DOI: 10.17323/2587-814X.2022.3.7.23

# Construction of an aggregated production function with implementation based on the example of the regions of the Central Federal District of the Russian Federation*

**Roman A. Zhukov**
E-mail: pluszh@mail.ru

**Nadezhda A. Kozlova**
E-mail: 95kno@mail.ru

**Evgeny V. Manokhin**
E-mail: emanfinun@gmail.com

**Maria A. Plinskaya**
E-mail: maria.plinskaya@gmail.com

Financial University under the Government of the Russian Federation, Tula Branch
Address: 1a, Oruzheynaya Street, Tula 300012, Russia

**Abstract**

A three-dimensional case is considered on the basis of a method developed for estimating the parameters of the aggregated production function used to calculate dynamic standards and build integral indicators of the performances of the functioning of socio-economic systems. The aggregated production function is determined by the quadratic convolution of the production functions of the results of the functioning of the elements of the subsystem and their correlation matrix. The parameters of the aggregated production function are determined from solving the problem of maximizing the likelihood function of a random variable – the residuals of production

* The article is published with the support of the HSE University Partnership Programme

functions aggregated according to a similar rule. On the example of a project subsystem within the framework of the Kleiner's spatial-temporal classification of socio-economic systems we obtained adjusted values of the parameters of a function that includes power-law multiplicative models of the relationship between the volume of gross domestic product by region for sections F (construction), G (wholesale and retail trade), K (financial activity) according to NACE 2 and the cost of fixed assets (total for section K, for sections F and G), the average annual number of employees (for sections F and G) and the average annual population (for section K), based on data for 2015–2020 (sections G, K) and 2018–2020 (section F) for the regions of the Central Federal District. The EFRA software package and Python's project were used as tools. The results obtained can be used by regional authorities in assessing the functioning of the regions and the formation of appropriate standards in the short term.

**Keywords:** socio-economic system, probability distribution density, aggregated production function, model, integral estimation

**Citation:** Zhukov R.A., Kozlova N.A., Manokhin E.V., Plinskaya M.A. (2022) Construction of an aggregated production function with implementation based on the example of the regions of the Central Federal District of the Russian Federation. *Business Informatics*, vol. 16, no. 3, pp. 7–23. DOI: 10.17323/2587-814X.2022.3.7.23

## Introduction

When assessing the results of the functioning of regional socio-economic systems (RSES), both private and integral indicators are used, the latter of which are reduced to aggregating the effective features of SES elements, including in the tasks of identifying significant factors, modeling and forecasting socio-economic processes [1, 2] in the context of sustainable, balanced development [3]. In most cases, aggregation is carried out in the form of averages and weighted averages of various types [4–7], or using generally recognized analysis of the functioning environment [8, 9], component analysis [10, 11]) and special algorithms for constructing integral indicators [12, 13] with an analysis of the correctness of their construction [14]. The volume of gross domestic product by region for the corresponding types of economic activity within the sectoral [15] or spatial-temporal [16] classifications of socio-economic subsystems is most often used as private indicators for assessing the functioning of the RSES, which are part of the integral indicator, to assess the balance of the functioning of the regions – subjects of the Russian Federation. According to the spatial-temporal classification of Kleiner, SES includes subsystems of four types: object (limited in space, not limited in time), environment (not limited in space or time), process (not limited in space, limited in time), project (limited in space and time). The regional project subsystem can be represented as three elements, each of which is a set of economic units – institutional units – residents of the region (in the terminology of the system of national accounts) contributing to the volume of gross domestic product by region (GDP by region) in sections F (construction), G (wholesale and retail trade), K (financial activity) in accordance with the All-Russian classifier of types of economic activity (OKVED, NACE). The second edition of OKVED 2 (NACE rev.2) has been used since 2017; previously OKVED 1 (NACE rev. 1) was used.

In order to model and predict the values of particular indicators – the results of the functioning of the elements of the RSES – we use models of the relationship of productive and factorial features in the form of economic and statistical models – production functions (PF) that establish the relationship between output volume and production factors, including linear [17, 18], quadratic [19], logarithmic [20], translogarithmic, constructed on the basis of the Cobb-Douglas function [21], transcendental [22], power multiplicative (most often found in publications), including taking into account the innovative component [23] etc. The choice of the functional form of models is most often determined by the researcher based on verification of a set of statistical hypotheses. It is also possible to introduce additional selection criteria related, for example, to the qualitative content of the model and the priorities of the RSES control centers if the models are used to set norms [24]. To estimate their parameters, conventional (OLS) and generalized least squares (GLS) and maximum likelihood (MLE) methods are traditionally used.

For an SES subsystem characterized by a set of models, its integral evaluation requires aggregation of the corresponding models of the functioning of the elements, in the simplest case determined by their simple or weighted average sum. However, due to the presence of interrelations between the elements, the procedure for searching for the parameters of the model that characterizes the result of the functioning of the subsystem as a whole, the parameters of the aggregated production function (APF) are already becoming unobvious.

The author's work [25] presents a method for estimating the parameters of the APF tested on the example of a two-component APF. In this article, we aim to apply the method for the three-dimensional case when the aggregated production function is determined by a quadratic convolution of three production functions, each of which characterizes the results of the functioning of an element of a three-element subsystem of the SES. Using the example of evaluating the results of the functioning of the project subsystems of the regions of the Central Federal District, we test the hypothesis about the possibility of using the method to refine the parameters of an aggregated production function based on the EFRA software package [26], as well as a software project specially developed in Python.

## 1. Methodology for evaluating the results of the subsystem functioning and constructing an aggregated production function

To evaluate the results of the functioning of the SES subsystem, we propose to use an integral indicator with the properties of monotony, identity, commensurability, dimensionlessness and transitivity, and also to take into account the relationships between the elements [27]:

$$\xi_{k,s_q}(t) = \frac{\sqrt{\sum_{i_1=1}^{I}\sum_{i_2=1}^{I} r_{i_1,i_2,s_q} \cdot y^0_{i_1,k,s_q}(t) \cdot y^0_{i_2,k,s_q}(t)}}{\sqrt{\sum_{i_1=1}^{I}\sum_{i_2=1}^{I} \widehat{r}_{i_1,i_2,s_q} \cdot \widehat{y}^0_{i_1,k,s_q}(t) \cdot \widehat{y}^0_{i_1,k,s_q}(t)}}, \quad (1)$$

where $r_{i_1,i_2,s_q}$, $\widehat{r}_{i_1,i_2,s_q}$ are corresponding values of the paired correlation coefficient between $i_1$-th $y^0_{i_1,s_q}$, $\widehat{y}^0_{i_1,s_q}$ and $i_2$-th $y^0_{i_2,s_q}$, $\widehat{y}^0_{i_2,s_q}$ variables (resultative features, respectively, actual and expected (normative), the values of the latter in the time period t, are determined using the production function (PF)) ($i_1$, $i_2 = 1, ..., I$, $I$ is the number of resultative features of a $k$-subsystem of type $s_q$); the index "0" shows that the values of variables are reduced to a scale from 0 to 1 by converting standardized (centered and normalized) values of absolute values:

$$(\cdot)^0_{i,k,s_q}(t) = \quad (2)$$

$$= \frac{(\cdot)^*_{i,k,s_q}(t) - \min\{y^*_{i,k,s_q}(t), \widehat{y}^*_{i,k,s_q}(t)\}}{\max\{y^*_{i,k,s_q}(t), \widehat{y}^*_{i,k,s_q}(t)\} - \min\{y^*_{i,k,s_q}(t), \widehat{y}^*_{i,k,s_q}(t)\}}.$$

Here $(\cdot)$ is $y_i$, $\widehat{y}_i$; $*$ means that the variables (we will consider them random variables) are centered and normalized:

$$(\cdot)^*_{k,s_q}(t) = \frac{(\cdot)_{k,s_q}(t) - M(y_{s_q})}{\sigma(y_{s_q})}, \qquad (3)$$

where $M(y_{s_q})$, $\sigma(y_{s_q})$ are the mean and standard deviation of the combined $k$ and $t$ samples.

If the value of the indicator is greater than or equal to one, then the functioning of the subsystem can be considered satisfactory. Similarly in formula (1), particular performance indicators are constructed determined by the ratio of actual and normative values (calculated by the production function (PF)) reduced to a scale from 0 to 1 in accordance with formulas (2) and (3).

The expression standing in the denominator (1) is an aggregated production function (APF) formed by a quadratic convolution of the production functions of the productive features $i_1$, $i_2$ and the corresponding correlation matrix.

The relationship between the values of effective features and factors can be represented as [25]:

$$y_{k,i}(t) = f_i(C_{i,j}, x_{k,i,j}(t)) + \varepsilon_{k,i}(t). \qquad (4)$$

where $k$ is the number of population elements, ($k = 1, \ldots, K \in N$);

$t$ is the observation time of the $k$-th of the population element ($t = 1, \ldots, T \in N$);

$i$ is the index of a random variable ($i = 1, \ldots, m \in N$);

$C_{i,j}$ are parameters of the function $f_i(\cdot) = \widehat{y}_i$;

$\varepsilon_{k,i}$ are the values of the stochastic random component $\varepsilon_i \sim N(0; \sigma^2_{\varepsilon_i})$:

$$f_{P,i}(\varepsilon_i) = \frac{1}{\sqrt{2\cdot\pi}\cdot\sigma_{\varepsilon,i}} \cdot \exp[-\frac{\varepsilon_i^2}{2\cdot\sigma^2_{\varepsilon,i}}] = \frac{1}{\sqrt{2\cdot\pi}\cdot\sigma_{y_i}} \cdot \exp[-\frac{(y_i - \widehat{y}_i)^2}{2\cdot\sigma^2_{y_i}}]. \qquad (5)$$

In order to eliminate the influence of units of measurement, we will consider standardized random variables $\varepsilon_i^*$ with a joint probability distribution density:

$$f_P(\varepsilon_1^*, \varepsilon_2^*, \ldots, \varepsilon_m^*) = \frac{1}{(2\cdot\pi)^{m/2}\cdot\sqrt{\Delta_r}} \times \times \exp[-\frac{1}{2\cdot\Delta_r}\sum_{i=1}^{m}\sum_{j=1}^{m} A_{r_{ij}} \cdot \varepsilon_i^* \cdot \varepsilon_j^*], \qquad (6)$$

where $\Delta_r$, $A_{r_{ij}}$ are the determinant and algebraic complements of the correlation matrix $\|r_{ij}\|$, correspondently, the elements of which are paired correlation coefficients; $\varepsilon_i^* = (y_i^* - \widehat{y}_i^*)^2 / (2\cdot\sigma^2_{y_i^*})$.

Then the probability distribution density of a random variable $\varepsilon^*$, which are aggregated random variables, will be defined as:

$$f_P(\varepsilon^*) = \frac{1}{(2\cdot\pi)^{m/2}\cdot\sqrt{\Delta_r}} \times \times \frac{d}{dy^*} \int\int\ldots\int_D \exp[-\frac{1}{2\cdot\Delta_r}\sum_{i=1}^{m}\sum_{j=1}^{m} A_{r_{ij}} \cdot \varepsilon_i^* \cdot \varepsilon_j^*] dD, \qquad (7)$$

where the region of integration $D$ depends on the combination $\varepsilon_i^*$.

We will conside $D$ as a quadratic convolution in two variants:

$$(y - \widehat{y})^2 = \left( \sqrt{\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} r_{i_1,i_2} \cdot y^*_{i_1} \cdot y^*_{i_2}} - \sqrt{\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} \widehat{r}_{i_1,i_2} \cdot \widehat{y}^*_{i_1} \cdot \widehat{y}^*_{i_2}} \right)^2 \le (\varepsilon^*)^2, \qquad (8)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m} r_{ij} \cdot \varepsilon_i^* \cdot \varepsilon_j^* \le (\varepsilon^*)^2. \qquad (9)$$

The formula (8) corresponds to the difference between the numerator and the denominator of the expression (1) used to calculate the integral performance indicator.

To simplify calculations, it is necessary to bring the quadratic form to a canonical form by calculating eigenvalues and eigenvectors, or use the Lagrange method.

For the two-dimensional case, an analytical expression of the probability distribution density is obtained [25]. For the three-dimensional case, the density expression can be represented in quadratures in a spherical coordinate system [28]:

$$f_P(\varepsilon^*) = \frac{1}{(2\cdot\pi)^{3/2}\cdot\Delta}\cdot\int_0^{2\cdot\pi}\int_0^{\pi}(\varepsilon^*)^2\cdot\sin\theta\times$$
$$\times\exp[-\frac{1}{2}\cdot(\varepsilon^*)^2\times(c_{11}\cdot\cos^2\phi\cdot\sin^2\theta+$$
$$+c_{22}\cdot\sin^2\phi\cdot\sin^2\theta+c_{22}\cdot\cos^2\theta+$$
$$+c_{12}\cdot\sin2\phi\cdot\sin^2\theta+c_{13}\cdot\cos\phi\cdot\sin2\theta+$$
$$c_{23}\cdot\sin\phi\cdot\sin2\theta)]d\theta d\phi, \quad (10)$$

where $c_{11} = K_{11}$;

$$c_{22} = \frac{1}{K_{33}\cdot\Delta}\cdot(K_{11}\cdot r_{12}^2+K_{22}-2\cdot K_{12}\cdot r_{12});$$
$$c_{33} = \frac{1}{K_{33}}\cdot(r_{12}\cdot K_{23}+r_{13}\cdot K_{33})\cdot(K_{11}\cdot r_{12}\cdot K_{23}+$$
$$+K_{11}\cdot r_{13}\cdot K_{33}-2\cdot K_{12}\cdot K_{23}-2\cdot K_{13}\cdot K_{33})+$$
$$+K_{33}^2+2\cdot K_{23}^2+\frac{K_{22}\cdot K_{23}^2}{K_{33}};$$
$$c_{12} = \frac{1}{\sqrt{K_{33}\cdot\Delta}}\cdot(-K_{11}\cdot r_{12}+K_{12}); \quad (11)$$
$$c_{13} = \frac{1}{\sqrt{K_{33}}}\cdot(-K_{11}\cdot r_{12}\cdot K_{23}-K_{11}\cdot r_{13}\cdot K_{33}+$$
$$+K_{12}\cdot K_{23}+K_{13}\cdot K_{33});$$
$$c_{23} = \frac{1}{K_{33}\cdot\sqrt{\Delta}}\cdot(K_{11}\cdot r_{12}\cdot(r_{12}\cdot K_{23}+r_{13}\cdot K_{33})+K_{22}\cdot K_{23}-$$
$$-K_{12}\cdot(2\cdot r_{12}\cdot K_{23}+r_{13}\cdot K_{33})-$$
$$-r_{12}\cdot K_{13}\cdot K_{33}+K_{23}\cdot K_{33});$$

$K_{ij}^{-1} = A_{ij}/\Delta$ are elements of the inverse covariance matrix.

In the first case, the parameters of the aggregated production function $C^*_{i,j[part]}$ can be determined by methods of OLS or MLE applied for each of the considered PF $f_i(\cdot) = \hat{y}_i$. In the second case, the parameters of the APF are found using MLE for a density of the form (7) with a likelihood function $\ln L(y^*\left|C^*_{i,j}, x^*_{i,j}(t), \sigma_{y*}\right.)$:

$$\ln L(y^*\left|C^*_{i,j}, x^*_{i,j}(t), \sigma_{y*}\right.) =$$
$$= \sum_{k=1}^{K}\sum_{t=1}^{T} f(y^*\left|C^*_{i,j}, x^*_{i,j}(t), \sigma_{y*}\right.) \to \max. \quad (12)$$

The maximum (12) is determined by $C^*_{i,j}$ ((the parameters of the standardized *i*-th PF) with constraints of the form:

$$\sum_{k=1}^{K}\sum_{t=1}^{T}(\varepsilon_k^*(t))^2 \le \sum_{k=1}^{K}\sum_{t=1}^{T}(\varepsilon_k^*(t))^2_{[part]}, \quad (13)$$

where $\varepsilon_k^*(t)$ are the values of a random variable calculated for defined by (12);

[part] (partial) are the values of a random variable calculated usin $C^*_{i,j[part]}$.

The method presented makes it possible to refine the parameters of the APF by solving the optimization problem, with a joint search for the coefficients of the models which makes it possible to increase the reliability of estimates when constructing the normative values of the results of the functioning of the SES subsystem.

## 2. Conceptual scheme and algorithm of the method realization

To find the initial values of the model parameters, the EFRA software package is used. This allows you to find coefficients and conduct a number of statistical tests justifying the possibility of using models to develop standards or forecasts, including evaluating the significance of the model (according to the Fisher criterion), evaluating the significance of model parameters (Student's criterion), checking for the absence of heteroscedasticity (according to Spearman's rank correlation coefficient).

The conceptual scheme of the method for the three-dimensional case consists of five generalized blocks (*Fig. 1*) that implement the corresponding algorithm in a software project (module) in the open source programming language Python [29].

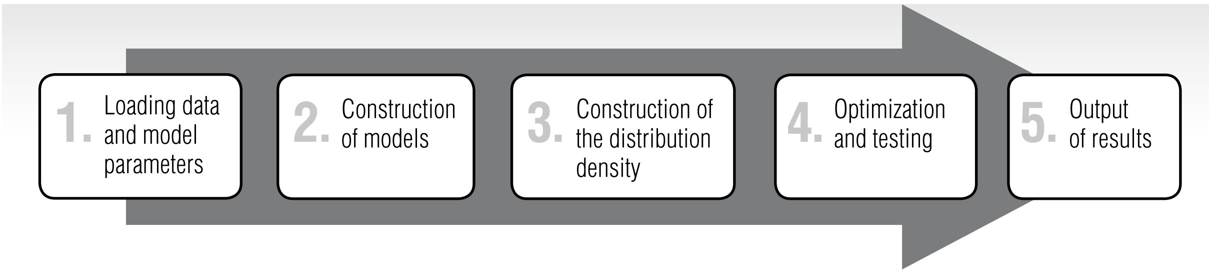



*Fig. 1.* Conceptual generalized scheme of the algorithm.

For the software module to work, it is necessary to connect additional Python libraries: numpy, pandas, scipy, matplotlib, opnpyxl and datetime.

Base file main.py loads additional modules and also contains the initial parameters of the models to run the corresponding blocks of the algorithm.

The first block provides loading of data and initial values of model parameters. A file has been developed for this purpose InputData.py containing the Input Data class, in which data is loaded from a specially generated statistical data file in the * format.xlsx with fields for element names, evaluation periods, and attributes. The class, like other project classes, contains the __init__() function containing data variables and the number of observations. The parameters of the initial and final evaluation period, the names of the effective and factor signs are passed to the data input function. The input parameters are sampled from the data file.

The second block provides the following functions included in the file Models.py with the Models class.

1. Formation of power multiplicative PF models according to (4) in absolute and logarithmic forms by adding the corresponding def functions to the program text with the transmitted values of the factors and coefficients of the models: def func_abs(*self, y, x1, x2, a0_3*), def func_std(*self, x1, x2, a0_3*), where *y* is the resultative feature, *x1*, *x2* are factor features, *a0_3* are coefficients of models.

2. Construction of aggregated random variables which are the residuals of the APF defined by the ratio (8) are 2 functions integr_std(s*elf, y, y_teor, m* = 3) and residuals(*self, y, y_teor*) and formula (9) is 1 function def integr_std_y_y_teor(*self, y, y_teor, m* = 3). Here $m = 3$ is the number of transmitted variables which are resultative features; *y, y_teor* are the actual and calculated values of productive features based on PF models presented as an array for the three-dimensional case (contains three variables). The functions, in addition to the values of the aggregated random variable, return the sum of the squares of the residuals and the correlation matrix $\|r_{ij}\|$ for the case when the values of the elements $\|r_{ij}\|$ are calculated at each iteration of solving the optimization problem using a set of variable parameters of the *a0_3* model corresponding to $C_{i,j}$.

3. The function res_test(s*elf, mu, sigma, size, m* = 3) generates a set of normally distributed random variables, with mean named as *mu*, standard deviation named as sigma and volume named as size.

The third block is represented by a file DistributionDensity.py containing the DistributionDensity class, the purpose of which is to calculate the distribution density of a three–component aggregated random variable. This module contains the following functions.

1. Integrand function def UnderIntFuncti

onQuadraqticForm(*self, theta1, theta2, r_3, z*), which corresponds to formulas (10) and (11) with angle variables *theta1, theta2* and *z* which are values of the aggregated random variable $\varepsilon^*$.

2. Integrand function def UnderIntFunctionLA(*self, theta 1, theta 2, r_3, z, m* = 3), which allows converting the integrand expression in formula (7) to canonical form by calculating eigenvalues and eigenvectors of the correlation matrix $\|r_{ij}\|$. The function can be extended to the case of *m* variables by means of the built-in double sum generation cycle in (7).

3. The function Density_3var(*self, bins, r_3, t* = 1) calculates the double integral in (7) for the values of bins named as $\varepsilon^*$. The variable *t*, which can be equal to 0 or 1, is responsible for choosing the integrand function UnderIntFunctionQuadraqticForm or UnderIntFunctionLA.

The fourth module is represented by two files: ObjectiveFunction.py and TestFunction.py .

The first file contains the ObjectiveFunction class, designed to form the objective function objective_func_SearchParameters_3var(*a*) and a system of constraints: a) upper and lower bounds for changing parameters *a* are parameters of the aggregated production function *lb* and *ub*, the values of which are calculated using "EFRA"; b) a system of nonlinear constraints corresponding to expression (13) named as def inequality_constraint_3var(*a*). Moreover, the calculation of the right side of the inequality (13) is the separate function f01() in order to reduce the execution time of the algorithm, that is, a single calculation of f01() and transfer the result to the system of nonlinear constraints (13).

The second file contains the TestFunction class, which presents 2 testing functions: a) daf hd2(*self, res, alpha* = 0.05) checks a number of *res*($\varepsilon^*$) residuals for compliance with the normal distribution law according to the criterion $\chi^2$; b) hi2_plotn_3var(*self, res, r_3, alfa* = 0.05) checks a number of residuals ($\varepsilon^*$) for compliance with the law with the density of the probability distribution (7) according to the criterion $\chi^2$. The alfa variable sets the significance level of the criterion. Splitting the sample into intervals for calculating the distribution frequencies can be done automatically or using the Sturdess formula (*n* = 1 + np.trunc(3.322 * np.log10(*n_res*))), np.trunc cuts off the fractional part, np.log10 is the decimal logarithm, *n_res* is the number of observed values $\varepsilon^*$.

The fifth output block is implemented through the main file main.py, in which optimization procedures, testing and plotting of frequencies $\varepsilon^*$, normal distribution and distribution with density as (10) are started.

The built-in minimize function from the Scipy library is used for optimization. The SLSQP (Sequential Least SQuares Programming) method is used as the basic algorithm [30]. The optimization results are output to the console and to the *.txt text file. Models with adjusted parameters are tested for compliance with the normal law and the law with the density described by formula (7), and are entered into the console.

The architecture of the project is shown in *Fig. 2.*

The necessary Python libraries are installed in the venv directory. At the same time, the project is isolated from other projects by creating its own virtual environment.

Thus, the developed Python module in combination with the EFRA software package allows is: to correct the parameters of the production functions and the aggregated production function, to test the hypothesis of the correspondence of a number of residues to the given distribution laws using the criterion $\chi^2$ based on the method developed for estimating the parameters of the APF for the three-dimensional case.

| | |
|---|---|
| Search Parameters | venv... |
| | DataImport_BD.xlsx |
| | DataInput.py |
| | DistributionDensity.py |
| | main.py |
| | Models.py |
| | ObjectiveFunction.py |
| | Results.txt |
| | TestFunction.py |

*Fig. 2.* Project architecture.

## 3. Results of the realization of the method for estimating the parameters of a three-component APF on the example of the Central Federal District regions

### 3.1. Testing on model data

Initially, three normally distributed random variables $\varepsilon_i \sim N(0; 1)$ of 50, 100, 1000 and 10000 observations were generated to test the algorithm and test the hypothesis that the aggregated random variable according to (9) distribution law corresponds to the density determined by the formula (7) (*Fig. 3*).

*Figure 3* shows that as the number of observations increases, the significance level for the normal law decreases, and for the law with density (7), on the contrary, it increases. At the same time, for a sample of 50 observations, the distribution of the aggregated random variable corresponds to the distribution with density (7) and is significant at the level of 0.091. That is, on samples of 50 and 100 observations, hypotheses about compliance with both the normal distribution and the distribution with density (7) cannot be rejected.

### 3.2. Construction of models for the SES project subsystem

In accordance with Kleiner's spatio-temporal classification, the socio-economic system includes four subsystems: object, environment, process and project type, the interaction between which forms the level of systemic balance of the economy [16]. At the same time, the project subsystem is characterized by the volume of gross regional product (GRP) according to sections F (construction), G (wholesale and retail trade), K (financial activity) according to NACE rev. 2; previously according to NACE rev. 1 sections were designated as F, G and J, respectively. In this sense, the project subsystem is three-component and can serve as an object of evaluation using the method developed.

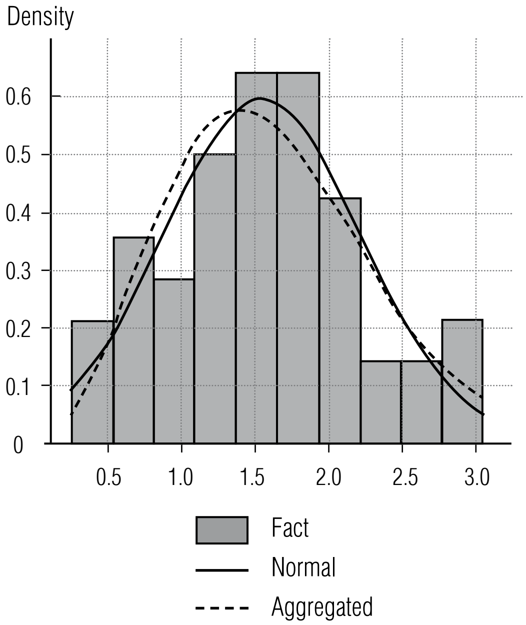
Based on previous studies [31], power-law multiplicative models were chosen as the functional form of the models, linking the volume of GDP by region in sections F(F), G(G), K(J) with the cost of fixed assets (total for section K(J), for sections F(F) and G(G)), the average annual number of employed (for sections F(F) and G(G)) and the average annual population (for section K), represented by the formula (14) and in linearized form by the formula (15):

$$\hat{y}_i = C_{i,0} \cdot x_{i,1}^{C_{i,1}} \cdot x_{i,2}^{C_{i,2}}, \quad (14)$$

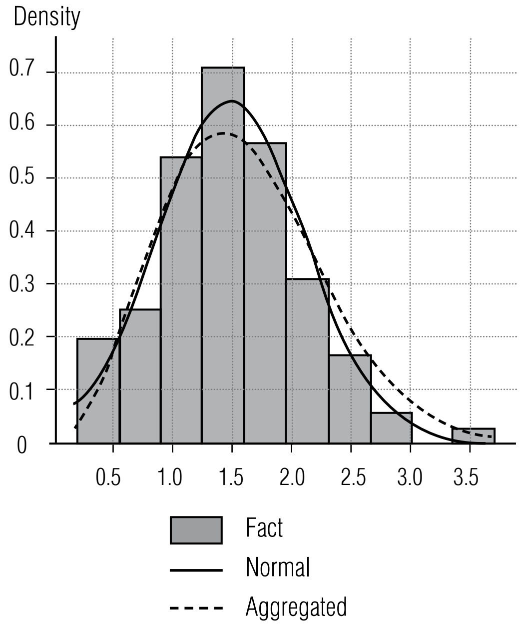
$$\ln(\hat{y}_i) = \ln(C_{i,0}) + C_{i,1} \cdot \ln(x_{i,1}) + C_{i,2} \cdot \ln(x_{i,2}). \quad (15)$$

The basis for the formation of models was Rosstat's open data for 17 regions of the Central Federal District (excluding Moscow) for the period from 2007 to 2020 [32]. All cost indicators were adjusted for the level of inflation and brought to the level of 2007 according to the formula:

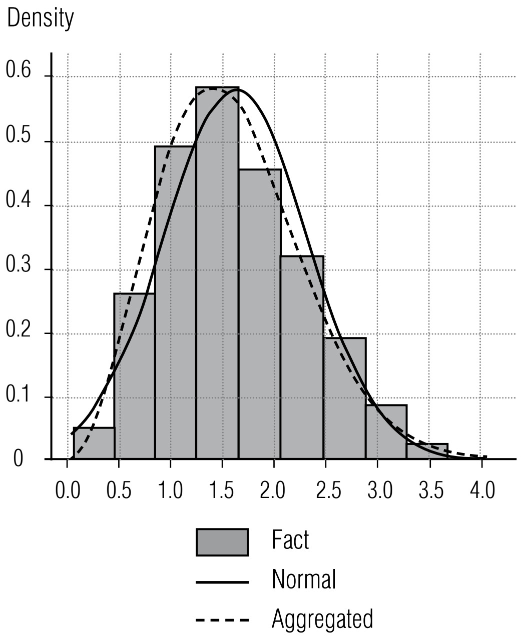
$$(\cdot)_t = (\cdot) / \prod_{i=2}^{t} (1 + \pi_i / 100). \quad (16)$$



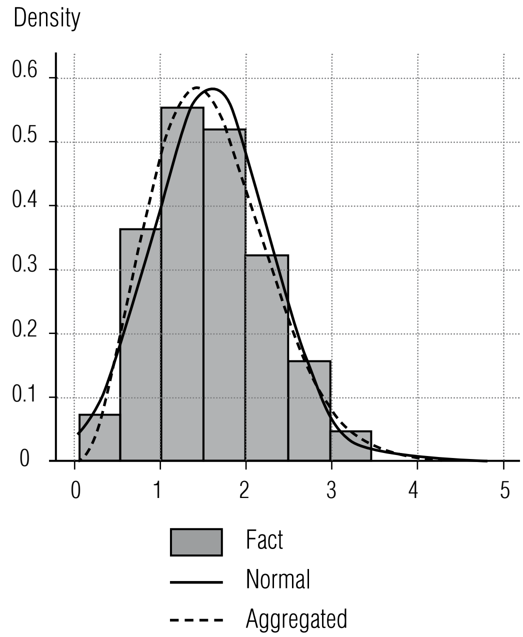

a) Residuals – 3var, p–value = 0.688 (normal law), p–value = 0.091 (law with density (7)).




b) Residuals – 3var, p–value = 0.351 (normal law), p–value = 0.316 (law with density (7)).




c) Residuals – 3var, p–value = 0.000 (normal law), p–value = 0.608 (law with density (7)).




d) Residuals – 3var, p–value = 0.000 (normal law), p–value = 0.905 (law with density (7)).

*Fig. 3.* Frequency diagram of the evaluation results of a three–component random variable with $\varepsilon_i \sim$ N(0; 1) volume a) 50 b) 100, c) 1000, d) 10000 observations; p–value is significance level according to the criterion $\chi^2$; number of intervals 10.

Here $\pi_i$ is the inflation rate in the *i*-th period ($i = 2$ correspond 2008).

Preliminary results showed the significance of the coefficients of the models and the coefficient of determination $R^2$. However, with further testing of models and their adequacy for a number of residuals (randomness, equality of zero mathematical expectation, the presence of autocorrelation, compliance with the normal distribution law, homoscedasticity test), the sample size had to be reduced until 2018–2020 (section F(F)) and 2015–2020 (sections G(G) and K(J)). The results of the parameter evaluation are presented in *Table 1.*

Using the method presented, the parameters of APF and PF were estimated in two variants: the aggregated random variable was determined by formulas (8) and (9) with distribution density (7) and (10), likelihood function (12) and constraints (13). The volume of the combined sample for the period 2018–2020 was 51. At the same time, the algorithm by which the distribution density was calculated using eigenvalues and eigenvectors to transform the integration domain to a canonical form (the function UnderIntFunctionLA) turned out to be almost three times slower than the algorithm in which the distribution density was determined in quadratures (the function UnderIntFunctionQuadraqticForm), which led to the conclusion that it is advisable to use the Lagrange method instead of the first option when converting variables to spherical coordinates with the maximum possible analytical description of the integrand expression of the density of the distribution of an aggregated random variable.

The results of the evaluation of the models are presented in *Table 2.*

*Table 1.*

**The main statistical characteristics of the assessed models**

| Model | $C_0$ | $C_1$ | $C_2$ | $R_2$ | v | rnd | M(e) | DW | W | $r_{x1}$ / $r_{x2}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| F(F) | 93.306 | 0.248 | 0.711 | 0.961 | 48 | 26/31 | 1.182 | 1.809 | 0.973 | 0.141 0.145 |
| p–value | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.050 | 0.243 | 0.050 | 0.282 | 0.322 0.310 |
| G(G) | 10.365 | 0.415 | 0.838 | 0.961 | 99 | 47/51 | 1.744 | 1.953 | 0.097 | 0.098 0.222 |
| p–value | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.050 | 0.084 | 0.050 | 0.027 | 0.326 0.025 |
| K(J) | 0.007 | 0.264 | 1.077 | 0.928 | 99 | 58/68 | 1.543 | 1.944 | 0.973 | 0.085 0.099 |
| p–value | 0.000 | 0.001 | 0.000 | 0.00 | – | 0.050 | 0.126 | 0.050 | 0.035 | 0.395 0.319 |

Note: the letter designations in the first column are the model for the NACE rev. 2; () is NACE rev. 1; p–value is the level of statistical significance; $C_i$ are the parameters models values; $R^2$ is the coefficient of determination; v is degrees of freedom; rnd is critical (for the significance level of 0.05) and the estimated number of turning points (check for the randomness of a number of residuals); M(e) is t–statistics (checking the equality of 0 of the mathematical expectation of a number of residuals); DW is the Darbin–Watson criterion (checking the absence of autocorrelation of a number of residuals, significant at the specified level); W is the Shapiro–Fork criterion (checking for the normality of a number of residues); $r_{xi}$ is t–statistics using Spearman's rank correlation coefficient of factor xi (homoscedasticity test). Residuals were constructed for linearized models.

*Table 2.*

**The main statistical characteristics of the assessed models (aggregated estimate)**

| Model | $C_0$ | $C_1$ | $C_2$ | $R_2$ | v | rnd | M(e) | DW | W | $r_{x1}$ / $r_{x2}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| F(F)[a] | 90.101 | 0.253 | 0.708 | 0.960 | 48 | 26/31 | 1.241 | 1.810 | 0.972 | 0.140 0.145 |
| p–value[a] | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.050 | 0.220 | 0.050 | 0.275 | 0.327 0.310 |
| F(F)[b] | 97.373 | 0.241 | 0.715 | 0.961 | 48 | 26/31 | 1.418 | 1.869 | 0.973 | 0.155 0.151 |
| p–value[b] | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.050 | 0.162 | 0.050 | 0.291 | 0.277 0.289 |
| G(G)[a] | 11.961 | 0.347 | 0.945 | 0.970 | 99 | 58/57 | 1.847 | 1.979 | 0.969 | 0.125 0.317 |
| p–value[a] | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.010 | 0.068 | 0.050 | 0.017 | 0.212 0.001 |
| G(G)[b] | 9.578 | 0.381 | 0.921 | 0.974 | 99 | 58/59 | 1.759 | 1.988 | 0.973 | 0.176 0.346 |
| p–value[b] | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.050 | 0.082 | 0.050 | 0.037 | 0.077 0.000 |
| K(J)[a] | 0.008 | 0.283 | 1.035 | 0.927 | 99 | 58/66 | 1.355 | 1.967 | 0.973 | 0.072 0.064 |
| p–value[a] | 0.000 | 0.000 | 0.000 | 0.000 | – | 0.050 | 0.178 | 0.050 | 0.037 | 0.473 0.520 |
| K(J)[b] | 0.0245 | 0.181 | 1.085 | 0.900 | 99 | 58/65 | 0.607 | 1.994 | 0.972 | 0.068 0.002 |
| p–value[b] | 0.000 | 0.039 | 0.000 | 0.000 | – | 0.050 | 0.545 | 0.050 | 0.030 | 0.494 0.981 |

Note: a) the aggregated random variable was determined by the formula (8); b) by the formula (9).

The results of the evaluations of the models presented in *Table 2* show that the models remained adequate after adjusting their parameters, both for the first and second options.

The results of the evaluation of the aggregated random variable and the values of the likelihood function are presented in *Table 3*.

The table shows that the value of the likelihood function after optimization increased by 4.809% and 7.437%, respectively, which confirms the hypothesis of an increase in the reliability of estimates using APF with parameters adjusted after optimization. That is, the method makes it possible to establish more reasonable norms for the results of the functioning of SES subsystems, in particular for three-element subsystems. However, the reliability of estimates for the compliance of the aggregated random variable with the distribution laws under consideration decreases, although it remains significant for the case d). When using $\varepsilon^*$, calculated by the

*Table 3.*

**The results of the assessment of the aggregated random variable**

| Characteristic / APF | $\varepsilon^{*a}$ | $\varepsilon^{*b}$ | $\varepsilon^{*c}$ | $\varepsilon^{*d}$ |
|---|---|---|---|---|
| $Ln(L(\varepsilon^*))$ | –69.147 | –66.049 | –65.821 | –61.136 |
| $\Delta Ln(L(\varepsilon^*))$ | – | – | 3.325 (4.809%) | 4.912 (7.437%) |
| $\chi^2_{norm}$ | 10.477 | 16.800 | 10.079 | 39.352 |
| p–value$_{norm}$ | 0.063 | 0.005 | 0.073 | 0.000 |
| $\chi^2_{agg}$ | 169.125 | 7.841 | 168.701 | 12.848 |
| p–value$_{agg}$ | 0.000 | 0.165 | 0.000 | 0.025 |

Note: $Ln(L(\varepsilon^*))$ is the value of the likelihood function; $\Delta Ln(L(\varepsilon^*))$ is change of the likelihood function; $\chi^2_{norm}$ is criterion value $\chi^2$ according to the normal law; $\chi^2_{agg}$ is criterion value $\chi^2$ to comply with the law with the density (7); p–value (norm, agg) are statistical significant levels, respectively $\chi^2_{norm}$ and $\chi^2_{agg}$; a) is $\varepsilon^*$ calculated by the formula (8) before optimization; b) is $\varepsilon^*$ calculated by the formula (9) before optimization; c) and d) are $\varepsilon^*$ calculated by the formulae (8) and (9) after optimization, respectively.

formula (8), the aggregated random variable is closer to the normal law; when using (9) it is closer to the law with density (7). These results are consistent with the conclusions obtained earlier for a two-component aggregated random variable [25].

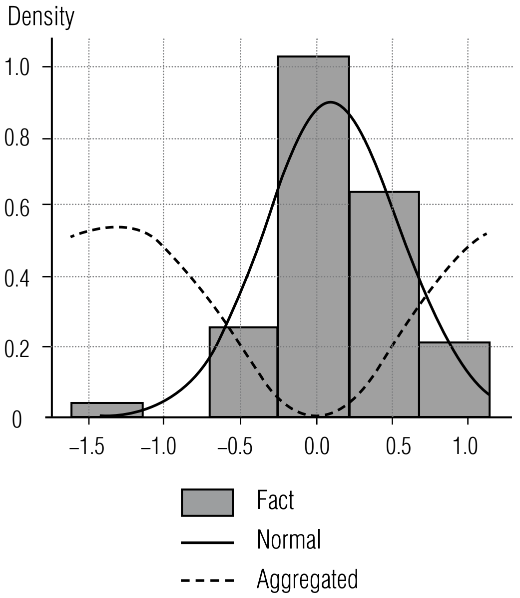
Diagrams of the aggregated random variable obtained by formulas (8) and (9) with p-value before and after optimization are shown in *Figs. 4 (a), (b), (c), (d)*, respectively.

The results of the evaluation of PF and APF parameters obtained were used to calculate the partial and integral performance indicator of the functioning of the project subsystem of the Central Federal District regions for 2018–2020 in three variants. The calculation results are presented on an external resource[1].

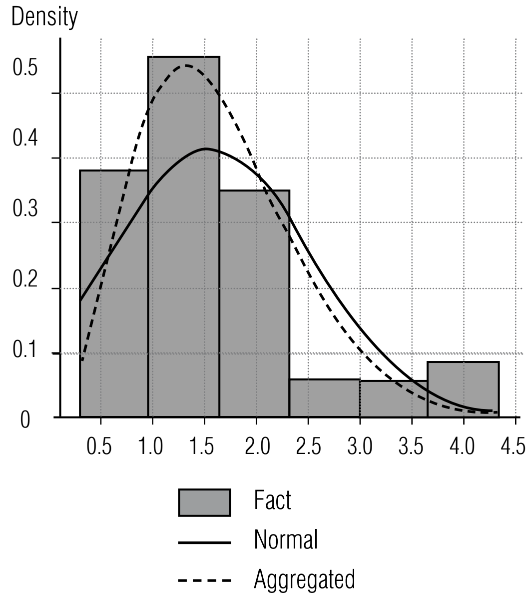
*Figure 5* shows the results of the calculation of indicators for the Tula Region in 2020.

*Figure 5* shows that the values of the indicators calculated for different variants are close to each other, and in the first approximation, estimates of the parameters calculated for each of the sections separately can be used to evaluate the project subsystem. If it is necessary to establish more reliable standards for the subsystem, it is advisable to adjust the parameters of the APF and PF using the distribution density of the form (9), since the likelihood function based on it is maximal among the other options.

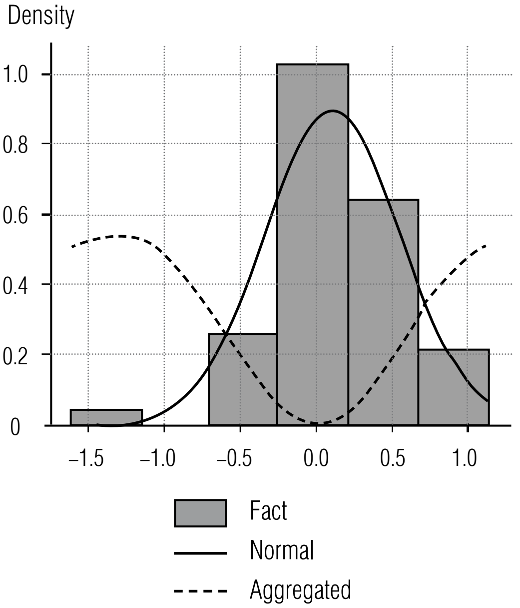
[1] The values of partial and integral performance indicators for the Central Federal District regions for 2018–2020 [Electronic resource]: https://disk.yandex.ru/i/NWOKIpcJG-sULA



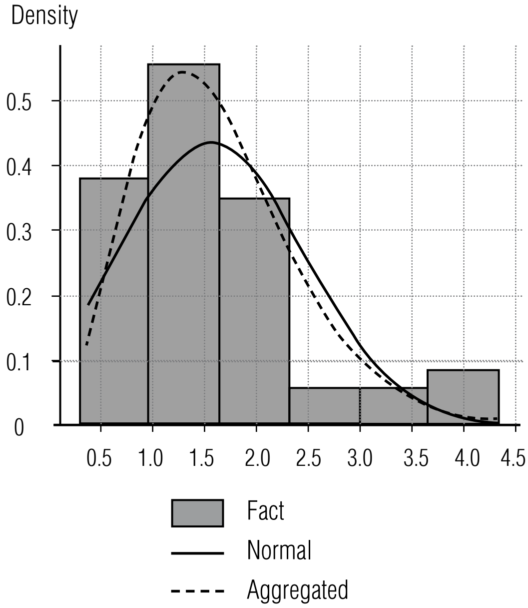

a) Residuals – 3var, p–value = 0.063 (normal law),
p–value = 0.000 (law with density (7)).




b) Residuals – 3var, p–value = 0.005 (normal law),
p–value = 0.165 (law with density (7)).




c) Residuals – 3var, p–value = 0.073 (normal law),
p–value = 0.000 (law with density (7)).




d) Residuals – 3var, p–value = 0.000 (normal law),
p–value = 0.025 (law with density (7)).

*Fig. 4.* Frequency diagram of the evaluation results of a three–component random variable with $\varepsilon_i$
a) is formula (8) before optimization, b) is formula (9) before optimization,
c) and d) are formulas (8) and (9) after optimization;
fact is histogram $\varepsilon_i$; Normal is normal law; Aggregated is a law with density (7);
p–value is the significance level according to the criterion $\chi^2$;
the number of intervals is 6 (calculated by the Sturgess formula), the sample size is 51.

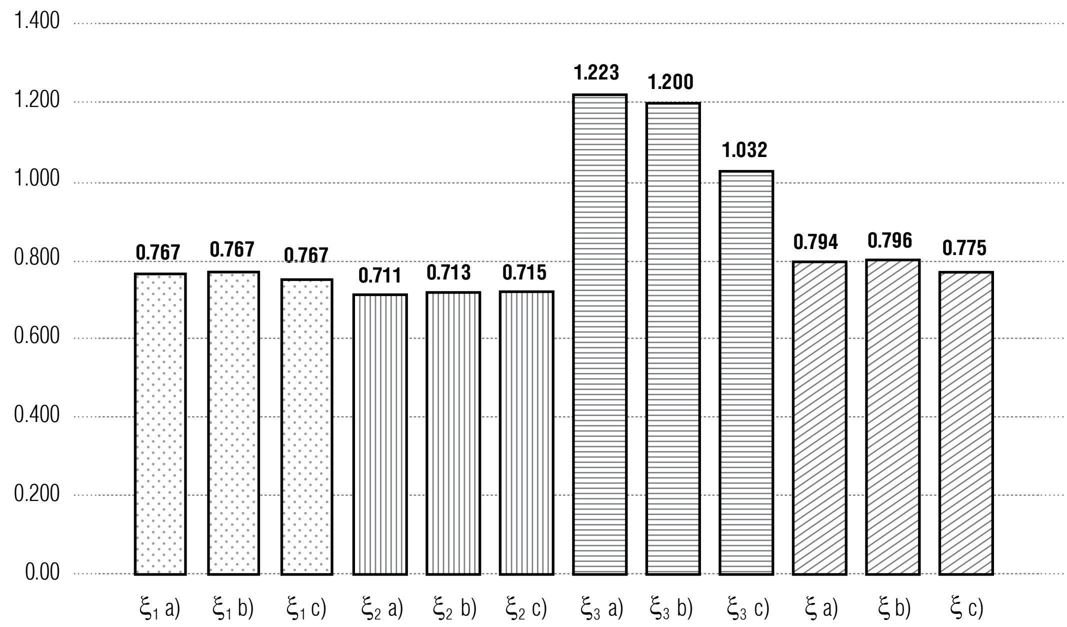



*Fig. 5.* The values of performance indicators for the Tula Region in 2020:
$\xi_1$, $\xi_2$, $\xi_3$ are partial performance indicators for sections F(F), G(G) and K(J) according to NACE rev. 2 (NACE rev. 1);
$\xi$ is integral performance indicator of the project subsystem;
a) are PF parameters determined separately;
b) formula (8) was used to assess the parameters of APF and PF,
c) formula (9) was used.

## Conclusion

This article presents a method for estimating the parameters of an aggregated production function used to calculate the standards for the results of the functioning of SES subsystems, implemented for a three-component APF. The difference of the method is the joint acquisition of the PF parameters of the SES elements, which ensures the consistency of the PF within one subsystem.

The application of the method for the regions of the Central Federal District using the developed and tested Python software project made it possible to adjust the parameters of PF and APF and obtain statistically appropriate models that can be used to build standards for elements of design subsystems and subsystems in general for the regions of the Central Federal District. This confirmed the earlier hypothesis about the possibility of using the method for three-component subsystems.

The results of the assessment of the functioning of the Central Federal District regions for 2018–2020 with the help of partial and integral indicators can be useful to regional governments for subsequent analysis and synthesis of solutions that make it possible to ensure that the actual and normative values of the resulting features correspond to a given degree of accuracy by changing and (or) intensifying the use of factors included in the models developed. ■

## Acknowledgments

The study was funded by a grant from the Russian Science Foundation № 22-28-20061, https://rscf.ru/project/ 22-28-20061/ and Tula Region.

## References


1. Maslennikov D.A., Mityakov S.N., Kataeva L.Y., Fedoseeva T.A. (2019) Identification of the characteristics of the regional strategic development based on the indicators' statistical analysis. *Economy of region*, vol. 15, no. 3, pp. 707–719 (in Russian). https://doi.org/10.17059/2019-3-7
2. Lokosov V.V., Ryumina E.V., Ulyanov V.V. (2018) Population quality and regional economy: direct and indirect correlation. *Economic and Social Changes: Facts, Trends, Forecast*, vol. 11, no. 1, pp. 32–42 (in Russian). https://doi.org/10.15838/esc/2018.1.55.2
3. Tretyakova E.A., Osipova M.Y. (2018) Evaluation of sustainable development indicators for regions of Russia. *Studies on Russian Economic Development*, vol. 29, no. 2, pp. 124–134 (in Russian).
4. Mishra S.K. (2010) A Brief History of Production functions. *The UIP Journal of Managerial Economics, UIP Publications*, vol. 8, no. 4, pp. 6–34.
5. Palash S.V. (2017) Structural balance of the economy: government programs for industrial development in the Russian Federation. *St. Petersburg State Polytechnic University Journal of Engineering Science and Technology*, vol. 10, no. 1, pp. 53–72 (in Russian). https://doi.org/10.18721/JE.10105
6. Kutyshkin A.V., Orlova D.S. (2019) Diagnostics of the sustainable development of regional socio-economic systems on the example of the Khanty-Mansiysk autonomous okrug – Yugra. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*, vol. 19, no. 2, pp. 103–116 (in Russian). https://doi.org/10.14529/ctcr190209
7. Sun X., Liu X., Li F., Tao Y., Song Y. (2017) Comprehensive evaluation of different scale cities sustainable development for economy, society, and ecological infrastructure in China. *Journal of Cleaner Production*, vol. 163, pp. 329–337. https://doi.org/10.1016/j.jclepro.2015.09.002
8. Krivonozhko V.E., Lychev A.V. (2010) *Analysis of the activities of complex socio-economic systems*. Moscow: Maks Press (in Russian).
9. Charnes A., Cooper W.W., Rhodes E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444.
10. Makarov V.L., Aivazyan S.H., Afanasiev M.Yu., Bakhtizin A.R., Nanavyan A.M. (2014) The estimation of the regions' efficiency of the Russian Federation including the intellectual capital, the characteristics of readiness for innovation, level of well-being, and quality of life. *Economy of Region*, vol. 4. pp. 9–30 (in Russian). https://doi.org/10.17059/2014-4-1
11. Aivazian S.A., Afanasiev M.Yu., Kudrov A.V. (2018) Indicators of economic development in the basis of the characteristics of regional differentiation. *Applied Econometrics*, vol. 50, pp. 4–22 (in Russian).
12. Vinogradova N.A. (2016) Integrated index of regional development. *Regional Economics: Theory and Practice,* vol. 2, pp. 68–83 (in Russian).
13. Dmitriev V.V., Kaledin N.V. (2016) Russian Northwest: an integral assessment of the conditions of regional social, environmental and economic systems and quality of life. *Baltic Region*, vol. 8, no. 2, pp. 125–140 (in Russian). https://doi.org/10.5922/2074-9848-2016-2-7
14. Zhgun T.V., Lipatov A.V., Chalov G.A. (2019) On the correctness of the computational problem of composite indices construction. *Modern Information Technologies and IT-Education*, vol. 15, no. 2, pp. 441–455 (in Russian). https://doi.org/10.25559/SITITO.15.201902.441-455
15. Kolesnikov N.G., Tolstoguzov O.V. (2016) Structural changes in the economy of the Russian Northwest: Spatial dimension. *Baltic Region*, vol. 8, no. 2, pp. 30–47 (in Russian). https://doi.org/10.5922/2074-9848-2016-2-2
16. Kleiner G.B., Rybachuk M.A. (2019) System balance of the Russian economy: Regional perspective. *Economy of Region*, vol. 15, no. 2, pp. 309–323 (in Russian). https://doi.org/10.17059/2019-2-1
17. Dreyer J.K., Schmid P.A. (2017) Growth effects of EU and EZ memberships: Empirical findings from the first 15 years of the Euro. *Economic Modelling*, vol. 67, pp. 45–54. https://doi.org/10.1016/j.econmod.2016.09.007

18. Sayaria N., Saria R., Hammoudehb S. (2018) The impact of value added components of GDP and FDI on economic freedom in Europe. *Economic Systems*, vol. 42, no. 2, pp. 282–294. https://doi.org/10.1016/j.ecosys.2017.03.003

19. Charfeddine L., Mrabet Z. (2017) The impact of economic development and social-political factors on ecological footprint: A panel data analysis for 15 MENA countries. *Renewable and Sustainable Energy Reviews,* vol. 76, pp. 138–154. https://doi.org/10.1016/j.rser.2017.03.031

20. Lin B., Benjamin I.N. (2018) Causal relationships between energy consumption, foreign direct investment and economic growth for MINT: Evidence from panel dynamic ordinary least square models. *Journal of Cleaner Production*, vol. 197, pp. 708–720. https://doi.org/10.1016/j.jclepro.2018.06.152

21. Zhenhua W., Guangsheng Z. (2016) Industrial policy, production efficiency improvement and the Chinese county economic growth. *Proceedings of Rijeka Faculty of Economics: Journal of Economics and Business*, vol. 34, no. 2, pp. 505–528. https://doi.org/10.18045/zbefri.2016.2.505

22. Afanasiev A.A., Ponomareva O.S. (2020) The macroeconomic production function of Russia in 1990–2017. *Economics and Mathematical Methods*, vol. 56, no.1, pp. 67–78 (in Russian). https://doi.org/10.31857/S042473880006708-7

23. Makarov V., Ayvazyan S., Afanasyev M., Bakhtizin A., Nanavyan A. (2016) Modeling the development of regional economy and an innovation space efficiency. *Foresight and STI Governance*, vol. 10, no. 3, pp. 76–90 (in Russian). https://doi.org/10.17323/1995-459X.2016.3.76.90

24. Zhukov R., Kuznetsov G., Gorodnichev S., Manokhin E., Nazyrova E., Melay E. (2019) Comparative analysis of results of assessing the Central Federal District's regions' economic development by using linear and non-linear models. *Statistika: Statistics and Economy Journal*, vol. 99, no. 3, pp. 272–286.

25. Zhukov R.A. (2021) Method for assessing the results of hierarchical socio-economic systems' functioning based on the aggregated production function. *Economics and Mathematical Methods*, vol. 57, no. 3, pp. 17–31 (in Russian). https://doi.org/10.31857/S042473880016428-9

26. Zhukov R.A. (2020) An approach to assessing the functioning of hierarchical socio-economic systems and decision-making based on the EFRA software package. *Business Informatics*, vol. 14, no. 3, pp. 82–95 (in Russian). https://doi.org/10.17323/2587-814X.2020.3.82.95

27. Zhukov R.A. (2019) *Socio-ecological-economic systems: Theory and practice*. Moscow: INFRA-M (in Russian). https://doi.org/10.12737/monography_5b7516626665a8.43347695

28. Zhukov R.A., Kozlova N.O. (2022) On the probability distribution densities of an aggregated random variable for evaluating the functioning of complex systems: a three-dimensional case. In digest *XXI international conference "Algebra, number theory, discrete geometry and multiscale modeling, modern problems, applications and problems of history dedicated to the 85th anniversary of the birth of A.A. Karatsuba"*, pp. 268–270 (in Russian).

29. *Programming language Python*. Available at: http://python.org (accessed 10 August 2022).

30. *SLSQP Python method*. Available at: https://docs.scipy.org/doc/scipy/tutorial/optimize.html (accessed 10 August 2022).

31. Zhukov R.A. (2018) Model of socio-ecological and economic system: The central federal district regions of the Russian Federation. *Statistika: Statistics and Economy Journal*, vol. 98, no. 3, pp. 237–261.

32. *Federal State Statistics Service*. Available at: http://gks.ru (accessed 10 August 2022).

## About the authors

**Roman A. Zhukov**

Cand. Sci. (Phys.-Math.), Associate Professor;

Researcher, Associate Professor, Department of Mathematics and Informatics, Financial University under the Government of the Russian Federation, Tula Branch, 1a, Oruzheynaya Street, Tula 300012, Russia;

E-mail: pluszh@mail.ru

ORCID: 0000-0002-2280-307X

**Nadezhda A. Kozlova**

Cand. Sci. (Tech.), Assistant Professor;

Assistant Professor, Department of Mathematics and Informatics, Financial University under the Government of the Russian Federation, Tula Branch, 1a, Oruzheynaya Street, Tula 300012, Russia;

E-mail: 95kno@mail.ru

ORCID: 0000-0001-6808-2880

**Evgeny V. Manokhin**

Cand. Sci. (Phys.-Math.), Associate Professor;

Associate Professor, Department of Mathematics and Informatics, Financial University under the Government of the Russian Federation, Tula Branch, 1a, Oruzheynaya Street, Tula 300012, Russia;

E-mail: emanfinun@gmail.com

ORCID: 0000-0001-6711-3737

**Maria A. Plinskaya**

Student of the direction "Business Informatics", Financial University under the Government of the Russian Federation, Tula Branch, 1a, Oruzheynaya Street, Tula 300012, Russia;

E-mail: maria.plinskaya@gmail.com

ORCID: 0000-0002-1307-0935