On linear regression for fuzzy data of different quality

Serhii Mashchenko[a], Oleksandr Marchenko[a]

[a] *Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska Street, City of Kyiv, 01601, Ukraine*

Abstract

The present paper is devoted to the linear regression for a fuzzy set of fuzzy data samples. This model allows one to take into account the data of different quality. It is shown that regression parameters are type-2 fuzzy sets. Furthermore, the corresponding type-2 membership functions are given. The decomposition approach is used to investigate the T2FSs of linear regression parameters. It is shown that each T2FS of regression parameter can be decomposed according to secondary membership grades into a finite collection of fuzzy numbers. Each of these fuzzy number is the corresponding fuzzy regression parameter for a set of data numbers. This set is the corresponding α-cut of the original fuzzy set of fuzzy data samples. The illustrative example is given.

Keywords [1]

Linear regression, fuzzy least squares estimator, fuzzy number, type-2 fuzzy set.

1. Introduction

The classical regression analysis is based on crisp data and a crisp relationship between the dependent variable and the independent variables. In practice, there are many situations in which observations cannot be measured as crisp quantities, because the information is often fuzzy, incomplete, linguistic or noisy. Fuzzy regression analysis is a non-statistical method based on a fuzzy set theory rather than probability theory (see [1]). In a general model of a fuzzy regression both input and output are fuzzy. In this regard, the fuzzy regression model contains fuzzy parameters instead of error terms.

The three main fields can be distinguished in fuzzy regression analysis. These are possibilistic regression analysis, fuzzy least squares methods and machine learning techniques. The probabilistic approach in fuzzy regression analysis was first proposed by Tanaka et al. [2]. Unlike conventional regression analysis, where deviations between observed and predicted values reflect a measurement error, deviations in fuzzy regression reflect the uncertainty of the system structure expressed by fuzzy parameters of the regression model. Fuzzy parameters of the model are considered to be distributions of possibilities and determined by solving a linear programming problem that allows one to minimize fuzzy deviations subject to membership degrees constraints. Since the membership functions (MFs) of fuzzy sets (FSs) can be viewed as probability distributions, this approach was called ‘possibilistic regression analysis.’ The possibilistic approach was explored and improved by many authors. Reviews of possibilistic regression analysis can be found in D’Urso [3].

Fuzzy regression analysis was also considered from the viewpoint of generalizing the classical least squares method to the case of fuzzy data. The idea of this approach is to minimize in some sense the different distance measures between the predicted fuzzy values and the given fuzzy data. Fuzzy least squares methods were first proposed by Celmins [4]. Later, this approach was significantly developed by many researchers. In [1] one can find qualitative reviews on the fuzzy least squares and fuzzy least absolute methods.

The Sixth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2023), May 3, 2023, Zaporizhzhia, Ukraine
EMAIL: s.o.mashchenko@gmail.com (S. Mashchenko); rozenkrans17@gmail.com (O. Marchenko)
ORCID: 0000-0003-4863-2763 (S. Mashchenko); 0000-0002-5408-5279 (O. Marchenko)

Machine learning techniques made it possible to generalize fuzzy regression analysis through the use of genetic algorithms, neural networks, and support vector machines. Relevant references can be found in Chukhrova and Johannssen [1] and Hastie et al. [5].

Often, when solving applied problems, data of different quality can be used [6]. For example, according to [7], wind tunnel experiments provide high simulation accuracy (a source of high-fidelity data). Also, experiments based on computational physical models have a higher error (source of low fidelity data). In some applications, a significantly more accurate regression model can be built if low-precision data are also used. In this case, the problem arises of constructing a regression based on data of different quality.

Using data of different quality with the aim of improving model accuracy is not a new concept. For instance, Hevesi et al. [8] predict average annual precipitation values near a potential nuclear waste disposal site using a set of precipitation measurements from the region along with more easily obtainable elevation map of the area. Kennedy and O'Hagan [9] approach the subject from the perspective of model construction using data resulting from computational simulations of varying fidelities and costs. In [7], to construct a Gaussian regression model, the problem of planning an experiment is solved with the choice of the ratio between the sizes of samples of low-precision and high-precision data. Also, to process data with a variable degree of certainty the methods of transfer learning [10], space mapping [11], and others [12] are used.

Data quality (degree of usability of data) is a complex concept. It is characterized by objectivity, integrity, relevance, measurability, controllability, etc. In some cases, the quality of data may not be crisp defined [13]. According to [14], where using data without expert knowledge, the choice of a representative sample becomes an NP-complete problem. Therefore, samples have to be found within a reasonable time, and this justifies the use of fuzzy methods that formalize expert knowledge expressed in natural language words. For instance, in the framework of quality control fuzzy expert assessments are used in [15] to construct acceptance sampling plans (how many units can be selected from a consignment and how many defective units are allowed in this sample).

In this article, we intend to investigate the method of constructing a fuzzy linear regression in the case when fuzzy data samples are of different qualities. Furthermore, the degrees of membership to a FS are known for these samples. This leads to a possibility that the regression takes into account fuzziness of the quality assessments of different samples, rather than just uncertainty of data. Examples of such FSs data samples could be: 'High quality data samples', 'Questionable data samples', 'Actual data samples', etc.

The main result of the article justifies the fact that a FS of different quality samples of type-1 fuzzy data generates a type-2 fuzzy regression model. In this model, the regression parameters are T2FSs on the real line with constant secondary grades. Although, in general, a T2FS is a rather complicated mathematical object, T2FSs with constant secondary grades are simple enough for practical use. This feature allows us to decompose this set by secondary grades into a collection of corresponding fuzzy numbers. Each of them represents the corresponding fuzzy regression parameter for a crisp set of data samples. The set in the focus is the α-cut of the original fuzzy set of data samples. We note that the well-known type 2 fuzzy regression models use crisp collections of type 2 data sets, while the model proposed in the article uses a fuzzy collection of type 1 fuzzy data samples. This is a principal difference between them. It should also be added that this article continues the line of research in the field of mathematical operations with a fuzzy set of operands, first introduced in the context of intersections and unions of fuzzy sets [16, 17].

2. Materials and Methods

In this section, we briefly review some existing theories and definitions.

2.1. Linear regression analysis for fuzzy input and output data using the extension principle

The article focuses on a linear regression in the case when data samples form a FS. We stress that one could have exploited different known methods of constructing a linear regression for fuzzy data. As an alternative, we intend to modify the method [18] based on the extension principle.

Let $K=\{1,...,|K|\}$ be the set of indices of data samples $\{y_i, x_{i1},...,x_{ip}\}$, $i\in K$, where $|K|$ is the cardinality of K. A crisp statistical linear regression has the form

$$y_i = x_i\,(\varphi(K))^T + \varepsilon_i,\; i\in K, \tag{1}$$

where for each $i\in K$, y_i is the dependent variable; $x_i=(1,x_{i1},...,x_{ip})$ is the vector of independent variables (factors, regressors) x_{il}, $l=1,...,p$; ε_i is the independent normal random variable. The symbol ^{T} denotes the transpose, p is the number of independent variables, $\varphi(K)=(\varphi_0(K),...,\varphi_p(K))=(\varphi_l(K))_{l=0,...,p}$ is the vector of regression parameters. Let $y(K)=X(K)(\varphi(K))^T+\varepsilon(K)$ be the matrix notation of equations (1) with $X(K)=\{x_{is}\}_{i\in K, s=0,...p}$ and $x_{i0}=1$ for all $i\in K$, $y(K)=(y_i)^T_{i\in K}$ and $\varepsilon(K)=(\varepsilon_i)^T_{i\in K}$. For the convenience of presentation, the set K of sample indices is indicated hereinafter as a parameter in these formulae. According to the least squares method, the estimate $\hat\varphi(K)=(\hat\varphi_l(K))_{l=0,...,p}$ of the parameter vector $\varphi(K)=(\varphi_l(K))_{l=0,...,p}$ has the form

$$\hat\varphi(K)=[X^T(K)X(K)]^{-1}X^T(K)y(K). \tag{2}$$

Assume that the data is fuzzy. To generalize formula (2) we denote by

$$f_l(X(K),y(K))=\hat\varphi_l(K) \tag{3}$$

the l-th element, $l=0,...,p$ of the estimate $\hat\varphi(K)$. We also denote by

$$\tilde X(K)=\{\tilde x_{is}\}_{i\in K, s=0,...p},\; \tilde y(K)=(\tilde y_i)^T_{i\in K} \tag{4}$$

the matrix of independent variables and the vector of dependent variables, respectively, where for each $i\in K$, $\tilde x_{i0}=1$ is the crisp number which is equal to 1; $\tilde x_{is}$, $s=1,...,p$ and $\tilde y_i$ are fuzzy numbers (FNs) with the MFs $\mu_{\tilde x_{is}}(x_{is})$, $x_{is}\in\mathbb{R}$, $s=1,...,p$ and $\mu_{\tilde y_i}(y_i)$, $y_i\in\mathbb{R}$, $i\in K$, respectively. Here, $\mathbb{R}$ is the real line.

Remark 1. Recall that a FN is a normal FS on $\mathbb{R}$ with the upper semicontinuous and quasi-concave MF (for example, see [19]).

The vector $\tilde\varphi(K)=(\tilde\varphi_l(K))_{l=0,...,p}$ of fuzzy parameters of the regression has the form $\tilde\varphi(K)=[\tilde X^T(K)\tilde X(K)]^{-1}\tilde X^T(K)\tilde y(K)$ according to the least squares method [18]. For each $l=0,...,p$, $\tilde\varphi_l(K)=f_l(\tilde X(K),\tilde y(K))=\{(r,\mu_{\tilde\varphi_l(K)}(r)): r\in\mathbb{R}\}$ is the FN with the MF

$$\begin{aligned}\mu_{\tilde\varphi_l(K)}(r)=&\max_{X(K),y(K)}\{\min_{i\in K, s=0,...,p}\{\mu_{\tilde x_{is}}(x_{is}),\mu_{\tilde y_i}(y_i)\}:\; r=f_l(X(K),y(K)),\\ &X(K)=\{x_{is}\}_{i\in K; s=0,...,p}\in\mathbb{R}^{p|K|}, y(K)=\{y_i\}_{i\in K}\in\mathbb{R}^{|K|}\}\end{aligned} \tag{5}$$

by Zadeh's extension principle [20]. Here, $f_l(X(K),y(K))$ is the *l*-th element of the vector $\hat\varphi(K)=[X^T(K)X(K)]^{-1}X^T(K)y(K)$ by (2) and (3). According to Remark 1, the maximum in (5) exists. As shown in [21], the representation of FNs by u-cuts is simpler for calculations than the functional approach. Therefore, for each $l=0,...,p$, we represent the MF of the FN $\tilde\varphi_l(K)$ in the form

$$\mu_{\tilde\varphi_l(K)}(r)=\max_{u\in[0,1]} u 1_{[\tilde\varphi_l(K)]_u}(r), \tag{6}$$

where $[\tilde\varphi_l(K)]_u$ is the u-cut of the FN $\tilde\varphi_l(K)$. This u-cut is the set $[\tilde\varphi_l(K)]_u=\{r\in\mathbb{R}: \mu_{\tilde\varphi_l(K)}(r)\ge u\}$ with the MF

$$1_{[\tilde\varphi_l(K)]_u}(r)=\begin{cases}1, & r\in[\tilde\varphi_l(K)]_u;\\ 0, & r\notin[\tilde\varphi_l(K)]_u;\end{cases} \tag{7}$$

$r\in\mathbb{R}$, $u\in[0,1]$. According to [18] and Remark 1, formula (5) implies that u-cut $[\tilde\varphi_l(K)]_u$ of the FN $\tilde\varphi_l(K)$ has the form

$$[\tilde{\varphi}_l(K)]_u = \{f_l(X(K), y(K)) : x_{is} \in [\tilde{x}_{is}]_u, s = 0,..., p;\ y_i \in [\tilde{y}_i]_u,\ i \in K\}, \tag{8}$$

where for each $i \in K$, the u-cuts of the FNs $\tilde{x}_{is}$, $s = 0,...p$ and $\tilde{y}_i$ are closed intervals

$$[\tilde{x}_{il}]_u = [[\tilde{x}_{is}]_u^D, [\tilde{x}_{is}]_u^H],\ s = 0,..., p \text{ and } [\tilde{y}_i]_u = [[\tilde{y}_i]_u^D, [\tilde{y}_i]_u^H], \tag{9}$$

respectively, of the real line $\mathbb{R}$. Since $\tilde{\varphi}_l(K)$ is a FN, then its u-cut $[\tilde{\varphi}_l(K)]_u$ is the interval $[\tilde{\varphi}_l(K)]_u = [[\tilde{\varphi}_l(K)]_u^D, [\tilde{\varphi}_l(K)]_u^H] \subset \mathbb{R}$ too. Equality (8) entails

$$[\tilde{\varphi}_l(K)]_u^D = \min\{f_l(X(K), y(K)) :\ x_{is} \in [\tilde{x}_{is}]_u, s = 0,..., p;\ \ y_i \in [\tilde{y}_i]_u,\ i \in K\}, \tag{10}$$

$$[\tilde{\varphi}_l(K)]_u^H = \max\{f_l(X(K), y(K)) :\ x_{is} \in [\tilde{x}_{is}]_u,\ , s = 0,..., p;\ y_i \in [\tilde{y}_i]_u,\ i \in K\}. \tag{11}$$

Thus, (6) ensures that formula (5) and the FN $\tilde{\varphi}_l(K)$ have the forms

$$\mu_{\tilde{\varphi}_l(K)}(r) = \max\{u \in [0,1] :\ [\tilde{\varphi}_l(K)]_u^D \le r \le [\tilde{\varphi}_l(K)]_u^H\},\ r \in \mathbb{R}, \tag{12}$$

$$\tilde{\varphi}_l(K) = \{(r,u) : r \in [[\tilde{\varphi}_l(K)]_u^L, [\tilde{\varphi}_l(K)]_u^H], u \in [0,1]\} = \{([[\tilde{\varphi}_l(K)]_u^L, [\tilde{\varphi}_l(K)]_u^H], u) : u \in [0,1]\},$$

respectively. Problems (10) and (11) are rather complicated. In view of this, it is suggested in [18] to use, for $l = 0,..., p$, the approximate value of the l-th fuzzy parameter

$$\tilde{\beta}_l(K) = \{([[\tilde{\beta}_l(K)]_u^L, [\tilde{\beta}_l(K)]_u^H], u) : u \in [0,1]\}, \tag{13}$$

with the MF

$$\mu_{\tilde{\beta}_l(K)}(r) = \max\{u \in [0,1] :\ [\tilde{\beta}_l(K)]_u^D \le r \le [\tilde{\beta}_l(K)]_u^H\},\ r \in \mathbb{R}, \tag{14}$$

where

$$\begin{aligned}[\tilde{\beta}_l(K)]_u^D = \min\{&f_l([\tilde{X}(K)]_u^D, [\tilde{y}(K)]_u^D), f_l([\tilde{X}(K)]_u^D, [\tilde{y}(K)]_u^H, \\ &f_l([\tilde{X}(K)]_u^H, [\tilde{y}(K)]_u^D), f_l([\tilde{X}(K)]_u^H, [\tilde{y}(K)]_u^H)\},\end{aligned} \tag{15}$$

$$\begin{aligned}[\tilde{\beta}_l(K)]_u^H = \max\{&f_l([\tilde{X}(K)]_u^D, [\tilde{y}(K)]_u^D), f_l([\tilde{X}(K)]_u^D, [\tilde{y}(K)]_u^H, \\ &f_l([\tilde{X}(K)]_u^H, [\tilde{y}(K)]_u^D), f_l([\tilde{X}(K)]_u^H, [\tilde{y}(K)]_u^H)\}.\end{aligned} \tag{16}$$

Here, the matrices $[\tilde{X}(K)]_u^D$, $[\tilde{X}(K)]_u^H$ and the vectors $[\tilde{y}(K)]_u^D$, $[\tilde{y}(K)]_u^H$ are comprised in of the elements $[\tilde{x}_{is}]_u^D, [\tilde{x}_{is}]_u^H$, $s = 0,..., p;\ [\tilde{y}_i]_u^D, [\tilde{y}_i]_u^H$, $i \in K$, respectively (see (9)). It is clear that $[\tilde{\varphi}_l(K)]_u^D \le [\tilde{\beta}_l(K)]_u^D$, $[\tilde{\varphi}_l(K)]_u^H \ge [\tilde{\beta}_l(K)]_u^H$. Therefore, the inclusion $[\tilde{\beta}_l(K)]_u \subseteq [\tilde{\varphi}_l(K)]_u$ is hold for the closed interval $[\tilde{\beta}_l(K)]_u = [[\tilde{\beta}_l(K)]_u^D, [\tilde{\beta}_l(K)]_u^H]$. This entails $\mu_{\tilde{\beta}_l(K)}(r) \le \mu_{\tilde{\varphi}_l(K)}(r)$ and thereupon the value $\mu_{\tilde{\beta}_l(K)}(r)$ is a lower bound of $\mu_{\tilde{\varphi}_l(K)}(r)$ for $r \in \mathbb{R}$.

2.2. Type-2 fuzzy sets

Zadeh [22] introduced the notion of type-2 fuzzy set (T2FS) as a generalization of type-1 fuzzy set (T1FS) (that is, an ordinary FS). Unlike a T1FS, the membership degree of elements in a T2FS is a FS on closed interval [0, 1]. Based on the ideas of Karnik and Mendel [23], Mendel and John [24] gave a different definition. The T2FS $\tilde{C}$ on a set X is the collection

$$\tilde{C} = \{((x,u), \eta_{\tilde{C}}(x,u)) :\ x \in X,\ u \in J_x \subseteq [0,1]\}, \tag{17}$$

where $\eta_{\tilde{C}}(x,u)$ is a type-2 membership function (T2MF), J_x is the set of primary membership degrees u of $x \in X$ to $\tilde{C}$. The value $\eta_{\tilde{C}}(x,u)$ is a crisp number from the closed interval [0, 1] which is called the secondary grade of the pair (x,u).

Remark 2. According to the comments of Harding et al. [25] and Aisbett et al. [26], one has to define the T2MF $\eta_{\tilde{C}}(x,u)$ for all $x \in X$ and $u \in [0,1]$. To this end, one should put $\eta_{\tilde{C}}(x,u) = 0$ for all $u \notin J_x$, $x \in X$.

Remark 3. The primary membership degree *u* is deemed as the degree of manifestation of some property (which determines the given FS) of $x \in X$. According to [27], we interpret the secondary grade $\eta_{\tilde{C}}(x,u)$ as the degree of truth of the corresponding primary degree *u* of this property for *x*.

Following [24], we define embedded T2FSs and T1FSs for a T2FS

$\tilde{C} = \{((x,u), \eta_{\tilde{C}}(x,u)) : x \in X, u \in [0,1]\}$. Assume that $u_x = \mu_{C^{e1}}(x) \in [0,1]$ is a unique primary degree of membership, for each $x \in X$, where $\mu_{C^{e1}}(x)$, $x \in X$ is the MF of the T1FS $C^{el} = \{(x, \mu_{C^{e1}}(x)) : x \in X\}$. This T1FS is called embedded in the T2FS $\tilde{C}$. We define the embedded T2FS $\tilde{C}^{e2}$ in $\tilde{C}$ in the form $\tilde{C}^{e2} = \{((x,u_x), \eta_{\tilde{C}^{e2}}(x,u_x)) : x \in X\}$ with $\eta_{\tilde{C}^{e2}}(x,u_x)) = \eta_{\tilde{C}}(x, \mu_{C^{e1}}(x)))$ for all $x \in X$.

Remark 4. According to [24], each element of the type-2 fuzzy collection $\tilde{C} = \{((x,u), \eta_{\tilde{C}}(x,u)) : x \in X, u \in [0,1]\}$ is interpreted as a subset. Thus, the collection is represented as the classical union of its elements in the sense of T1FSs. In [24], Mendel and John stated that each T2FS can be represented as a collection of all possible embedded T2FSs.

2.3. Collections of T2FSs with constant secondary grades

We shall need one special case of a T2FS to be defined according to [28, 29, 30]. Let $A = \{\eta_{\tilde{C}}(x,u) : \eta_{\tilde{C}}(x,u) > 0, x \in X, u \in [0,1]\}$ be the set of all possible positive values of secondary grades for the T2FS $\tilde{C} = \{((x,u), \eta_{\tilde{C}}(x,u)) :\ x \in X,\ u \in [0,1]\}$. Assume that the set A is finite.

Definition 1. We say that an embedded T2FS $\tilde{C}^{e2}(\alpha) = \{((x,u_x), \eta_{\tilde{C}^{e2}(\alpha)}(x,u_x)) : x \in X\}$ in the T2FS $\tilde{C}$ has a constant secondary grade $\alpha \in A$ if, for each $x \in X$, the unique primary degree $u_x = \mu_{C^{e1}(\alpha)}(x) \in [0,1]$ exists for which $\eta_{\tilde{C}^{e2}(\alpha)}(x,u_x) \equiv \alpha$, i.e. $\tilde{C}^{e2}(\alpha) = \{((x, \mu_{C^{e1}(\alpha)}(x)), \alpha) :\ x \in X\}$.

Here $\mu_{C^{e1}(\alpha)}(x)$, $x \in X$ is the MF of the embedded T1FS $C^{el}(\alpha) = \{(x, \mu_{C^{e1}(\alpha)}(x)) : x \in X\}$ in the T2FS $\tilde{C}$.

Remark 5. Obviously, for the T2FS $\tilde{C}$ and each $\alpha \in A$, there is the unique embedded T1FS $C^{el}(\alpha) = \{(x, \mu_{C^{e1}(\alpha)}(x)) :\ x \in X\}$ which is corresponding to the embedded T2FS $\tilde{C}^{e2}(\alpha)$ with constant secondary grade α. Hence, $\tilde{C}^{e2}(\alpha) = \{(C^{el}(\alpha), \alpha)\} = \{(\{(x, \mu_{C^{e1}(\alpha)}(x)) :\ x \in X\}, \alpha)\} =$
$= \{((x, \mu_{C^{e1}(\alpha)}(x)), \alpha) :\ x \in X\}$.

3. Formulation of the problem and the main idea

Let $N = \{1,...,n\}$ be the set of indices of fuzzy data samples $\{\tilde{y}_i, \tilde{x}_{i1}, ..., \tilde{x}_{ip}\}$, $i \in N$ in the form of FNs with the MFs $\mu_{\tilde{x}_{is}}(x_{is})$, $x_{is} \in \square$, $s = 1,...,p$, $\mu_{\tilde{y}_i}(y_i)$, $y_i \in \square$, $i \in N$, respectively. The matrix $\tilde{X}(N)$ of independent variables and the vector $\tilde{y}(N)$ of dependent variables are given by formula (4). Assume that qualities of data samples $\{\tilde{y}_i, \tilde{x}_{i1}, ..., \tilde{x}_{ip}\}$, $i \in N$ are different. Furthermore, the degrees $\mu_I(j), j \in N$ of membership to the FS $I = \{(j, \mu_I(j)) : j \in N\}$ of data samples indices are known. The following question arises: 'what are linear regression parameters in the case when fuzzy data samples are involved in the calculation with the corresponding degrees $\mu_I(j)$, $j \in N$ of membership?' In other words: 'what are the fuzzy parameters of the regression for the FS I of the data samples indices? We investigate this problem for the l-th regression parameter. First, we generalize formula (12) for the case of an arbitrary subset $K \subseteq N$ of sample indices and represent it in a convenient form for us. For each $r \in \square$, we consider the mapping $U_l^r : 2^N \to [0,1]$ given by

$$U_l^r(K) = \max\{u \in [0,1] :\ [\tilde{\beta}_l(K)]_u^D \le r \le [\tilde{\beta}_l(K)]_u^H\},\ K \subseteq N. \tag{18}$$

According to (12), the mapping U_l^r associates each subset $K \subseteq N$ of data sample indices with the value of the MF $\mu_{\tilde{\beta}_l(K)}(r)$ of the fuzzy parameter $\tilde{\beta}_l(K)$ (see (13)). The latter is the FN

$$\tilde{\beta}_l(K)=\{(r,\mu_{\tilde{\beta}_l(K)}(r)):r\in\mathbb{R}\}, \tag{19}$$

with the MF

$$\mu_{\tilde{\beta}_l(K)}(r)=U_l^r(K),\ \ r\in\operatorname{supp}(\tilde{\beta}_l(K))=\{r\in\mathbb{R}:U_l^r(K)\neq 0\}, \tag{20}$$

where $\operatorname{supp}(\tilde{\beta}_l(K))$ is the support of the FN $\tilde{\beta}_l(K)$. Next, we generalize formulae (19) and (20) to the case of the FS I of sample indices. We denote by $\tilde{B}_l$ the l-th regression parameter, and by $M_{\tilde{B}_l}(r)$, $r\in\mathbb{R}$ its MF for the FS I of data samples indices. In this case, the value of the MF $M_{\tilde{B}_l}(r)$ for each fixed $r=r^*$ coincides with the image $U_l^{r^*}(I)$ of the FS I under $U_l^{r^*}$, i.e. $M_{\tilde{B}_l}(r^*)=U_l^{r^*}(I)$. According to Zadeh's extension principle [20], the image of the FS I under the mapping $U_l^{r^*}:2^N\to[0,1]$ (see (18)) is the FS $U_l^{r^*}(I)=\{(u,\mu_{U_l^{r^*}(I)}(u)):u\in[0,1]\}$ with the MF

$$\mu_{U_l^{r^*}(I)}(u)=\max\{\alpha\in[0,1]:\ u=U_l^{r^*}(I(\alpha))\},\ u\in[0,1]. \tag{21}$$

Here, $I(\alpha)=\{j\in N:\mu_I(j)\geq\alpha\}$ is the α-cut, $\alpha\in[0,1]$ of the FS $I=\{(j,\mu_I(j)):\ j\in N\}$ of sample indices;

$$U_l^{r^*}(I(\alpha))=\mu_{\tilde{\beta}_l(I(\alpha))}(r^*), \tag{22}$$

is the image of the α-cut $I(\alpha)$, $\alpha\in[0,1]$ of the FS I of the samples indices in the mapping $U_l^{r^*}$ (see (18)). The value $U_l^{r^*}(I(\alpha))$ is equal to the MF value $\mu_{\tilde{\beta}_l(I(\alpha))}(r^*)$ of the l-th fuzzy parameter $\tilde{\beta}_l(I(\alpha))$ for the set $I(\alpha)$ of sample indices.

Remark 6. Let $A=\{\mu_I(j):\ j\in N\}$ be the set of membership degrees values of the fuzzy set $I=\{(j,\mu_I(j)):\ j\in N\}$ of sample indices. Note that the cardinality of the set A is |A| ≤ n. The situation |A| < n may occur if the degrees of membership $\mu_I(j)$ coincide for different indices $j\in N$ of samples. It is clear that while obtaining α-cuts $I(\alpha)=\{j\in N:\mu_I(j)\geq\alpha\}\neq\varnothing$ of the fuzzy set I we can assume that $\alpha\in A$ rather than $\alpha\in[0,1]$.

Thus, according to (20) and (21), for fixed $r=r^*$, values of $M_{\tilde{B}_l}(r^*)$ form the T1FS $\{(u,\mu_{M_{\tilde{B}_l}(r^*)}(u)):u\in[0,1]\}$ on [0, 1] with the MF $\mu_{M_{\tilde{B}_l}(r^*)}(u)=\mu_{U_l^{r^*}(I)}(u)=\max\{\alpha\in A:u=U_l^{r^*}(I(\alpha))\}$, $u\in\operatorname{supp}(M_{\tilde{B}_l}(r^*))=\{u\in[0,1]:\ u=U_l^{r^*}(I(\alpha)),\ \alpha\in A\}$. Then (22) entails

$$\mu_{M_{\tilde{B}_l}(r^*)}(u)=\max\{\alpha\in A:\ u=\mu_{\tilde{\beta}_l(I(\alpha))}(r^*)\}, \tag{23}$$

where $u\in\operatorname{supp}(M_{\tilde{B}_l}(r^*))=\{u\in[0,1]:\ u=\mu_{\tilde{\beta}_l(I(\alpha))}(r^*),\ \alpha\in A\}$. Therefore, we conclude that $\tilde{B}_l$ is a FS on $\mathbb{R}$ with the MF whose values form T1FS on $[0,1]$. Then, according to [22], $\tilde{B}_l$ is the T2FS on $\mathbb{R}$. In the manner of vertical slices [27] the T2FS $\tilde{B}_l$ on $\mathbb{R}$ has the form:

$$\tilde{B}_l=\{(r,M_{\tilde{B}_l}(r)):r\in R\}=\{(r,\{(u,\mu_{M_{\tilde{B}_l}(r)}(u)):\ u\in J_r\}):\ r\in\mathbb{R}\}, \tag{24}$$

where $\mu_{M_{\tilde{B}_l}(r)}(u)$, $u\in[0,1]$ is the MF of the T1FS $M_{\tilde{B}_l}(r)=\{\{(u,\mu_{M_{\tilde{B}_l}(r)}(u)):\ u\in[0,1]\}$ of values of fuzzy degree of membership of the element $r\in\mathbb{R}$ to the T2FS $\tilde{B}_l$ and $J_r=\operatorname{supp}(M_{\tilde{B}_l}(r))$ is the set of primary membership degrees, where $\operatorname{supp}(M_{\tilde{B}_l}(r))$ is the support of the T1FS $M_{\tilde{B}_l}(r)$ for $r\in\mathbb{R}$. According to Section 2.2, we can also characterize the T2FS $\tilde{B}_l$ of the l-th regression parameter by means of the T2MF $\eta_{\tilde{B}_l}(r,u)=\mu_{M_{\tilde{B}_l}(r)}(u),\ u\in J_r$ and $\eta_{\tilde{B}_l}(r,u)=0,\ u\notin J_r$. This conclusion allows us to introduce the following notion.

Definition 2. By the regression parameter with index $l=0,...,p$ for the FS I of sample indices is meant the T2FS

$$\tilde{B}_l=\{((r,u),\eta_{\tilde{B}_l}(r,u)):\ u\in[0,1],\ r\in\mathbb{R}\}, \tag{25}$$

with the T2MF

$$\eta_{\tilde{B}_l}(r,u)=\begin{cases}\max\{\alpha\in A:\ u=\mu_{\tilde{\beta}_l(I(\alpha))}(r)\},\ u\in J_r;\\ 0,\qquad\qquad u\notin J_r.\end{cases} \tag{26}$$

Here,

$$J_r=\{u\in[0,1]:\ u=\mu_{\tilde{\beta}_l(I(\alpha))}(r),\ \alpha\in A\}, \tag{27}$$

is the set of primary membership degrees $u\in[0,1]$ with strictly positive secondary grades $\eta_{\tilde{B}_l}(r,u)$ which coincides with the support $\mathrm{supp}(M_{\tilde{B}_l}(r))$ (see (23)) of the T1FS $M_{\tilde{B}_l}(r)$ of fuzzy membership degrees of the element $r\in\mathbb{R}$;

$$\mu_{\tilde{\beta}_l(I(\alpha))}(r)=\max\{u\in[0,1]:\ [\tilde{\beta}_l(I(\alpha))]_u^D\le r\le[\tilde{\beta}_l(I(\alpha))]_u^H\}, \tag{28}$$

is the MF of the l-th fuzzy parameter

$$\tilde{\beta}_l(I(\alpha))=\{(r,\mu_{\tilde{\beta}_l(I(\alpha))}(r)):r\in\mathbb{R}\}, \tag{29}$$

for the set $I(\alpha)$ of sample indices (see (19)-(20) for $K=I(\alpha)$);

$$\begin{aligned}[\tilde{\beta}_l(I(\alpha))]_u^D=\min\{&f_l([\tilde{X}(I(\alpha))]_u^D,[\tilde{y}(I(\alpha))]_u^D),f_l([\tilde{X}(I(\alpha))]_u^D,[\tilde{y}(I(\alpha))]_u^H,\\ &f_l([\tilde{X}(I(\alpha))]_u^H,[\tilde{y}(I(\alpha))]_u^D),f_l([\tilde{X}(I(\alpha))]_u^H,[\tilde{y}(I(\alpha))]_u^H)\}\end{aligned} \tag{30}$$

and

$$\begin{aligned}[\tilde{\beta}_l(I(\alpha))]_u^H=\max\{&f_l([\tilde{X}(I(\alpha))]_u^D,[\tilde{y}(I(\alpha))]_u^D),f_l([\tilde{X}(I(\alpha))]_u^D,[\tilde{y}(I(\alpha))]_u^H,\\ &f_l([\tilde{X}(I(\alpha))]_u^H,[\tilde{y}(I(\alpha))]_u^D),f_l([\tilde{X}(I(\alpha))]_u^H,[\tilde{y}(I(\alpha))]_u^H)\}\end{aligned} \tag{31}$$

are the estimates of the lower and upper bounds (see (15)-(16) for $K=I(\alpha)$) of the u-cut $[\tilde{\beta}_l(I(\alpha))]_u$ of the FN $\tilde{\beta}_l(I(\alpha))$;

$$I(\alpha)=\{j\in N:\mu_I(j)\ge\alpha\} \tag{32}$$

is the α-cut of the FS I of sample indices;

A is the set of the membership degrees values $\mu_I(j), j\in N$ of the FS $I=\{(j,\mu_I(j)):j\in N\}$ of sample indices (see Remark 6). According to (26), the set A includes all possible positive values of secondary grades for the T2FS $\tilde{B}_l$ of the l-th regression parameter.

4. Regression for a fuzzy set of sample indices

4.1. Decomposition of T2FSs of regression parameters

For each l-th regression parameter, $l=0,...,p$, we apply a decomposition approach to represent the T2FS $\tilde{B}_l$ in a more convenient form. Theorem 1 justifies the representation of the T2FS $\tilde{B}_l$ in the form of a collection of the embedded T2FSs with constant secondary grades.

Theorem 1. The T2FS $\tilde{B}_l$ of the l-th regression parameter is represented in the form of the collection $\tilde{B}_l=\{\tilde{B}_l^{e2}(I(\alpha)):\ \alpha\in A\}$ of embedded T2FSs

$$\tilde{B}_l^{e2}(I(\alpha))=\{(\tilde{\beta}_l(I(\alpha)),\alpha)\}, \tag{33}$$

with the constant secondary grades $\alpha\in A$. For each $\alpha\in A$, the embedded T1FS $\tilde{\beta}_l(I(\alpha))=\{(r,\mu_{\tilde{\beta}_l(I(\alpha))}(r)):r\in\mathbb{R}\}$ is the FN which is the l-th fuzzy parameter for the set $I(\alpha)=\{j\in N:\mu_I(j)\ge\alpha\}$ of sample indices, with the MF $\mu_{\tilde{\beta}_l(I(\alpha))}(r)$ in form (28).

Proof. According to (25), the T2FS of the regression parameter with the index $l=0,...,p$ has the form $\tilde{B}_l=\{((r,u),\mu_{\tilde{B}_l}(r,u)):\ u\in[0,1],\ r\in\mathbb{R}\}$. Hence,

$$\tilde{B}_l=\{\{((r,u),\max\{\alpha\in A:\ u=\mu_{\tilde{\beta}_l(I(\alpha))}(r)\}):\ u\in J_r\}\bigcup\{((r,u),0):\ u\notin J_r\}:\ r\in\mathbb{R}\}, \tag{34}$$

by (26). Remark 4 allows us to ignore the pairs (r,u) that have secondary grades equal to 0. Thus,

$$\tilde{B}_l = \{((r,u), \max\{\alpha \in A:\ u = \mu_{\tilde{\beta}_l(I(\alpha))}(r)\}):\ u \in J_x,\ r \in \mathbb{R}\}, \tag{35}$$

and thereupon $\tilde{B}_l = \{((r, \mu_{\tilde{\beta}_l(I(\alpha))}(r)):\ \alpha \in A):\ r \in \mathbb{R}\}$ by (27). Note that the collection $\{(\mu_{\tilde{\beta}_l(I(\alpha))}(r), \alpha):\ \alpha \in A\}$ is the T1FS which is formed by the unique value $u = \mu_{\tilde{\beta}_l(I(\alpha))}(r)$ of the fuzzy degree of membership of $r \in \mathbb{R}$. The different values $\alpha \in A$ may correspond to $u = \mu_{\tilde{\beta}_l(I(\alpha))}(r)$. Therefore, the equality $\{(\mu_{\tilde{\beta}_l(I(\alpha))}(r), \alpha) : \alpha \in A\} = \{(u, \max_{\alpha \in A}\{\alpha :\ u = \mu_{\tilde{\beta}_l(I(\alpha))}(r)\})\}$ holds true. Further, regrouping elements yields

$$\tilde{B}_l = \{(r, (\mu_{\tilde{\beta}_l(I(\alpha))}(r), \alpha)) : \alpha \in A, r \in \mathbb{R}\} = \{\{((r, \mu_{\tilde{\beta}_l(I(\alpha))}(r)), \alpha) : r \in \mathbb{R}\} : \alpha \in A\}. \tag{36}$$

Finally, by virtue of formula (29) we conclude that $\tilde{B}_l = \{(\tilde{\beta}_l(I(\alpha)), \alpha):\ \alpha \in A\}$.

The proof of Theorem 1 is complete.

4.2. Calculation of T2FSs of regression parameters

First, we construct the set $A = \{\mu_I(j) : j \in N\}$ of membership degrees values of the FS $I = \{(j, \mu_I(j)) : j \in N\}$ of sample indices. For each $\alpha \in A$, according to (32), we construct the α-cut $I(\alpha) = \{j \in N : \mu_I(j) \ge \alpha\}$ of the FS I. Further, for each $l = 0, ..., p$, we use the representation of the T2FS $\tilde{B}_l$ in the form of a collection of embedded T2FSs with constant secondary grades (see Theorem 1). This leads to the following sequence of calculations for each $\alpha \in A$.

We construct the embedded T1FS $\tilde{\beta}_l(I(\alpha)) = \{(r, \mu_{\tilde{\beta}_l(I(\alpha))}(r)) : r \in \mathbb{R}\}$. This is the FN, which is the l-th fuzzy parameter of the regression for the set $I(\alpha)$ of sample indices. To construct the FN $\tilde{\beta}_l(I(\alpha))$ one can use any known methods. An application of the method worked out in Section 2.1 of [18] yields $\tilde{\beta}_l(I(\alpha)) = \{([[\tilde{\beta}_l(I(\alpha))]_u^D, [\tilde{\beta}_l(I(\alpha))]_u^H], u) : u \in [0,1]\}$ (see (13), (30), (31) for $K = I(\alpha)$) with the MF

$$\mu_{\tilde{\beta}_l(I(\alpha))}(r_l) = \max\{u \in [0,1]:\ [\tilde{\beta}_l(I(\alpha))]_u^D \le r_l \le [\tilde{\beta}_l(I(\alpha))]_u^H\},\ r_l \in \mathbb{R}. \tag{37}$$

Formula (37) is justified by (14) with $K = I(\alpha)$. Then, the corresponding embedded T2FS with the constant secondary grade α has the form $\tilde{B}_l^{e2}(I(\alpha)) = \{(\tilde{\beta}_l(I(\alpha)), \alpha))\}$ according to (33).

Once all embedded T2FSs $\tilde{B}_l^{e2}(I(\alpha))$ with constant secondary grades $\alpha \in A$ have been obtained, the resulting T2FS of the l-th regression parameter has the form $\tilde{B}_l = \{\tilde{B}_l^{e2}(I(\alpha)):\ \alpha \in A\} = \{(\tilde{\beta}_l(I(\alpha)), \alpha)):\ \alpha \in A\}$ by Theorem 1. According to Remark 3, the T2FSs $\tilde{B}_l$ can be interpreted as follows. For fuzzy data of different quality, the l-th regression parameter $\tilde{B}_l$ is equal to the l-th parameter $\tilde{\beta}_l(I(\alpha))$ of the regression for the corresponding crisp set $I(\alpha)$ of data samples with the degree of truth being equal to α, $\alpha \in A$.

5. Illustration and discussion

5.1. Example

This example is devised to illustrate our approach. We stress that this example does not use real data. All data are given to the second decimal place.

To conduct the historical research, we need to find out the dependence of weight of a middle-aged person on he or her tall in the 10th century. The data borrowed from four authentic historical documents are located in lines 1-4 of Table 1. There is a reason to believe that the values in line 3 are slightly less reliable than the rest. We assume that in those days the height and the weight were measured with an accuracy to ±1Kg and ±5%, respectively. We denote by $\tilde{a} = (a^L, a^C, a^U)$ a 'triangular' FN with the MF

$$\mu_{\tilde a}(r)=\begin{cases}(r-a^L)/(a^C-a^L), & r\in[a^L-a^C];\\ (a^R-r)/(a^R-a^C), & r\in[a^C-a^R];\\ 0, & \text{otherwise.}\end{cases} \quad (38)$$

Table 1 contains the fuzzy data in forms of the 'triangular' FNs $\tilde{x}_{i1}=(x_{i1}^L,x_{i1}^C,x_{i1}^U)$ and $\tilde{y}_{i1}=(y_i^L,y_i^C,y_i^U)$, $i\in\{1,...,4\}$. The fuzzy data $\tilde{x}_{i1}$ and $\tilde{y}_i$ be denoted as $\tilde{x}_{i1}=appr(x_{i1}^C)$ and $\tilde{y}_i=appr(y_i^C)$, and can be interpreted as 'approximately x_{i1}^C and y_i^C', respectively. The dependence of the weight on the height for majority of data in rows 1-4 of Table 1 may seem strange, since weight gain decreases as the height increases. We assume that this can be explained by the small sample size and its non-representativeness.

Table 1

Weight depending on height

i	Height $\tilde{x}_{i1}$	Weight $\tilde{y}_i$
1	(149,150,151)	(55,1;58,60,9)
2	(159,160,161)	(63,65;67;70,35)
3	(169,170,171)	(63,65;67;70,35)
4	(189,190,191)	(80,75;85;89,25)
5	150	50
6	160	60
7	170	70
8	180	80

To solve this problem, we supplement the data sample. We know from the medical sources [31] that the Broca's formula for determining the height and the optimal weight has the form $W = H - 100$, where H is the height in cm and W is the optimal weight in Kg. According to this formula, we calculate the optimal weights for four different heights and place these data in rows 5-8 of Table 1. Thus, we have three sources of data of different quality. The high-quality source is historical, which we fully trust. We evaluate the degree of reliability of data from this source. The results which are located in rows 1,2,4 of Table 1 have the values of the degree of reliability are equal to 1. The degree of less reliable data from the second source of historical data which are located in line 3 of Table 1 is estimated at 0,9. The third source of data is medical. It is known that modern persons are taller than persons living in the 10th century and having the same weight. As a consequence, the degree of reliability of data from a medical source is not very high (for example, we estimate it as 0,7), since this source only provides information on the ratio of the weight and the height for modern people. Thus, we construct the FS 'Reliable data samples' $I=\{(1;1),(2;1),(3;0,9),(4;1),(5;0,7),(6;0,7),(7;0,7),(8;0,7)\}$ on the set $N=\{1,...,8\}$ of data sample indexes with the MF values $\mu_I(1)=\mu_I(2)=\mu_I(4)=1$, $\mu_I(3)=0,9$ и $\mu_I(5)=...=\mu_I(8)=0,7$.

In view of Remark 6, the set of membership degrees values of the FS I takes the form $A=\{0,7;0,9;1\}$. For $\alpha=1,0$; $\alpha=0,9$ and $\alpha=0,7$, according to (32), we construct the corresponding α-cuts $I(1)=\{1,2,4\}$, $I(0,9)=\{1,...,4\}$ and $I(0,7)=\{1,...,8\}$ of the FS I of data samples indices. Next, we construct FNs of parameters of linear least-squares regressions (see Section 2.1). In Example 1, these FNs are of the 'triangular' type (in general case, this is not necessary). Here, these FNs are of the 'triangular' type (in general case, this is not necessary).

For $\alpha=1$, we get embedded T1FSs $\tilde{\beta}_0(I(1))=(-86,51;-82,2;-79,6)=appr(-82,2)$ and $\tilde{\beta}_1(I(1))=(0,88;0,91;0,94)=appr(0,91)$ in the T2FSs $\tilde{B}_0^{e2}(I(1))=\{(\tilde{\beta}_0(I(1)),1))\}$ and $\tilde{B}_1^{e2}(I(1))=\{(\tilde{\beta}_1(I(1)),1))\}$, respectively, with the constant secondary grade (the degree of truth) being equal to 1.

For $\alpha=0,9$, we get embedded T1FSs $\tilde{\beta}_0(I(0,9))=(-68,47;-64,4;-60,41)=appr(-64,4)$ and $\tilde{\beta}_1(I(0,9))=(0,77;0,81;0,85)=appr(0,81)$ in the T2FSs $\tilde{B}_0^{e2}(I(0,9))=\{(\tilde{\beta}_0(I(0,9));0,9))\}$ and

$\tilde{B}_1^{e2}(I(0,9)) = \{(\tilde{\beta}_1(I(0,9)); 0,9))\}$, respectively, with the degree of truth being equal to 0,9.

For $\alpha = 0,7$, we get embedded T1FSs $\tilde{\beta}_0(I(0,7)) = (-81,79; -77; -72,3) = appr(-77)$ and $\tilde{\beta}_1(I(0,7)) = (0,85; 0,9; 0,95) = appr(0,9)$ in the T2FSs $\tilde{B}_0^{e2}(I(0,7)) = \{(\tilde{\beta}_0(I(0,7)); 0,7))\}$ and $\tilde{B}_1^{e2}(I(0,7)) = \{(\tilde{\beta}_1(I(0,7)); 0,7))\}$, respectively, with the degree of truth) being equal to 0,7.

According to Theorem 1, the resulting T2FSs of regression parameters have the forms

$$\tilde{B}_0 = \{(appr(-82,2);1), (appr(-64,4); 0,9), (appr(-77); 0,7)\}, \quad (39)$$

$$\tilde{B}_1 = \{(appr(0,91);1), (appr(0,81); 0,9), (appr(0,9); 0,7)\}. \quad (40)$$

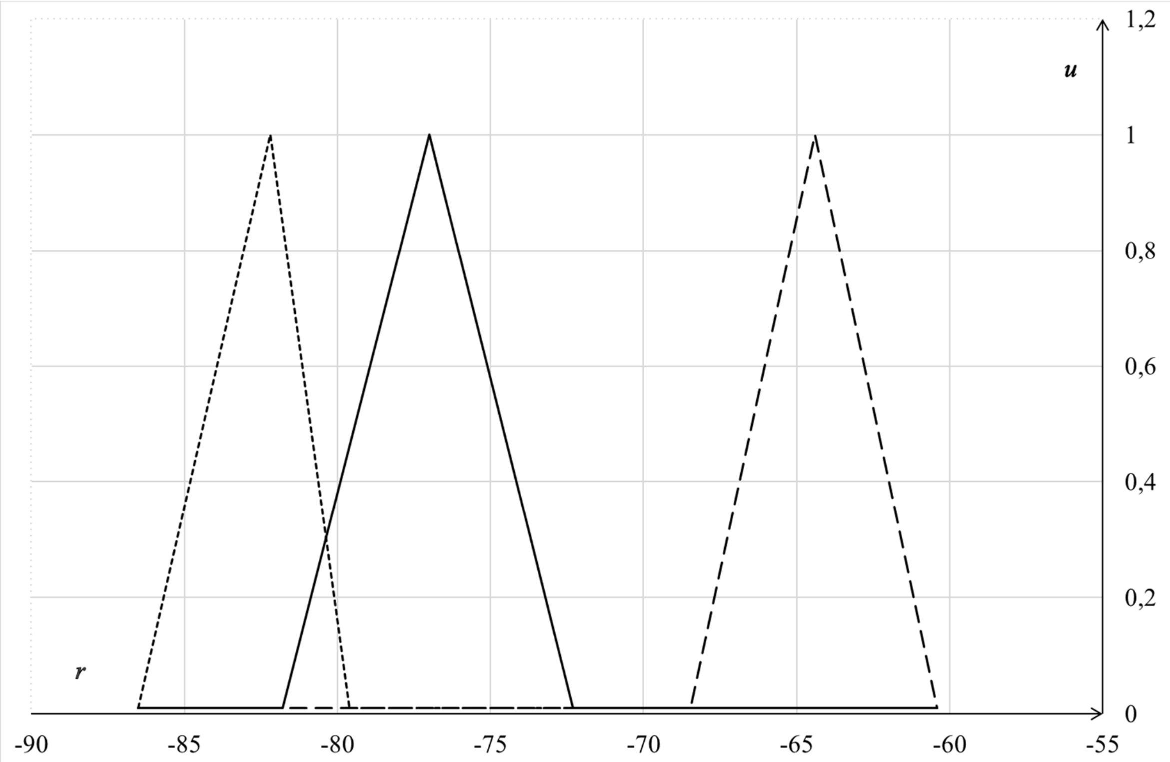

Figure 1. The lines of the levels $\alpha = 0,7$; $\alpha = 0,9$ and $\alpha = 1,0$ of the T2MF $\eta_{\tilde{B}_0}(r,u)$

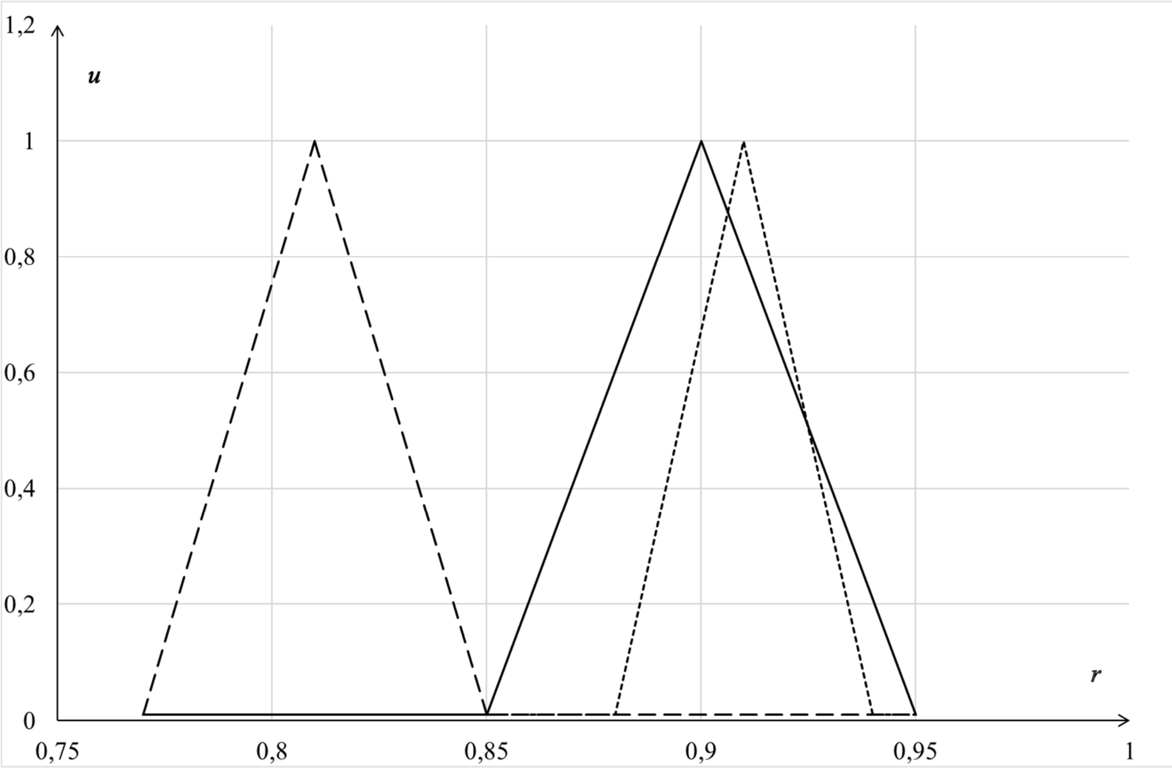

Figure 2. The lines of the levels $\alpha = 0,7$; $\alpha = 0,9$ and $\alpha = 1,0$ of the $\eta_{\tilde{B}_1}(r,u)$

The T2MFs $\eta_{\tilde{B}_0}(r,u)$ and $\eta_{\tilde{B}_1}(r,u)$ can be calculated with the help of formulae (26) and (27). Their levels $\alpha \in A = \{0,7;0,9;1\}$ are represented by solid (for $\alpha = 1$), dashed (for $\alpha = 0,9$) and dotted (for $\alpha = 0,7$) lines in Figure 1 for $\eta_{\tilde{B}_0}(r,u)$ and in Figure 2 for $\eta_{\tilde{B}_1}(r,u)$.

The obtained T2FSs can be interpreted as follows. For fuzzy data of different quality, the T2FSs $\tilde{B}_0$ and $\tilde{B}_1$ of fuzzy regression parameters values are equal to:

- the FNs $\tilde{\beta}_0(I(0,7)) = appr(-77)$ and $\tilde{\beta}_1(I(0,7)) = appr(0,9)$, respectively, for the crisp set $I(0,7) = \{1,...,8\}$ of data sample indices with the degree of truth being equal to 0,7;
- the FNs $\tilde{\beta}_0(I(0,9)) = appr(-64,4)$ and $\tilde{\beta}_1(I(0,9)) = appr(0,81)$, respectively, for the crisp set $I(0,9) = \{1,...,4\}$ of data sample indices with the degree of truth being equal to 0,9 and
- the FNs $\tilde{\beta}_0(I(1)) = appr(-82,2)$ and $\tilde{\beta}_1(I(1)) = appr(0,91)$, respectively, for the crisp set $I(1) = \{1,2,4\}$ of data sample indices with the degree of truth being equal to 1.

5.2. The discussion of the results

Let us consider a graphical interpretation obtained regression with T2FSs parameters. On Figure 3 thin solid lines demonstrate the graphs of the regression functions

$$y = (\tilde{\beta}_1(I(1))_0^L + (\tilde{\beta}_0(I(1))_0^L x = -86,51 + 0,88x, \quad (41)$$

$$y = (\tilde{\beta}_1(I(1)))_0^H + (\tilde{\beta}_0(I(1)))_0^H x = -79,6 + 0,94x \quad (42)$$

with parameters which correspond to the lower and upper bounds of the 0-cuts $[(\tilde{\beta}_0(I(1))_0^L, (\tilde{\beta}_0(I(1))_0^H] = [-86,51;-79,6]$ and $[(\tilde{\beta}_1(I(1))_0^L, (\tilde{\beta}_1(I(1))_1^H] = [0,88;0,94]$ of the FNs $\tilde{\beta}_0(I(1)) = appr(-82,2)$ and $\tilde{\beta}_1(I(1)) = appr(0,91)$, respectively. These FNs are corresponding

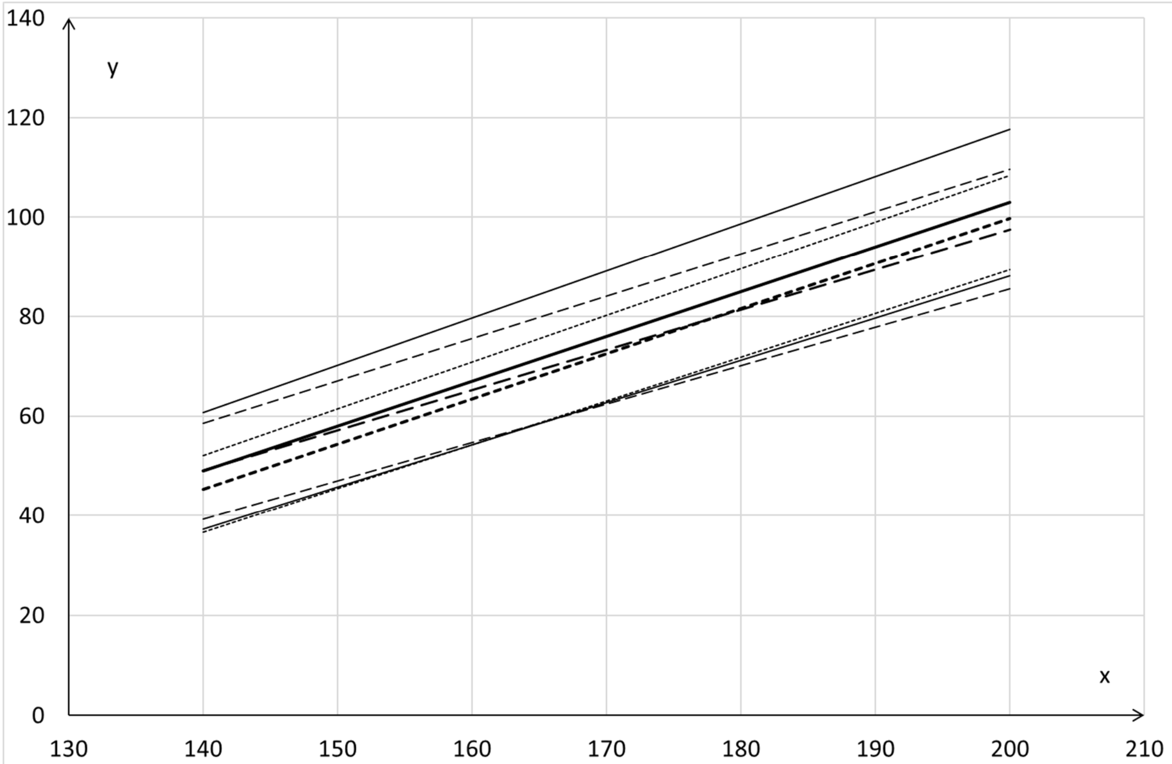

Figure 3. The illustration of the regression with the parameter T2FSs

parameters of the fuzzy linear regression for crisp set $I(1)=\{1,2,4\}$ of data sample indices. For the same set of sample indices, the thick solid line shows the graph of the regression function $y=(\tilde{\beta}_1(I(1))_1^L+(\tilde{\beta}_0(I(1))_1^L x=-82,2+0,91x$ with parameters which correspond to the 1-cuts $[-82,2;-82,2]$ and $[0,91;0,91]$ of the FNs $\tilde{\beta}_0(I(1))=appr(-82,2)$ and $\tilde{\beta}_1(I(1))=appr(0,91)$, respectively. For the crisp set $I(0,9)=\{1,...,4\}$ of data sample indices, the dashed lines demonstrate the graphs of the regression functions $y=(\tilde{\beta}_1(I(0,9))_0^L+(\tilde{\beta}_0(I(0,9))_0^L x=-68,47+0,77x$ and

$$y=(\tilde{\beta}_1(I(0,9))_0^H+(\tilde{\beta}_0(I(0,9))_0^H x=-60,4+0,85x \tag{43}$$

(thin lines), and

$$y=(\tilde{\beta}_1(I(0,9))_1^L+(\tilde{\beta}_0(I(0,9))_1^L x=-64,4+0,81x \tag{44}$$

(a thick line). For the crisp set $I(0,7)=\{1,...,8\}$ of data sample indices, the dotted lines demonstrate the graphs of the regression functions $y=(\tilde{\beta}_1(I(0,7))_0^L+(\tilde{\beta}_0(I(0,7))_0^L x=-81,79+0,85x$ and

$$y=(\tilde{\beta}_1(I(0,7))_0^H+(\tilde{\beta}_0(I(0,7))_0^H x=-72,3+0,95x \tag{45}$$

(thin lines), and

$$y=(\tilde{\beta}_1(I(1)))_0^H+(\tilde{\beta}_0(I(1)))_0^H x=-79,6+0,94x \tag{46}$$

(a thick line).

Figure 3 allows us to draw the following conclusions. Regressions corresponding to data of different quality may differ from each other and express a different relationship between the independent variables and predicted output. Therefore, the resulting regression with T2FSs of parameters should be taken as a whole as a collection of regressions corresponding to α-cuts $I(\alpha)$, $\alpha\in[0,1]$ of the FS I of data sample indices. Only in this case we get an idea about dependence between the regression and the quality of the data used. Sometimes understanding this is important.

If we need some 'maximizing' regression, which has parameters values with the primary membership degrees equal to 1, then it should be considered as a collection of regressions corresponding to α-cuts $I(\alpha),\ \alpha \in [0,1]$ of the FS I of data sample indices. For these regressions, we take the parameters values with the membership degrees equal to 1. In Example 1, these are regressions with graphs depicted by thick lines in Figure 3.

The regression for fuzzy data of different quality can be represented as a set of fuzzy regressions with the degree of truth $\alpha \in A$. These fuzzy regressions correspond to data samples of different quality with indices belonging to α-cuts $I(\alpha),\ \alpha \in A$ of the FS I. Thus, the degree of truth of the regression is determined by the quality of the data for which it is constructed. The higher the quality of data samples, the greater the degree of membership of their indices to the FS I and the higher the degree of truth of the corresponding fuzzy regression in the collection that forms the regression for fuzzy data of different quality. On the other hand, increasing the amount of low-quality data reduces the degree of truth of the corresponding fuzzy regression in the collection.

6. Conclusion

According to the proposed approach, the linear regression parameters for fuzzy data of different quality are T2FSs with constant secondary grades. Although, in general, a T2FS is a rather complicated mathematical object, T2FSs with constant secondary grades are simple enough for practical use. Therefore, to represent these sets in a form which is convenient for understanding and applications, we have used a decomposition method. The results obtained have allowed us to decompose the T2FSs of linear regression parameters according to secondary grades into finite sets of FNs. These are fuzzy parameters of the regressions which correspond to different α-cuts of FSs of data sample indices. One direction of future investigation of regressions with a fuzzy set of data sample indices may be related to the development of a similar approach for possibilistic regression analysis.

7. Acknowledgements

This work has been supported by Ministry of Education and Science of Ukraine: R&D Project 0122U001844 for the period of 2022-2024 at Taras Shevchenko National University of Kyiv.

8. References

[1] N. Chukhrova, A. Johannssen, Fuzzy regression analysis: Systematic review and bibliography, Appl. Soft Comput. J. 84 (2019) 1–29. doi:10.1016/j.asoc.2019.105708.

[2] H. Tanaka, S. Uejima, K. Asai, Linear regression analysis with Fuzzy model, IEEE Trans. Syst. Man Cybern. 12(6) (1982) 903–907.

[3] P. D'Urso, Exploratory multivariate analysis for empirical information affected by uncertainty and modeled in a fuzzy manner: a review, Granular Comput. 2(4) (2017) 225–247. doi:10.1007/s41066-017-0040-y.

[4] A. Celmins, Least squares model fitting to fuzzy vector data, Fuzzy Sets and Syst. 22(3) (1987) 245–269. doi:10.1016/0165-0114(87)90070- 4.

[5] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY, 2009. doi:10.1007/978-0-387-84858-7.

[6] A. Forrester, A. Sobester, A. Keane, Multi-fidelity optimization via surrogate modelling, Proceedings of the Royal society A: Mathematical, Physical and Engineering Sci. 463 (2007) 3251-3269.

[7] A. Zaytsev, E. Burnaev, Minimax Approach to Variable Fidelity Data Interpolation, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, volume 54, Fort Lauderdale, FL, USA, 2017, pp.652–661. https://proceedings.mlr.press/v54/zaytsev 17a.html.

[8] J. Hevesi, A. Flint, J. Istok, Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: isohyetal maps. J. Appl. Meteorol. 31 (1992) 677–688. doi:10.1175/1520-0450(1992)031!0677:PEIMTUO2.0.CO;2.

[9] M. Kennedy, A. O'Hagan, Predicting the output from a complex computer code when fast approximations are available, Biometrika 87(1) (2000) 1–13. doi:10.1093/biomet/87.1.1.

[10] S. Pan, Q. Yang, A survey on transfer learning. IEEE Trans. on Knowledge and Data Engineering 22(10) (2010) 1345–1359. doi:10.1109/TKDE.2009.191.

[11] J.W. Bandler, Q.S. Cheng, S.A. Dakroury, A.S. Mohamed, M.H. Bakr, K. Madsen, J. Sondergaard, Space mapping: the state of the art, IEEE Trans. on microwave theory and techniques 52(1) (2004) 337–361. https://ieeexplore.ieee.org/document/ 1262727.

[12] C. E. Rasmussen, C. K. I. Williams, Gaussian processes for machine learning, The MIT Press, 2006.

[13] N. Arruda, J. Alcantara, V. Vidal, A. Brayner, M. Casanova, V. Pequeno, W.A. Franco, Fuzzy Approach for Data Quality Assessment of Linked Datasets, in: Proceedings of the 21th International Conference on Enterprise Information Syst. (ICEIS 2019), volume 1, 2019, pp. 399–406. doi:10.5220/0007718803990406.

[14] J. Gamez, F. Modave, O. Kosheleva, Selecting the Most Representative Sample is NP-Hard: Need for Expert (Fuzzy) Knowledge, in: Proceedings of the IEEE International Conference on Fuzzy Syst. (IEEE World Congress on Computational Intelligence), 2008, pp. 1069–1074. doi:10.1109/FUZZY.2008. 4630502.

[15] M. Z. Khan, M. F. Khan, M. Aslam, A. R. Mughal, Design of Fuzzy Sampling Plan Using the Birnbaum-Saunders Distribution, Mathematics 7(1) (2019) 9. doi:10.3390/math7010009.

[16] S. Mashchenko, Intersections and unions of fuzzy sets of operands, Fuzzy Sets and Syst. 352(1) (2018) 12–25. doi:10.1016/j.fss.2018.04.006.

[17] S.O. Mashchenko, D.O. Kapustian, Decomposition of intersections with fuzzy sets of operands, in: V.A. Sadovnichiy, M.Z. Zgurovsky (Ed.), Contemporary Approaches and Methods in Fundamental Mathematics and Mechanics. Understanding Complex Systems, Springer, Cham, 2020, pp. 417–432. https://link.springer.com/book/10.1007/ 978-3-030-50302-4.

[18] H.-C. Wu, Linear regression analysis for fuzzy input and output data using the extension principle, Computers and Mathematics with Applications 45(12) (2003) 1849–1859. doi:10.1016/S0898-1221(03)90006-X.

[19] S. Heilpern, Representation and application of fuzzy numbers, Fuzzy Sets and Syst. 91 (1997) 259–268.

[20] L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning – I, Inform. Sci. 8 (1975) 199–249. doi:10.1016/0020-0255(75)90036-5.

[21] D. Dubois, H. Prade, Operations on fuzzy numbers, International J. of Syst. Sci. 9(6) (1978) 613–626.

[22] L. A. Zadeh, Quantitative fuzzy semantics, Inform. Sci. 3 (1971) 159–176. doi:10.1016/S0020-0255(71)80004-X.

[23] N. N. Karnik, J.M. Mendel, Introduction to type-2 fuzzy logic systems, IEEE International Conference on Fuzzy Syst. 2 (1998) 915–920.

[24] J. M. Mendel, R. I. John, Type-2 fuzzy sets made simple, IEEE Trans. on Fuzzy Syst. 10 (2002) 117–127.

[25] J. Harding, C. Walker, E. Walker, The variety generated by the truth value algebra of T2FSs, Fuzzy Sets and Syst. 161 (2010) 735–749.

[26] J. Aisbett, J.T. Rickard, D.G. Morgenthaler, Type-2 fuzzy sets as functions on spaces, IEEE Trans. on Fuzzy Syst. 18(4) (2010) 841–844.

[27] J. M. Mendel, Type-2 fuzzy sets: some questions and answers, IEEE Connections. Newsletter of the IEEE Neural Networks Society 1 (2003) 10–13.

[28] S.O. Mashchenko, Sums of fuzzy set of summands, Fuzzy Sets and Syst. 417 (2020) 140–151. doi:10.1016/j.fss.2020.10.006.

[29] S.O. Mashchenko, Sum of discrete fuzzy numbers with fuzzy set of summands, Cybernetics and Syst. Analysis 57(3) (2021) 374–382. doi:10.1007/s10559-021-00362-w.

[30] S.O. Mashchenko, Minimum of fuzzy numbers with a fuzzy set of operands, Cybernetics and Syst. Analysis 58(2) (2022) 210–219. doi: 10.1007/s10559-021-00362-w.

[31] J. C. Segen, Concise dictionary of modern medicine, McGraw-Hill, New York, NY,2006.