УДК 519.681

COMPARATOR REGRESSION ANALYSIS IN CONDITIONS OF FUZZY INITIAL DATA

КОМПАРАТОРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ В УМОВАХ НЕЧІТКИХ ВИХІДНИХ ДАНИХ

Raskin L.G.[1], Sukhomlyn L.V.[2], Sokolov D.D.[3] , Vlasenko V.V.[4]
Раскін Л.Г.[1], Сухомлин Л.В.[2], Соколов Д.Д.[3] , Власенко В.В.[4]

[1,3,4] National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine,
[2] Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine
ORCID: [1] http://orcid.org/0000-0002-9015-4016, [2]https://orcid.org/0000-0001-9511-5932,
[3]https://orcid.org/0000-0002-4558-9598, [4]https://orcid.org/0000-0001-5427-0223
E-mail: [1]: topology@ukr.net, [2] lar.sukhomlyn@gmail.com, [3] sokolovddd@gmail.com, [4] vitalik.vlasenko.000@gmail.com

DOI: 10.15673/atbp.v16i3.2926

Abstract. *The canonical task of regression analysis consists in studying the problem of estimation and mathematical description of the influence caused by one or more independent variables (predictors, explanatory variables) on the dependent (explained) variable. The subject of comparator identification problem is a specific modification of the standard regression analysis problem in the case when, for some reason, it is not possible to measure values of the dependent, explained variable. However, the problem of reconstructing a regression polynomial can be solved if it is possible to qualitatively compare the results of experiments and rank them, for example, in the order of increasing. Taking into account the results of ranking the values of the explained variable in different experiments, a corresponding system of inequalities is formed. The purpose of the research consists in developing a method for mathematical processing of the resulting system of inequalities for analytical assessment of the level of influence of explanatory variables on the explained variable. The method for solving the problem is based on transforming a system of inequalities into a system of linear algebraic equations. A significant drawback of the known approach to solving this problem consists in the fact that the resulting system of linear algebraic equations is an underdetermined one (the number of unknowns in this system exceeds the number of equations). At the same time, the system has an infinite number of solutions, among which there may be an uncontrollable number of unacceptable ones. The exhaustive search method for finding acceptable solutions is futile. In this regard, the problem of developing a correct method of comparator identification remains relevant. Another difficulty that accompanies the actual procedure of comparator identification when solving practical problems consists in the fact that measurements of values of the controlled parameters are not accurate. The uncertainty that arises in the conditions of a small sample of initial data can be removed using the tools of fuzzy mathematics. It is assumed that, based on preliminary research, a mathematical description of the corresponding membership function can be obtained for each of the controlled parameters. As a result, effective approaches to solving emerging problems have been proposed, based on the use of developed optimization procedures in the conditions of a small sample of fuzzy initial data.*

Анотація. *Канонічне завдання регресійного аналізу полягає в дослідженні проблеми оцінювання та математичному описі впливу однієї або кількох незалежних змінних (предикторів, пояснювальних змінних) на залежну (пояснювальну) змінну. Предметом задачі компараторної (порівняльної) ідентифікації є специфічна модифікація типової задачі регресійного аналізу на випадок, коли з якихось причин не є можливим вимірювання значень залежної, пояснюваної змінної. Однак завдання відновлення регресійного полінома може бути розв'язане, якщо існує можливість якісного порівняння результатів експериментів та їх ранжування в порядку, наприклад, зростання. З урахуванням результатів ранжування значень пояснюваної змінної в різних дослідах формується відповідна система нерівностей. Мета дослідження полягає в розробленні методу математичної обробки отриманої системи нерівностей для аналітичної оцінки рівня впливу пояснювальних змінних на пояснювану. Метод розв'язання задачі ґрунтується на перетворенні системи нерівностей на систему лінійних алгебраїчних рівнянь. Істотний недолік відомого підходу до розв'язання цієї задачі полягає в тому, що отримана система лінійних алгебраїчних рівнянь є недовизначеною (кількість невідомих у цій системі перевищує кількість рівнянь). При цьому система має незліченну безліч розв'язків, серед яких може бути неконтрольоване число неприпустимих. Перебірний метод пошуку прийнятних рішень безперспективний. У зв'язку з цим завдання розроблення коректного методу компараторної ідентифікації залишається актуальним. Інша складність, що супроводжує реальну процедуру компараторної ідентифікації під час розв'язання практичних завдань, полягає в тому, що вимірювання значень контрольованих параметрів не точні. Невизначеність, що виникає при цьому, в умовах малої вибірки вихідних даних може бути знята при використанні інструментарію нечіткої математики. При цьому передбачається, що на підставі попередніх досліджень для кожного з контрольованих параметрів може бути отримано математичний опис відповідної функції належності. У результаті запропоновано ефективні підходи до розв'язання задач, що виникають, які ґрунтуються на використанні розроблених процедур оптимізації в умовах малої вибірки нечітких вихідних даних.*

Key words: comparator identification, underdetermined system of equations, fuzziness of initial data, small sample, fuzzy optimization.

Ключові слова: компараторна ідентифікація, невизначена система рівнянь, нечіткість вихідних даних, мала вибірка, нечітка оптимізація.

Introduction. Comparator regression analysis is a special unconventional version of canonical regression analysis [1, 2] applied in the following situation [3]. Let, as a result of a series of experiments containing n experiments obtained are corresponding values $y_1, y_2,..., y_n$ of a variable describing the process under study (behavior or state of the object under study). At the same time, each value of the explained variable y_j, $j = 1,2,..,n$ corresponds to a set of values of factors that presumably influence the y. The problem consists in developing a mathematical model that describes the relationship between the explanatory variables $F_1, F_2,..., F_m$ and the explained variable y in the form of a regression polynomial [1,2]

$$y_j = x_0 + x_1F_{j_1} + x_2F_{j_2} + ... + x_mF_{j_m}. \quad j = 1,2...n \tag{1}$$

The desired set $x = (x_1, x_2,.., x_m)$ is calculated using the least squares method. At that, introduced are

$$H = \begin{pmatrix} 1 & F_{11} & F_{12}...F_{1m} \\ 1 & F_{21} & F_{22}...F_{2m} \\ . & . & ...... \\ 1 & F_{n1} & F_{n2}...F_{nm} \end{pmatrix}, \quad X = \begin{pmatrix} x_0 \\ x_1 \\ ... \\ x_m \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix},$$

as well as a quadratic criterion $J = (HX - Y)^T (HX - Y)$ for the proximity of the observed values of the variables y_j, $j = 1,2,..,n$, and the values of these variables, calculated according to the formula [2]

$$X = (H^T H)^{-1} H^T Y \tag{2}$$

The problem is solved. Now, for some reason, it is impossible to determine the exact values of the explained variables y в in each experiment, but these values can be ranked, for example, in the in the order of increasing. Then, taking into account the numbering of variables y, we obtain a chain of inequalities:

$$y_1 < y_2 < ... < y_n \tag{3}$$

Now the aim is to find a vector x_1, that, using model (1), would ensure the fulfillment of inequalities (3). The problem formulated in this way is called the comparator identification problem [3-5].

Analysis of known results. Articulation of issue. The following method for solving this problem is known. Using (3) and taking into account (1), the following system of inequalities is formed [6.7]:

$$\begin{aligned} & y_1 - y_2 = (x_0 + x_1F_{11} + x_2F_{12} + ... + x_mF_{1m}) - \\ & -(x_0 + x_1F_{21} + x_2F_{22} + ... + x_{n_1}F_{2m}) = \\ & = x_1(F_{11} - F_{21}) + x_2(F_{12} - F_{22}) + ... + x_m(F_{1m} - F_{2m}) < 0, \\ & y_2 - y_3 = (x_0 + x_1F_{21} + x_2F_{22} + ... + x_mF_{2m}) - \\ & -(x_0 + x_1F_{32} + x_2F_{32} + ... + x_mF_{3m}) = \\ & = x_1(F_{21} - F_{31}) + x_2(F_{22} - F_{32}) + ... + x_m(F_{2m} - F_{3m}) < 0, \\ & y_{n-1} - y_n = (x_0 + x_1F_{n-1,1} + x_2F_{n-1,2} + ... + x_mF_{n-1,m}) - \\ & -(x_0 + x_1F_{n1} + x_2F_{n2} + ... + x_mF_{nm}) = \\ & = x_1(F_{n-1,1} - F_{n1}) + x_2(F_{n-1,2} - F_{n2}) + ... + x_m(F_{n-1,m} - F_{nm}) < 0. \end{aligned} \tag{4}$$

The resulting system of inequalities is the basic basis for calculating the values of the unknown parameters of the regression relationship that establishes the relationship between the explanatory variables $x_1, x_2,.., x_n$ and the explained variable y.

The system of inequalities (4) (by adding a positive variable to each of these inequalities) is transformed into a system of equalities [8.9]:

$$\begin{aligned} & x_1(F_{11} - F_{21}) + x_2(F_{12} - F_{22}) + ... + x_m(F_{1m} - F_{2m}) + x_{m+1} = 0, \\ & x_1(F_{21} - F_{31}) + x_2(F_{22} - F_{32}) + ... + x_m(F_{2m} - F_{3m}) + x_{m+2} = 0, \\ & x_1(F_{n-1,1} - F_{n1}) + x_2(F_{n-1,2} - F_{n3}) + ... + x_m(F_{n-1,m} - F_{n,m}) + x_{m+n} = 0. \end{aligned} \tag{5}$$

Thus, a system of linear algebraic equations has been obtained, and the solution of this system gives the required set x.

Let's consider a simple example. Let us investigate a system effectiveness of which y presumably depends on the values of two factors F_1 and F_2 in accordance with the relation

$$y = x_1F_1 + x_2F_2.$$

Suppose that during three experiments the corresponding values of the factors were recorded and compiled into a matrix

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 8 & 2 \\ 10 & 1 \end{pmatrix}.$$

As a result of expert assessment of the system's effectiveness in these three experiments, their ranking was obtained as follows: $y_1 < y_2 < y_3$, and based on (1), (3)-(4), this allows us to write the set:

$$\begin{aligned} y_1 &= x_1F_{11} + x_2F_{12}, \\ y_2 &= x_1F_{21} + x_2F_{22}, \\ y_3 &= x_1F_{31} + x_2F_{32}. \end{aligned} \quad (6)$$

$$\begin{aligned} y_1 - y_2 &= (x_1F_{11} + x_2F_{12}) - (x_1F_{21} + x_2F_{22}) = \\ &= x_1(F_{11} - F_{21}) + x_2(F_{12} - F_{22}) < 0, \\ y_2 - y_3 &= (x_1F_{21} + x_2F_{22}) - (x_1F_{31} + x_2F_{32}) = \\ &= x_1(F_{21} - F_{31}) + x_2(F_{22} - F_{32}) < 0. \end{aligned} \quad (7)$$

By means of substituting the numerical values of the factors into (7), we obtain

$$\begin{aligned} -4x_1 + 4x_2 &< 0, \\ -2x_1 + x_2 &< 0. \end{aligned} \quad (8)$$

By adding a positive variable to each of the inequalities, we transform them into the equalities:

$$\begin{aligned} -4x_1 + 4x_2 + x_3 &= 0, \\ -2x_1 + x_2 + x_4 &= 0. \end{aligned} \quad (9)$$

The resulting system of linear algebraic equations is underdetermined (it contains two equations and four variables) and therefore it has an infinite number of solutions, for example, this one:

$$x_1 = 0,5; \quad x_2 = 0,2; \quad x_3 = 1,2; \quad x_4 = 0,8.$$

Solution check:

$$-4 \cdot 0,5 + 4 \cdot 0,2 + 1,2 = -2 + 0,8 + 1,2 = 0$$
$$-2 \cdot 0,5 + 0,2 + 0,8 = 0.$$

At the same time, we should note that the set of solutions of system (9) may also contain solutions that are unacceptable, taking into account (4), for example the following:

$$x_1 = 0,2; \quad x_2 = 1,6; \quad x_3 = -1,6; \quad x_4 = -0,2.$$

Solution check:

$$-4 \cdot 0,2 + 4 \cdot 0,6 - 1,6 = -0,8 + 2,4 - 1,6 = 0$$
$$-2 \cdot 0,2 + 0,6 - 0,2 = 0.$$

In this solution, the complementary variables x_3 and x_4 are negative, which is unacceptable.

The solution to the comparator identification problem proposed in [8] reduces it to minimization of the quadratic form. This changes the computational procedure, but does not reduce the risk of obtaining an unacceptable solution. Let us consider another approach to obtaining a solution to system (5) that satisfies the requirement that complementary variables be positive; this approach consists in initially setting an arbitrary positive value for these variables Δ. Herewith, the system of equations (8) will take the following form:

$$-4x_1 + 4x_2 + \Delta = 0,$$
$$-2x_1 + x_2 + \Delta = 0.$$

The numerical value of the parameter Δ is arbitrary. Let it be, for example, as follows $\Delta = 1$. The corresponding system of equations will take the following form:

$$-4x_1 + 4x_2 = -1,$$
$$-2x_1 + x_2 = -1.$$

Let us solve this system of equations using Cramer formula [10-12]. Let us introduce the following

$$A = \begin{pmatrix} -4 & 4 \\ -2 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad B = \begin{pmatrix} -1 \\ -1 \end{pmatrix}. \quad (10)$$

The system of equations (9) in matrix form is written as:

$$AX = B. \quad (11)$$

Let us further introduce

$$detA = det\begin{pmatrix} -4 & 4 \\ -2 & 1 \end{pmatrix} = -4 + 8 = 4.$$

$$detA_1 = det\begin{pmatrix} -1 & 4 \\ -1 & 1 \end{pmatrix} = -1 + 4 = 3.$$

$$detA_2 = det\begin{pmatrix} -4 & -1 \\ -2 & -1 \end{pmatrix} = 4 - 2 = 2.$$

Herewith

$$x_1 = \frac{detA_1}{detA} = \frac{3}{4} = 0,75, \quad x_2 = \frac{detA_2}{detA} = \frac{2}{4} = 0,5$$

Let us check

$$-4 \cdot 0,75 + 4 \cdot 0,5 = -3 + 2 = -1,$$

$$-2 \cdot 0,75 + 0,5 = -1,5 + 0,5 = -1,$$

which is what was required

The numerical value of the positive parameter Δ can be chosen in any way. However, if the goal of the research consists in obtaining a ratio that will distinguish the states $y_1, y_2, .., y_n$ of the system with maximum contrast, then it is advisable to increase the value Δ .

Let us now consider the possibility of improving the described technology for high-dimensional comparator identification with a large number of states. The specific structure of the mathematical model of the comparator identification problem makes it possible to reduce its dimension to obtain an approximate solution. This possibility is based on the following obvious statement.

If $A < B$, with a certain level of confidence p_1 , and $B < C$, with a certain level of confidence p_2 , then $A < C$ with a level of confidence $p > \max\{p_1, p_2\}$.

Let us use this statement when considering the chain of inequalities (4). It follows from the statement that a pair of inequalities $y_1 < y_2$ and $y_2 < y_3$ can be replaced with a single one, namely $y_1 < y_3$, at the same time reducing the number of inequalities by one. It is clear that by repeating this technique, the total number of inequalities can be easily brought to m - the total number of unknowns in system (4). As a result, by adding a positive parameter Δ , to each inequality, obtained will be a standard system of m linear algebraic equations with m unknowns $x_1, x_2, .., x_m$.

The problem of comparator identification becomes significantly more complicated if the source data is not accurate. Let, for example, the values of the factors in the ongoing series of experiments be fuzzy numbers. In this case, we will assume that by means of using all available numerical information, the membership functions of these numbers are obtained. It is clear that the described standard method for solving the problem of comparator identification in conditions of fuzzy initial data cannot be directly applied.

Purpose and Objectives. In accordance with this, the purpose of the research consists in developing a method for solving the problem of comparator identification in conditions of vaguely specified initial data. To achieve the purpose, it is necessary to solve the following tasks.

1 Form a correct formulation of the problem of comparator identification in conditions of fuzzy initial data.

2. Justify the quality criterion for solving a fuzzy comparator identification problem.

3. Develop a method for solving the problem of comparator identification under conditions of fuzzy initial data.

Data and methods. Developing a method for solving the problem of comparator identification under conditions of fuzzy initial data.

Mathematical formulation of the problem.

It is assumed that a series of experiments have been carried out, and in each of these experiments measured are fuzzy values of influencing factors with membership functions $\mu(F_{ji}),\ j = 1, 2, .., n,\ i = 1, 2, .., m.$ These values of the explained variable corresponding to each of the experiments are ranked. As a result of this a system of inequalities is obtained (3). It is required to find a vector X , that provides calculation of the explained variable, satisfies (3) and optimizes the selected criterion.

Let us consider the following scheme for solving the posed problem of comparator identification. First, we will select the type of analytical description of the membership functions of fuzzy source data. Then we will build a mathematical model of the problem, and on the basis of this model we will form a naturally interpretable and easily calculated quality criterion for solving the identification problem. To complete the solution, we will develop a methodology for optimizing the selected criterion.

In practice, solving various problems of fuzzy mathematics, an extensive set of types of fuzzy number membership functions is used, for example, triangular one, trapezoidal one, Gaussian one and many others. At the same time, recently the method of constructing membership functions has been increasingly used, this method is based on the approximation of known types of membership functions by the so-called functions of $(L-R)$ type. Attractiveness of using $(L-R)$–type membership functions is explained by two important properties of these functions. Firstly, these functions are easily and conveniently defined by a fixed set of numerical parameters. Secondly, in relation to numbers with a $(L-R)$–type membership function, a strict system of rules for performing algebraic operations has been developed [13]. To implement these rules, it is necessary and sufficient to find sets of numerical values of the parameters of the membership functions of fuzzy numbers - operands. At the same time, in most practical problems it is acceptable to use these $(L-R)$–type three-parameter membership functions that have the following form

$$\mu(x)=\begin{cases}0, & x<m-\alpha,\\ L\left(\dfrac{m-x}{\alpha}\right), & m-\alpha\le x<m,\\ R\left(\dfrac{x-m}{\beta}\right), & m\le x\le m+\beta,\\ 0, & x>m+\beta.\end{cases} \quad (12)$$

where m is the modal value of the fuzzy number, x , α is the left fuzzy coefficient of the number x , β is the right fuzzy coefficient of the number x .

In the future, while using the generally accepted notation, we will use significantly more concise symbolism (instead of the cumbersome relation (12)) to describe three-parameter fuzzy numbers of the $(L-R)$-type in cases where the type of analytical description of the membership function is specified in advance; and this symbolism is as follows:

$$y=<m,\alpha,\beta>. \quad (13)$$

In accordance with the proposed general scheme for solving the problem of comparator identification, let us introduce a regression model (1) and a system of inequalities (4). However, unlike the previous one, all components in (4) are fuzzy numbers. Therefore, all such differences $y_j-y_{j+1},\ j=1,2,..,n-1,$ are fuzzy. Let's introduce a set of variables $\xi_j=y_j-y_{j+1},\ j=1,2,..,n-1,$ and, in accordance with the rules for performing arithmetic operations on fuzzy numbers of $(L-R)$ type [8], we will calculate the membership functions of fuzzy numbers $\xi_j,\ j=1,2,..,n.$ We will write down the corresponding summary of the rules for performing binary operations. Let's introduce a couple of fuzzy numbers $F_1=<m_1,\alpha_1,\beta_1>$ and $F_2=<m_2,\alpha_2,\beta_2>$.

Addition operation:

$$Z=F_1+F_2=<m_z,\alpha_z,\beta_z>,$$
$$m_z=m_1+m_2,\ \alpha_z=\alpha_1+\alpha_2,\ \beta_z=\beta_1+\beta_2;$$

Subtraction operation:

$$Z=F_1-F_2=<m_z,\alpha_z,\beta_z>,$$
$$m_z=m_1-m_2,\ \alpha_z=\alpha_1+\beta_2,\ \beta_z=\beta_1+\beta_2;$$

Multiplication operation:

$$Z=F_1\cdot F_2=<m_z,\alpha_z,\beta_z>,$$
$$m_z=m_1m_2,\ \alpha_z=m_1\alpha_2+m_2\alpha_1-\alpha_1\alpha_2,\ \beta_z=m_1\beta_2+m_2\beta_1+\beta_1\beta_2; \quad (14)$$

Division operation:

$$m_z=\frac{m_1}{m_2},\ \alpha_z=\frac{m_2\alpha_1+m_1\beta_2}{m_2(m_2+\beta_2)},\ \beta_z=\frac{m_1\alpha_2+m_2\beta_1}{m_2(m_2-\alpha_2)}. \quad (15)$$

Next

$$\xi_j=y_j-y_{j+1}=x_1(F_{j1}-F_{j+1,1})+x_2(F_{j2}-F_{j+1,2})+$$
$$+\ldots+x_m(F_{j-1,m}-F_{jm}).$$

As far as

$$F_{ji}=<m_{ji},\alpha_{ji},\beta_{ji}>,\ F_{j+1,i}=<m_{j+1,i},\alpha_{j+1,i},\beta_{j+1,i}>,$$

then

$$F_{ji}-F_{j+1,i}=0,\ r_{ji}=<m_{rji},\alpha_{rji},\beta_{rji}>,$$
$$m_{rji}=m_{ji}-m_{j+1,i},\ \alpha_{rji}=\alpha_{ji}+\beta_{j+1,i},\ \beta_{rji}=\beta_{ji}+\alpha_{j+1,i},$$
$$j=1,2,\ldots,n-1,\ i=1,2,\ldots,m.$$

From relation (14), which determines the result of multiplying two fuzzy numbers, we obtain formulas for the result of multiplying a fuzzy number by a crisp number. Let us write a crisp number x_i , using the notation adopted for describing fuzzy numbers of $(L-R)$ type: $x_i=<x_i,0,0>$. Then the fuzzy result of multiplying the crisp number x_i by the fuzzy number r_{ji} will be equal to

$$x_i r_{ji} = \Delta_{ji} =< m_{\Delta ji}, \alpha_{\Delta ji}, \beta_{\Delta ji} >,$$

$$m_{\Delta ji} = x_i m_{rji}, \alpha_{\Delta ji} = x_i(\alpha_{ji} + \beta_{j+1,i}),$$

$$\beta_{\Delta ji} = x_i(\beta_{ji} + \alpha_{j+1,i}).$$

Then

$$\xi_j = \sum_{j=1}^{m} x_i r_{ji} =< m_{\xi j}, \alpha_{\xi j}, \beta_{\xi j} >,$$

$$m_{\xi j} = \sum_{i=1}^{m} x_i m_{rji}; \ \ \alpha_{\xi j} = \sum_{i=1}^{m} x_i(\alpha_{ji} + \beta_{j+1,i});$$

$$\beta_{\xi j} = \sum_{i=1}^{m} x_i(\beta_{ji} + \alpha_{j+1,i}).$$

and finally

$$\xi = \sum_{j=1}^{n-1} \xi_j = \sum_{j=1}^{n-1} \sum_{i=1}^{m} x_i r_{ji} =< m_\xi, \alpha_\xi, \beta_\xi >, \tag{16}$$

$$m_\xi = \sum_{j=1}^{n-1} m_{\xi j} = \sum_{j=1}^{n-1} \sum_{i=1}^{m} x_i m_{rji};$$

$$\alpha_\xi = \sum_{j=1}^{n-1} \alpha_{\xi j} = \sum_{j=1}^{n-1} x_i(\alpha_{ji} + \beta_{j+1,i});$$

$$\beta_\xi = \sum_{j=1}^{n-1} \beta_{\xi j} = \sum_{j=1}^{n-1} \sum_{i=1}^{m} x_i(\beta_{ji} + \alpha_{j+1,i}). \tag{17}$$

All analytical relationships required for the formation of a quality criterion for solving the comparator identification problem under conditions of fuzzy initial data have been obtained. In order to satisfy the normalization condition (6) the searched (target) set $(x_1, x_2, ..., x_m)$ must be chosen so that all fuzzy numbers ξ_j, $j = 1, 2, ..., n-1,$ should be non-positive.

It is clear that the requirement of non-positivity of fuzzy numbers ξ_j, $j = 1, 2, ..., n-1,$ cannot be satisfied if the fuzzy numbers F_{ji}, defining the values ξ_j are defined on an infinite carrier. However, if fuzzy numbers are given on a compact carrier, then the problem can be solved. Let, for example, fuzzy numbers F_{ji} have a triangular membership function

$$\mu(F_{ji}) = \begin{cases} 0, \ F_{ji} < m_{ji} - \alpha_{ji}, \\ \dfrac{F_{ji} - (m_{ji} - \alpha_{ji})}{\alpha_{ji}}, m_{ji} - \alpha_{ji} \le F_{ji} < m_{ji}, \\ \dfrac{(m + \beta_{ji}) - F_{ji}}{\beta_{ji}}, m_{ji} \le F_{ji} < m_{ji} + \beta_{ji}, \\ 0, \ F_{ji} > m_{ji} + \beta_{ji}. \end{cases}$$

In this case, the maximum value of the fuzzy number F_{ji} is equal to $m_{ji} + \beta_{ji}$, and the maximum value of the number $F_{ji} - F_{j+1}$ is equal to $m_{ji} = m_{j+1,i} + \beta_{ji} + \alpha_{j+1,i}$. Then the maximum possible value of ξ_j on the set $x_1, x_2, ..., x_m$ is determined by the following relation

$$\xi_{j\max} = \max_x \left[\sum_{i=1}^{m} x_i(m_{ji} - m_{j+1,i} + \beta_{ji} + \alpha_{j+1,i}) \right] \tag{18}$$

Now the problem can be formulated as follows: find a set $(x_1, x_2, ..., x_m)$, satisfying the system of equations

$$\sum_{i=1}^{m} x_i(m_{ji} - m_{j+1} + \beta_{ji} + \alpha_{j+1,i}) + \Delta = 0, \tag{19}$$

$$\sum_{i=1}^{m} x_i = 1, \ \ \Delta > 0. \tag{20}$$

$$x_i \ge 0, \ x_{m+j} \ge 0, \ \ i = 1, 2, .., m, \ \ j = 1, 2, .., n-1. \tag{21}$$

At the same time, the simultaneous fulfillment of conditions (19) and (21) ensures non-positivity of ξ_j, $j = 1, 2, .., n-1.$ In the system of equations (19)-(20), the number of unknowns exceeds the number of equations. Therefore, a preliminary operation of complexing inequalities was carried out to bring the mathematical model of the problem to a canonical form.

Practice shows that uncertainty models formed using standard tools of fuzzy mathematics give an opportunity to obtain an acceptable solution to a large number of possible problems. However, in many real-life situations the nature of uncertainty is more complex than the usual one. In particular, when trying to describe the uncertainty of demand for fuzzy numbers with a triangular membership function, it turns out that the values of the lower and upper boundaries of the uncertainty interval, as well as the modal value of this fuzzy number cannot be determined accurately based on the results of the source data statistical processing. These values, apparently, are also naturally to be considered fuzzy. To correctly describe uncertainty in this case, it is advisable to use fuzzy numbers whose membership function parameters are also defined in a fuzzy way [13]. The resulting numbers are called fuzzy numbers of type 2 (or fuzzy numbers of the second order, or bi-fuzzy numbers). Correct rules for performing arithmetic operations on second-order fuzzy numbers are proposed in [14,15]. Let us present those of them that will be used for the binary analogue of the objective function (19) in the comparator identification problem. Let us introduce a couple of bi-fuzzy numbers F_1 and F_2, which are equal to

$$F_1 = < m_1, \alpha_1, \beta_1 >, \quad F_2 = < m_2, \alpha_2, \beta_2 >$$
$$m_1 = < m_{m_1}, \alpha_{m_1}, \beta_{m_1} >, \quad \alpha_1 = < m_{\alpha_1}, \alpha_{\alpha_1}, \beta_{\alpha_1} >,$$
$$\beta_1 = < m_{\beta_1}, \alpha_{\beta_1}, \beta_{\beta_1} >,$$
$$m_2 = < m_{m_2}, \alpha_{m_2}, \beta_{m_2} >, \quad \alpha_2 = < m_{\alpha_2}, \alpha_{\alpha_2}, \beta_{\alpha_2} >,$$
$$\beta_2 = < m_{\beta_2}, \alpha_{\beta_2}, \beta_{\beta_2} >.$$

Then, in accordance with (14),

$$Z = F_1 - F_2 = < m_z, \alpha_z, \beta_z >,$$
$$m_z = m_1 - m_2 = < m_{m_1}, \alpha_{m_1}, \beta_{m_1} > -$$
$$- < m_{m_2}, \alpha_{m_2}, \beta_{m_2} > = < m_{m_z}, \alpha_{m_z}, \beta_{m_z} >,$$
$$m_{m_z} = m_{m_1} - m_{m_2}, \quad \alpha_{m_z} = \alpha_{m_1} + \beta_{m_2},$$
$$\beta_{m_z} = \beta_{m_1} + \alpha_{m_2};$$
$$\alpha_z = \alpha_1 + \beta_1 = < m_{\alpha_1}, \alpha_{\alpha_1}, \beta_{\alpha_1} > +$$
$$+ < m_{\beta_2}, \alpha_{\beta_2}, \beta_{\beta_2} > = < m_{\alpha_z}, \alpha_{\alpha_z}, \beta_{\alpha_z} >,$$
$$m_{\alpha_z} = m_{\alpha_1} + m_{\beta_2}, \alpha_{\alpha_z} =$$
$$= \alpha_{\alpha_1} + \alpha_{\beta_2}, \beta_{\alpha_z} = \beta_{\alpha_1} + \beta_{\beta_2};$$
$$\beta_z = \beta_1 + \alpha_2 = < m_{\beta_1}, \alpha_{\beta_1}, \beta_{\beta_1} > +$$
$$+ < m_{\alpha_2}, \alpha_{\alpha_2}, \beta_{\alpha_2} > = < m_{\beta_z}, \alpha_{\beta_z}, \beta_{\beta_z} >,$$
$$m_{\beta_z} = m_{\beta_1} + m_{\alpha_2}, \alpha_{\beta_z} =$$
$$= \alpha_{\beta_1} + \alpha_{\alpha_2}, \beta_{\beta_z} = \beta_{\beta_1} + \beta_{\alpha_2}.$$

The resulting relations define the parameters of the membership function of the bi-fuzzy number $\xi_1 = F_1 - F_2$. Relations for bi-fuzzy numbers $\xi_j = F_j - F_{j+1}, \quad j = 1, 2, .., n-1.$ will be written similarly. Let us now write down the result of multiplying the number ξ_1 by the constant x_1. By means of using (14), it is easy to show that when multiplying a bi-fuzzy number by a constant, a bi-fuzzy number will be obtained, all the parameters of the membership functions of which will be equal to the corresponding parameters of the bi-fuzzy factor, multiplied by this constant. Then

$$\Delta_1 = \xi_i x_1 = < m_\Delta, \alpha_\Delta, \beta_\Delta >,$$
$$m_\Delta = < m_{m_1} x_1, \alpha_{m_1} x_1, \beta_{m_1} x_1 >;$$
$$\alpha_\Delta = < m_{\alpha_1} x_1, \alpha_{\alpha_1} x_1, \beta_{\alpha_1} x_1 >;$$
$$\beta_\Delta = < m_{\beta_1} x_1, \alpha_{\beta_1} x_1, \beta_{\beta_1} x_1 >.$$

Finally, we obtain relations for calculating the parameters of the sum of bi-fuzzy numbers $\Delta_i = \xi_i x_i$. At the same time

$$\Delta = \sum_{i=1}^{m} \Delta_i = \sum_{i=1}^{m} \xi_i x_i = < m_\Delta, \alpha_\Delta, \beta_\Delta >,$$
$$\Delta_i = \xi_i x_i = < m_{\Delta_i}, \alpha_{\Delta_i}, \beta_{\Delta_i} >,$$

As far as

$$m_{\Delta_i} = < m_{m_{\Delta_i}} x_i, \alpha_{m_{\Delta_i}} x_i, \beta_{m_{\Delta_i}} x_i >,$$
$$\alpha_{\Delta_i} = < m_{\alpha_{\Delta_i}} x_i, \alpha_{\alpha_{\Delta_i}} x_i, \beta_{\alpha_{\Delta_i}} >,$$
$$\beta_{\Delta_i} = < m_{\beta_{\Delta_i}} x_i, \alpha_{\beta_{\Delta_i}} x_i, \beta_{\beta_{\Delta_i}} >,$$

then

$$m_\Delta =< \sum_{i=1}^{m} m_{m_{\Delta_i}} x_i, \sum_{i=1}^{m} \alpha_{m_{\Delta_i}} x_i, \sum_{i=1}^{m} \beta_{m_{\Delta_i}} x_i >,$$

$$\alpha_\Delta =< \sum_{i=1}^{m} m_{\alpha_{\Delta_i}} x_i, \sum_{i=1}^{m} \alpha_{\alpha_{\Delta_i}} x_i, \sum_{i=1}^{m} \beta_{\alpha_{\Delta_i}} x_i >,$$

$$\beta_\Delta =< \sum_{i=1}^{m} m_{\beta_{\Delta_i}} x_i, \sum_{i=1}^{m} \alpha_{\beta_{\Delta_i}} x_i, \sum_{i=1}^{m} \beta_{\beta_{\Delta_i}} x_i > .$$

Now, similar to the previous one, let's write down the maximum value ξ_j .

$$\xi_{j\max} = \max\left[\sum_{i=1}^{m} m_{m_{\Delta_i}} x_i + \sum_{i=1}^{m} \beta_{m_{\Delta_i}} x_i + \sum_{i=1}^{m} m_{\beta_{\Delta_i}} x_i + \right.$$

$$\left. + \sum_{i=1}^{m} \beta_{\beta_{\Delta_i}} x_i = \max_x \sum_{i=1}^{m} m_{m_{\Delta_i}} + \beta_{m_{\Delta_i}} + m_{\beta_{\Delta_i}} + \beta_{\beta_{\Delta_i}} \right] x_i .$$

Then the problem is formed as follows: find a set $x_i, i = 1, 2, .., m,$ that satisfies the system of equations

$$\sum_{i=1}^{m} x_i \left(m_{m_{\Delta_i}} + \beta_{m_{\Delta_i}} + m_{\beta_{\Delta_i}} + \beta_{\beta_{\Delta_i}} \right) + \Delta = 0, \tag{22}$$

$$\sum_{i=1}^{m} x_i = 1, \quad x_i \ge 0, \quad i = 1, 2, .., m, \quad j = 1, 2, .., n-1 \tag{23}$$

It is clear that joint satisfaction of conditions (22), (23) ensures non-positivity $\xi_i, \ j = 1, 2, .., n-1,$ which is what was required.

Discussing results of solving the problem of comparator identification under conditions of fuzzy initial data. Comparator identification is a method for solving the problem of regression analysis, the purpose of which consists in establishing the dependence of the explained variable on explanatory variables - factors influencing the explained variable. The problem of comparator identification and the method for solving this problem have been developed and applied in cases where, during collection of initial data, it is not possible to measure values of the explained variable, but results of the experiments can be ranked. The work considers an important practical version of the comparator identification problem, when measurements of the value of factors in each experiment are not clearly defined. Solving this problem using traditional methods of regression analysis is impossible. The method proposed in the work reduces the problem to solving a system of linear algebraic equations, the parameters of which are not clearly defined. For the case when all fuzzy parameters are specified on a compact carrier, the method provides a solution to a fuzzy system of linear algebraic equations. The proposed computational procedure has been developed and it can be applied in a situation where uncertainty in the value of factors is described in terms of second-order fuzzy numbers.

A possible direction for further research consists in developing a method for solving the comparator identification problem for the case when the initial data are not specified accurately [15].

Conclusions.

1. A correct formulation of the comparator identification problem is presented for the case when the initial data are specified in terms of fuzzy mathematics. In this connection, to describe the membership functions of the corresponding fuzzy numbers, functions of *(L – R)* type are selected.

2. A criterion for the effectiveness of solving the problem is proposed, based on the calculation of the membership functions of the experimental results.

3. Developed has been a method for solving the problem of comparator identification when the initial data are fuzzy numbers of the *(L – R)* type. The proposed method is generalized to the case when the initial data are second-order fuzzy numbers.

References

[1].Kao. J., Subramaniam K. Combining independent estimators and astimation in linear regression unequal variances// Biometrics,–1997.–vol.27.–p.971.

[2].Petrov, K.E., Komparatorna strukturna ta parametrichna IdentifIkatsIya skalyarnih bagatofaktornih modeley otsInyuvannya [Comparator structural and parametric identification of scalar multifactor estimation models] / K.E. Petrov, V.V. Kryuchkovsky. Kryuchkovsky. - Kherson Oldi-Plus, 2009.-294p.

[3].Petrov E., Petrov K., Komparatorna IdentifIkatsIya bagatofaktornih modeley otsInyuvannya [Comparator identification of multifactor estimation models]/ Palmarium Academic Publishing, 2014.-p.224.

[4].Bondarenko M.F., Shabanov-Kushnarenko Y.P., Sharonova N.V., InstrumentarIy komparatornoyi IdentifIkatsiyi [Comparator identification toolkit]. // Bionics of Intelligence, 2010,№2.p.74-86.

[5].Suprun T.S., Shabanov–Kushnarenko S.Y. Isomorphism of predicate model comparator identification, Radio Electronics, Computer science, Control, 2014.p.14–24.

[6].D. Recordon, M.Rubinstein, N. Mora, G. Lurgin A comparator – based technique for Identification of international electromagnetic interference attacks.–(2014)IEEE International Symposium on Electromagnetic Compatibility, Gothenburg.

[7].Abed A.A. (2018) Functional Structure of comparators predicate in the compartment identification method. Advanced information systems, 2(2), 78–83.
[8].Bondarenko M.F., Shabanov – Kushnarenko S., Shabanov–Kushnarenko Yu.P. (2009), Practical applications of comparative identification of linear finite– dimensional objects, Bionika intellekta, №2 (71), pp.5–12.
[9].Poole, David (2006), Linear Algebra: A Modern Introduction (2nd ed.), Brooks/Cole, ISBN 0-534-99845-3.-pp. 336.
[10]. Anton, Howard (2005), Elementary Linear Algebra (Applications Version) (9th ed.), Wiley International..–pp. 432.
[11]. Leon, Steven J. (2006), Linear Algebra With Applications (7th ed.), Pearson Prentice Hall.– pp.367.
[12]. Raskin L., Sira O., Katkova T. Comparator identification in the conditions of bifuzzy initial data. Eureka. Physics and Engineering, 2021 (1), 113–124.
[13]. Lev Raskin, Oksana Sira. Performing arithmetic operations over the (L–R) Type fuzzy numbers. 2000.–Vol 3. issue 4.–p.6–//Eastern European.
[14]. Lev Raskin, Oksana Sira. Execution of arithmetic operations involving the second order fuzzy numbers//Eastern European Journal of Enterprise Technologies.– 2020.–Vol.4,issue 4.–p.14–20.
Отримана в редакції 26.07.2024. Прийнята до друку 17.08.2024. Received 26 July 2023. Approved 17 August 2024. Available in Internet 23 October 2024

УДК 004.415.2:004.94

WATERFALL MODEL OVER PCD.UCT MODEL REVIEW

МОДЕЛЬ ВОДОПАДУ НАД ОГЛЯДОМ МОДЕЛІ PCD.UCT

NS Wisidagama[1], Faiz Marikkar[2]
NS Wisidagama[1], Faiz Marikkar[2]

[1]Department of Computer Science, Faculty of Computing, General Sir John Kotelawala Defence University,
[2]Staff Development Centre, General Sir John Kotelawala Defence University
E-mail: [2]faiz@kdu.ac.lk

DOI: 10.15673/atbp.v16i3.2927

Abstract. *The software development life cycle (SDLC) has seen the evolution of numerous models, each tailored to meet specific project needs. Among these, the Waterfall Model has been a longstanding, traditional approach, characterized by its linear and sequential phases. In contrast, the PcD.UcT Model, a relatively newer framework, advocates for a more iterative and user-centered approach. This paper presents a comprehensive review of the Waterfall Model, evaluating its strengths and limitations in modern software development. It contrasts these findings with the PcD.UcT Model, analyzing how its principles better align with contemporary project demands, particularly in terms of flexibility, user engagement, and iterative refinement. Through this comparative analysis, the paper aims to provide insights into the applicability and effectiveness of each model, offering guidance for software development teams in choosing the appropriate methodology for their projects. Since the waterfall development methodology, the key requirements in system development strategies have shifted from processes to users. Because the process as well as the customers are both equally important, employing either a predictive or an adaptive methodology is extremely difficult. The primary goal of this work is mainly combining and then evaluating all the relevant methodologies in order to create a HydroGIS that accurately automates the complexities of the hydrology process in a HydroGIS environment while meeting all user requirements.*

Анотація. *Життєвий цикл розробки програмного забезпечення (SDLC) бачив еволюцію численних моделей, кожна з яких адаптована для задоволення конкретних потреб проекту. Серед них модель водоспаду є давнім традиційним підходом, який характеризується лінійними та послідовними фазами. Навпаки, модель PcD.UcT, відносно новіша структура, виступає за більш ітеративний і орієнтований на користувача підхід. Ця стаття представляє комплексний огляд моделі водоспаду, оцінюючи її сильні сторони та обмеження в сучасній розробці програмного забезпечення. Він порівнює ці висновки з моделлю PcD.UcT, аналізуючи, як її принципи краще узгоджуються з вимогами сучасного проекту, зокрема з точки зору гнучкості, залучення користувачів та ітераційного вдосконалення. Завдяки цьому порівняльному аналізу ця стаття має на меті надати розуміння застосовності та ефективності кожної моделі, пропонуючи керівництво для команд розробників програмного*