-0,152	-0,067	-0,062	-0,166	0,276	0,142	0,086
Teplicky	-0,171	-0,121	-0,169	0,057	-0,269	0,118	-0,052	0,498	0,152	0,145	0,202
Tyvrivsky	0,053	0,009	0,038	0,157	0,159	-0,195	0,118	-0,094	-0,083	-0,013	0,098
Tomashpil-sky	-0,149	-0,101	0,067	-0,168	0,276	-0,011	-0,042	0,145	-0,148	0,083	-0,119
Trostyanet-sky	0,276	-0,074	0,097	-0,068	-0,113	0,126	-0,076	0,056	-0,189	0,084	-0,083
Tulchinsky	0,098	0,189	-0,080	0,169	0,093	-0,139	-0,042	-0,145	-0,176	-0,141	-0,047
Khmelnyt-sky	0,387	-0,294	-0,237	0,206	-0,033	0,286	-0,078	-0,004	0,319	-0,273	-0,005
Chernivtsky	-0,164	-0,110	-0,296	-0,417	-0,055	0,454	-0,069	-0,154	0,071	0,215	0,271
Chethelnit-sky	-0,109	-0,183	-0,149	-0,105	-0,334	0,118	-0,079	0,024	-0,043	0,242	0,287
Shargorod-sky	-0,249	-0,167	0,244	0,132	-0,130	-0,283	-0,075	0,077	-0,091	0,121	-0,045
Yampilsky	-0,046	-0,004	-0,100	-0,068	-0,135	-0,179	-0,075	0,025	-0,236	0,154	0,103

of population, UAH (P2); Natural increase (reduction) of the population (P3); Registered unemployment rate (P4); Average monthly salary, UAH (P5); Provision of housing by the population, m^2 per person (P6); The ratio of m^2 of built housing to the population (P7); Preschool establishments per unit of population

(P8); General educational institutions per unit of population (P9); Number of crimes per 1000 people (P10); Emissions of pollutants (P11). It should be noted that the list of indicators, depending on the goals and objectives of the assessing, may change, thereby changing its emphasis. Thus, for the Vinnytsia region we have a matrix of initial socio-economic indicators in the size of 27×11 (27 districts of the region and 11 indicators). Listed in a single scale of measurements and normalized (2) values of socio-economic indicators of districts are presented in Table 1. On the basis of the normalized matrix of socio-economic indicators (Table 1), the pairwise correlations matrix $\mathbf{\Omega}$ of indicators is dimensioned 11×11. For the pairwise correlations matrix of indicators, we determine eigenvalues λ (Table 2) and eigenvektors $\mathbf{X}$.

The matrix of factors F is obtained by multiplying the normalized matrix of socio-economic indicators (Table 1) into the matrix of the eigenvektors of the pairwise correlations matrix. The obtained factors are normalized by the formula (2). The normalized factor matrix is used to calculate the matrix of correlations between factors and indicators of socio-economic development, that is required for the interpretation of factors.

On the basis of the calculated eigenvalues of the pairwise correlations matrix $\mathbf{\Omega}$ (Table 2) and the given threshold L of the dispersion for normalized socio-economic indicators (Table 1), the formula (34) determines the number of N factors in the integral indicator. In this case, the number of main components (factors) must be used, which exhaust at least 60-70% of the variance of the initial random variables. For example, at a given threshold of 0.6, from Table 2 it is necessary to select N factors with maximal eigenvalues, the sum of values of which is not less than $0,6 \times 11 = 6,6$. The sum of the first three eigenvalues λ is 7.49, that is, the integral index consists of the first three factors ($N = 3$) that explain approximately 68% (see formula 35) of the variance of the initial data (Table 2). The calculate matrix of correlations $\mathbf{V}$ between the normalized socio-economic indicators and the factors shows, which indicators are included in the given three factors (with the value of the variance of the indicators should not be less than the given limit value of 0.6).

Table 3 shows the structure of factors: the coefficient of correlation between the indicators and factors in which they are included, statistical (20) and expert (31) weights coefficients and weighted average weight coefficient of factors (32). The first factor included the first four socio-economic indicators - 1) the number of cars per 1000 people; 2) services rendered per unit of population; 3) natural increase (reduction) of the population; 4) the level of registered unemployment. The second factor included the eleventh indicator - emissions of pollutants. The third factor entered the seventh indicator - the ratio m_2 of the built housing to the population.

To calculate the integral indicators, it is necessary to implement the mechanism for determining weight factor factors using expert evaluation. The basis for calculations is a table with score points of socio-economic indicators, put forward by experts. As experts participating in the evaluation, employees of the Regional Economic Development Department of Vinnytsia Region State Admin-

Table 2. Eigenvalues of the pairwise correlation matrix of indicators.

λ	4,566	1,334	1,193	0,931	0,765	0,587	0,495	0,364	0,244	0,098	0,022
$\%(F_i)$	41,51%	12,12%	10,85%	8,46%	6,95%	5,34%	4,50%	3,31%	2,22%	0,89%	0,20%
$\sum\%$	41,5%	53,6%	64,4%	72,9%	79,9%	85,2%	89,7%	93,1%	95,2%	96,2%	96,4%

Table 3. Structure of factors and their weight coefficients.

<table>
<tr><th>Factors</th><th>Socio-economic indicators</th><th>Correlation coefficient</th><th>Statistical weight coefficient</th><th colspan="2">Experts weight coefficient</th><th>Weighted average weight coefficient</th></tr>
<tr><td rowspan="4">F1</td><td>P1</td><td>0,805</td><td rowspan="4">2,137</td><td>0,016</td><td rowspan="4">0,105</td><td rowspan="4">0,370</td></tr>
<tr><td>P2</td><td>0,814</td><td>0,015</td></tr>
<tr><td>P3</td><td>0,720</td><td>0,021</td></tr>
<tr><td>P4</td><td>0,680</td><td>0,053</td></tr>
<tr><td>F2</td><td>P11</td><td>0,788</td><td>1,317</td><td colspan="2">0,034</td><td>0,188</td></tr>
<tr><td>F3</td><td>P7</td><td>0,690</td><td>1,092</td><td colspan="2">0,250</td><td>0,442</td></tr>
</table>

istration were used in this study. In order to minimize the subjectivity of expert assessments, an evaluation of the consensus of the experts' conclusions in the group and the definition of experts competencies is initially carried out. To do this, the Kendal Concordance Coefficient (KK) is used. Calculation is carried out by formulas (21) and (22). The result of calculating the coefficient Kendel shows that it is close to the unit KK = 0,815. That is, it can be concluded that expert assessments are consistent and the composition of the expert group need not be changed. To assess the competencies of experts, an approach based on the processing of normalized ball scores is used. First, on the basis of scores of experts, the amount of points calculated by a particular expert is calculated (23). Then, by dividing each ball into the sum of all the scores arranged by this esperty (24), we obtain a normalized ball scores of each expert. The average normalized value is calculated for each indicator. On the basis of the obtained results, by formula (25), we calculated the weighted sum of the relative points of each expert. In Table 4 the results of the weighted sum of expert assessments calculations are presented. Also, the sum of the weighted evaluations obtained by experts is (2.3463). Subsequently, by dividing the weighted sum of expert points by the sum of the weighted assessments of all experts, the competencies of each expert (27) and the average group competence of experts (28) are calculated. Analyzing the competencies of experts (Table 4), it is evident that 5 experts have competences that are closest to the group's core competencies. That is, the most competent experts are 1, 4, 10, 11 and 20 experts. It is the ballroom evaluation of the indicators by these experts (Table 4) will be used in the future for calculations.

Table 4. Weighted sums of relative ball scores and competence coefficients of experts.

Experts	**Weighted sum of relative points**	**Competence coefficient** β		Δ**=Average -** β
Expert 1	0,1174	0,050036427		0,03643
Expert 2	0,1192	0,050809829		0,80983
Expert 3	0,1209	0,051535191		1,53519
Expert 4	0,1168	0,049768327		0,23167
Expert 5	0,1158	0,049374578		0,62542
Expert 6	0,1135	0,048393521		1,60648
Expert 7	0,1211	0,051600554		1,60055
Expert 8	0,1206	0,051397059		1,39706
Expert 9	0,1137	0,048441221		1,55878
Expert 10	0,1166	0,049692931		0,30707
Expert 11	0,1159	0,049407296		0,59270
Expert 12	0,1142	0,048666712		1,33329
Expert 13	0,1195	0,050940928		0,94093
Expert 14	0,1211	0,051628416		1,62842
Expert 15	0,1206	0,051400438		1,40044
Expert 16	0,1157	0,049319329		0,68067
Expert 17	0,1210	0,051574862		1,57486
Expert 18	0,1137	0,048445995		1,55401
Expert 19	0,1129	0,048119791		1,88021
Expert 20	0,1160	0,049446595		0,55341
Summ	**2,3463**	**Average**	**0,0500**	

Table 5. Group expert scores for indicators and and their weighting coefficients.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Average	0,062	0,053	0,074	0,168	0,164	0,065	0,053	0,066	0,070	0,053	0,172
Median	0,063	0,058	0,077	0,173	0,154	0,063	0,058	0,063	0,064	0,058	0,170
Y_i	0,062	0,055	0,075	0,170	0,159	0,064	0,055	0,064	0,067	0,055	0,171
C_i	0,016	0,015	0,021	0,053	0,063	0,034	0,034	0,045	0,057	0,061	0,250

Experts estimate the medians and the average values of ball scores for indicators. Based on the median and the mean values of the ball scores, the formula (29) determines the group expert scores for each indicator Y_i (Table 5). The expert weight of each indicator C_i is determined by the formula (30), their values are given in Table 5. The weight coefficient q_i of the factor is estimated by the experts (31), as the sum of the expert weight of each indicator included in this factor (see Table 3). The statistical weight coefficient ν_i of factor is determined on the basis of the eigenvalues of the pairwise correlations matrix $\mathbf{\Omega}$ of normalized socio-economic indicators (Table 2) by the formula (20). The generalized weight of the factor w_i by the formula (32).

Table 6. Calculated factors and integral indicators of districts of Vinnytsia region.

Districts	F3	F2	F1	Integral indicators
Barsky	0,016	-0,079	-0,100	-0,164
Bershadsky	-0,116	0,028	-0,099	-0,187
Vinnitsky	0,119	0,019	-0,422	-0,284
Haysinsky	-0,153	0,004	-0,224	-0,374
Zhmerinsky	0,087	0,103	-0,012	0,179
Illinetsky	0,083	-0,038	0,021	0,066
Kalinovsky	0,079	-0,041	-0,213	-0,175
Koziatynsky	-0,190	0,053	-0,106	-0,243
Kryzhopolsky	-0,100	-0,049	-0,026	-0,175
Lipovetsky	0,004	-0,029	-0,050	-0,074
Litinsky	0,016	0,002	0,041	0,059
Mohyliv-Podilsky	0,164	0,070	0,050	0,284
Murovannarurilotsky	-0,026	-0,006	0,248	0,216
Nemirovsky	-0,144	0,023	-0,101	-0,222
Orativsky	0,061	0,001	0,272	0,334
Pischancky	0,004	-0,047	0,131	0,088
Pogrebishchensky	0,002	0,006	0,111	0,119
Teplicky	0,020	-0,010	0,193	0,203
Tyvrivsky	0,058	0,002	-0,076	-0,017
Tomashpilsky	-0,038	-0,056	-0,002	-0,096
Trostyanetsky	0,037	0,009	0,013	0,059
Tulchinsky	-0,032	0,020	-0,112	-0,124
Khmelnytsky	0,053	0,129	0,063	0,245
Chernivtsky	-0,167	0,008	0,203	0,044
Chethelnitsky	0,018	-0,012	0,171	0,177
Shargorodsky	0,123	-0,072	-0,015	0,037
Yampilsky	0,021	-0,037	0,040	0,023

To calculate the integral index I_j, it is necessary to combine the calculated factors. Calculations made show (Table 3) that weighted average weight coefficients of factors are: 0,3697, 0,1881 and 0,4422 for 1st, 2nd and 3rd factors, respectively. By multiplying the obtained factors by the corresponding weighted average weight coefficients of factors, by formula (33) we obtain the values of integral indicators that allow to rank the regions in terms of their socio-economic status. In table 6 results of calculations of factors and integral indicators of social and economic development of districts of Vinnytsia region are presented.

3 Conclusion

In the presented algorithm of evaluation of the socio-economic development level, the increase of its reliability is not due to a growing number of sources of initial data on the basis of which the factors are determined, but due to the implementation of expert evaluation procedure in the mechanism of determining weighting coefficients of factors. Thus, taking into account the knowledge and experience of experts in determining the weighting coefficients of factors, we introduce the importance of specific indicators in the factor model of the evaluation of the level of socio-economic development, or, in other words, the intensity of their influence on the state of the socio-economic system. The obtained correlation dependencies can be used, for example, to detect correlations between the indicators and features that determine the socio-economic development (regression) of individual regions, etc. The main advantages of the developed algorithm for determining the integral indicators are: the use of the whole set of initial data, which excludes the possibility of distorting the content of the socio-economic model; ensuring the possibility of operative work with large socio-economic data bulk; taking into account knowledge and experience of experts in building a single socio-economic development integral indicator. The proposed algorithm for the determination of integral indicators makes it possible to implement a unified approach to data analysis and to ensure the efficiency of constructing integral indicators. It should also be noted that in the context of processing automatization of socio-economic data, the expert-statistical algorithm proposed provides the possibility of program implementation of the procedure for determining the socio-economic development integral indicators. The developed computation algorithm is used in the teaching of the discipline "Prognostics of socio-economic processes" in conducting a laboratory workshop on the topic "Determining the level of socio-economic development on the basis of expert-statistical method" and "Modeling the influence of the socio-economic indicators values on the general level of regional socio-economic development".

References

1. Theodore, W.: An Introduction to Multivariate Statistical Analysis. 3rd ed. Wiley-Interscience, Boston (2003)
2. Tacq, J.: Multivariate Analysis Techniques in Social Science Research: From Problem to Analysis. 1st ed. SAGE Publications Inc., Thousand Oaks (1997)
3. Dunteman, G.H.: Principal Components Analysis (Quantitative Applications in the Social Sciences). 1st ed. SAGE Publications Inc., Thousand Oaks (1989)
4. Zhukovskaya, V.M., Muchnik, I.B.: Factor Analysis in Socio-Economic Research. Statistic, Moscow (1976) (in Russian)
5. Beshelev, S.D. Gurvich, F.G.: Mathematical and Statistical Methods of Expert Estimates. Statistic, Moscow (1980) (in Russian)
6. Goldstein, H., Lewis, T.: Assessment: Problems, Developments and Statistical Issues. 1st ed. Wiley-Blackwell, Boston (1996)
7. Ustinovieius, L.: Determining integrated weights of attributes. J. Civ. Eng. Manag. **7**(4), 321–326 (2001). doi: 10.3846/13921525.2001.10531743
8. Verner, I.Yev. (ed.): Statistical Yearbook of Ukraine for 2017. State Statistics Service of Ukraine, Kyiv (2017)