Estimation of a Simple Multivariate Linear Model for Fuzzy Random Sets

Dabuxilatu Wang and Miao Shi

School of Economics and Statistics
Research Center of Statistical Science Lingnan,
Guangzhou University, No.230 Waihuanxi road,
Higher Education Mega Center, Guangzhou, 510006, P.R. China

Abstract. A simple two-variate linear regression model with fuzzy random sets under concepts of functional data analysis is considered. The support function of a fuzzy random set establishes a useful embedding of the space of fuzzy random sets into a cone of a functional Hilbert space. Treating the fuzzy random sets as special functional data, we estimate the linear model within the cone. An example of the case of LR fuzzy random sets is given.

Keywords: Fuzzy random sets, linear models, Bootstrap distribution.

1 Introduction

In investigating the relationship between random elements, regression analysis enables to seek for some complex effect of several random elements upon another. Regression techniques have long been relevant to many fields [1]. The random elements considered actually in many practical application in public health, medical science, ecology, social or economic and financial problems sometimes involve vagueness, so the regression problems have to face with such a mixture of fuzziness and randomness. There are two main lines concerning regression modeling with fuzzy data in literature: namely, the so-called fuzzy or possibilistic regression proposed by Tanaka [11] and widely analyzed since then [4,5,11] and the so-called least squares problems of linear models [1,7,9,8,14] with fuzzy random sets [2,10,12,13]). In the former research line, the regression models are established based on possibilistic inclusion relationship between input and output of the systems rather than stochastic statistical settings. The last research line is based on statistical nonparametric settings, to consider both effects of randomness and fuzziness to the systems in the regression modeling, and the parameters (vector valued or fuzzy sets valued) estimation of the linear models are solved with least squares methods under metric between sets (see [1,2,7,9,8,14] and literature therein), and some concrete computational formulas for parameter estimation for simple linear regression model have been given. However, the same problems remain to be further investigated for the case of multivariate linear regression with fuzzy random sets.

P. Grzegorzewski et al. (eds.), *Strengthening Links between Data Analysis & Soft Computing*,
Advances in Intelligent Systems and Computing 315, DOI: 10.1007/978-3-319-10765-3_24

In this paper, we focus on a simple two-variate linear regression model with fuzzy random sets under concepts of functional data analysis. Based on the support function of the fuzzy random sets, we treat the fuzzy random sets as special functional data, and estimate the linear model within the support functional space. An example of the case of LR fuzzy random sets is given.

2 Preliminaries

2.1 Fuzzy Set on $\mathbb{R}^n$

A fuzzy set $\tilde{u}$ of $\mathbb{R}^n$ equivalents to its membership function $\tilde{u} : \mathbb{R}^n \to [0,1]$, where the number $\tilde{u}(x)$ represents the degree of membership that x belongs to $\tilde{u}$. By $F(\mathbb{R}^n)$ we denote the collection of all normal, convex and compact fuzzy sets on $\mathbb{R}^n$, i.e. for $\tilde{u} \in F(\mathbb{R}^n)$, (1) There exists $x_0 \in \mathbb{R}^n$ such that $\tilde{u}(x_0) = 1$; (2) The $\alpha-$cut of $\tilde{u}$, $\tilde{u}_\alpha := \{x \in \mathbb{R}^n : \tilde{u}(x) \geq \alpha\}$, $\alpha \in (0,1]$, is a convex and compact set of $\mathbb{R}^n$; (3) $\tilde{u}_0 := cl\{x \in \mathbb{R}^n : \tilde{u}(x) > 0\}$, the support of $\tilde{u}$, is compact.

Zadeh's extension principle [4] allows us to define addition and scalar multiplication on $F(\mathbb{R}^n)$:

$$(\tilde{u} \oplus \tilde{v})(x) = \sup_{s+t=x} \min(\tilde{u}(s), \tilde{v}(t)), x \in \mathbb{R}^n.$$

$$(a \odot \tilde{u})(x) = \begin{cases} \tilde{u}(\frac{x}{a}), a \neq 0 \\ 0, a = 0 \end{cases} \quad a \in \mathbb{R}.$$

and [9] for any $a, b \in \mathbb{R}$, it holds

$$(ab) \odot \tilde{u} = a \odot (b \odot \tilde{u}), a \odot (\tilde{u} \oplus \tilde{v}) = (a \odot \tilde{u}) \oplus (a \odot \tilde{v}).$$

But it holds only for $ab \geq 0, a, b \in \mathbb{R}$

$$(a + b) \odot \tilde{u} = (a \odot \tilde{u}) \oplus (b \odot \tilde{u}).$$

It indicates that $(F(\mathbb{R}^n), \oplus, \odot)$ is not a linear space. With Minkowski's sets operation it holds

$$(\tilde{u} \oplus \tilde{v})_\alpha = \tilde{u}_\alpha \oplus \tilde{v}_\alpha, \quad \alpha \in (0,1].$$

$$(a \odot \tilde{u})_\alpha = a \odot \tilde{u}_\alpha, \quad \alpha \in (0,1].$$

Definition 2.1 [14,2]. For $\tilde{u}, \tilde{v} \in F(\mathbb{R}^n)$, if there exists $\tilde{h} \in F(\mathbb{R}^n)$ such that $\tilde{u} = \tilde{v} \oplus \tilde{h}$, then $\tilde{h}$ is said to be Hukuhara difference between $\tilde{u}, \tilde{v}$ and denoted by $\tilde{h} := \tilde{u} \ominus_H \tilde{v}$.

The support function of $\tilde{u} \in F(\mathbb{R}^n)$ is defined as

$$S_{\tilde{u}_\alpha}(x) = \begin{cases} \sup_{t \in \tilde{u}_\alpha} \{x \cdot t\}, \alpha \in (0,1], \\ 0, \alpha = 0. \end{cases} \quad x \in S^{n-1} = \{x : \| x \| = 1\}.$$

where $\cdot$ denotes the inner product in the Euclidean space $\mathbb{R}^n$. It holds that for $\tilde{u}, \tilde{v} \in F(\mathbb{R}^n)$ and $a \in \mathbb{R}$,

$$S_{\tilde{u} \oplus \tilde{v}} = S_{\tilde{u}} + S_{\tilde{v}}.$$

$$S_{a\odot\tilde{u}}(x) = aS_{\tilde{u}}(x), a > 0; S_{a\odot\tilde{u}}(x) = -aS_{\tilde{u}}(-x), a < 0.$$

thus, it holds that

$$S_{((a\odot\tilde{u})\oplus(b\odot\tilde{v}))_\alpha}(x) = \begin{cases} (aS_{\tilde{u}_\alpha} + bS_{\tilde{v}_\alpha})(x), a, b > 0 \\ -(aS_{\tilde{u}_\alpha} + bS_{\tilde{v}_\alpha})(-x), a, b < 0. \end{cases}$$

where $\alpha \in [0, 1]$. Thus, the map $S : F(\mathbb{R}^n) \to L^2(S^{n-1} \times [0, 1])$, $\tilde{u} \mapsto S_{\tilde{u}_\alpha}(x)$ enables us to view the fuzzy set $\tilde{u}$ as a support function equivalently, i.e. the map S embeds $F(\mathbb{R}^n)$ into a cone of functional Hilbert space [7].

We will employ the distance between $\tilde{u}, \tilde{v}$ proposed by [4] by the L_2 metric δ_2,

$$\delta_2(\tilde{u}, \tilde{v}) := \Big(n \int_0^1 \int_{S^{n-1}} (S_{\tilde{u}_\alpha}(x) - S_{\tilde{v}_\alpha}(x))^2 \mu(dx) d\alpha\Big)^{1/2},$$

where μ is a normalized Lebesgue measure. This distance has been widely used in area of fuzzy set-valued analysis, and in recent years several alternative versions of which as new metrics between fuzzy values have been proposed in literature, see [2,13].

2.2 Fuzzy Random Sets (Fuzzy Random Variables)

Fuzzy random sets as an extension of the concept of random sets had been introduced by Puri and Ralescu [10], and other definitions of fuzzy random sets were also proposed by Kwakernaak, Kruse and Meyer and Krätschmer [9] in different setting.

Definition 3.1 [10]. Let $(\Omega, \mathcal{B}, P)$ be a complete probability space. The mapping $\tilde{X} : \Omega \to F(\mathbb{R}^n)$ is said to be a fuzzy random set (frs) if $\tilde{X}$ is $\mathcal{B} - \mathcal{A}$ measurable, where we assume $\mathcal{A}$ is a σ-algebra induced by $\tilde{X}$ associated with δ_2.

Let $\tilde{X}$ be a frs, then for $\alpha \in [0, 1]$, $S_{\tilde{X}_\alpha}$ is a special random element, for a fixed $x \in S^{n-1}$, $S_{\tilde{X}_\alpha}(x)$ is random variable: $\Omega \to \mathbb{R}$, $\omega \mapsto S_{\tilde{X}_\alpha(\omega)}(x)$. A sample $\tilde{x}$ from $\tilde{X}$ can be viewed as a fuzzy data, thus, $S_{\tilde{x}}$ is a special functional data, an equivalence of $\tilde{x}$ [7].

Definition 3.2 [10]. Let $\tilde{X}$ be a frs. The Aumann expectation of $\tilde{X}$ is defined as a fuzzy set $E\tilde{X} \in F(\mathbb{R}^n)$ satisfying

$$\forall \alpha \in [0, 1] : (E\tilde{X})_\alpha = E(\tilde{X}_\alpha),$$

Here $E(\tilde{X}_\alpha)$ is the Aumann expectation of the random set $\tilde{X}_\alpha$ defined by

$$E(\tilde{X}_\alpha) = \{E\eta : \eta(\omega) \in \tilde{X}_\alpha(\omega)\ P - a.e.\ and\ \eta \in L^1(\Omega, \mathcal{B}, P)\}.$$

Note that $E(S_{\tilde{X}_\alpha}) = S_{E(\tilde{X}_\alpha)}$ [13,14] if the expectation $E(\tilde{X}_\alpha)$ exists, where $E(\tilde{X}_\alpha)$ is an Aumann expectation of $(\tilde{X}_\alpha), \alpha \in [0, 1]$ [10,9].

In the sequel, we assume that frs $\tilde{X}$ is with second order, i.e.

$$E(\|\tilde{X}\|) := E(\delta_2^2(\tilde{X}, \{0\})) < +\infty,$$

The Fréchet variance of $\tilde{X}$ w.r.t distance δ_2 is given in [12] as

$$Var(\tilde{X}) := E(\delta_2^2(\tilde{X}, E(\tilde{X}))) = n\int_0^1 \int_{S^{n-1}} Var(S_{\tilde{X}_\alpha}(x))\mu(dx)d\alpha.$$

and the Fréchet covariance of frs's $\tilde{X}, \tilde{Y}$ is also given in [12] as

$$Cov(\tilde{X}, \tilde{Y}) := n\int_0^1 \int_{S^{n-1}} Cov(S_{\tilde{X}_\alpha}(x), S_{\tilde{Y}_\alpha}(x))\mu(dx)d\alpha.$$

Note that,

$$Cov((a \odot \tilde{X}) \oplus (b \odot \tilde{Y}), c \odot \tilde{Z}) = acCov(\tilde{X}, \tilde{Z}) + bcCov(\tilde{Y}, \tilde{Z})$$

holds only for $ac \geq 0, bc \geq 0, a, b, c \in \mathbb{R}$.

The independence of frs's can be followed by the independence of the random elements which is already defined by [13]. If two frs's $\tilde{X}$ and $\tilde{Y}$ are independent, then $Cov(\tilde{X}, \tilde{Y}) = 0$. However, if $Cov(\tilde{X}, \tilde{Y}) \neq 0$, then $\tilde{X}$ and $\tilde{Y}$ will be dependent in some sense of semi-linear or non-linear [3].

Remark 2.1. The Fréchet variance, covariance can be defined w.r.t. different distances for frs (see [2,4,13]), and in general these distances such as $d_\infty, \delta_2, D_\theta^\varphi$ [2,4,13] are not coincide each other except some special cases. We prefer to employ Näther's one since that the concerned distance δ_2 is standard and simple one used in functional analysis.

Fréchet Principle [12]. The $E(\tilde{X})$ is the solution of the optimization problem $\inf_{\tilde{Y}\in F(\mathbb{R}^n)} E(\delta_2^2(\tilde{X}, \tilde{Y}))$.

Let $\tilde{X}, \tilde{Y}$ be frs's, and let $\{\tilde{X}_i\}, \{\tilde{Y}_i\}, i = 1, \cdots, m$, be independent observations on $\tilde{X}, \tilde{Y}$, respectively. Then equivalently we have r.v. $S_{\tilde{X}_\alpha}(x), S_{\tilde{Y}_\alpha}(x)$ and the functional data sets $\{S_{\tilde{X}_{i\alpha}}(x)\}$, $\{S_{\tilde{Y}_{i\alpha}}(x)\}$, $i = 1, \cdots, m$, and the estimations of $E(S_{\tilde{X}_\alpha}(x))$, $Var(S_{\tilde{X}_\alpha}(x))$ and $Cov(S_{\tilde{X}_\alpha}(x), S_{\tilde{Y}_\alpha}(x))$ are respectively as follows,

$$E(\widehat{S_{\tilde{X}_\alpha}(x)}) = \frac{1}{m}\sum_{i=1}^m S_{\tilde{X}_{i\alpha}}(x), Var(\widehat{S_{\tilde{X}_\alpha}(x)}) = \frac{1}{m}\sum_{i=1}^m (S_{\tilde{X}_{i\alpha}}(x) - S_{\overline{\tilde{X}}_\alpha})^2,$$

$$Cov(S_{\tilde{X}_\alpha}\widehat{(x), S_{\tilde{Y}_\alpha}}(x)) = \frac{1}{m}\sum_{i=1}^m (S_{\tilde{X}_{i\alpha}}(x) - S_{\overline{\tilde{X}}_\alpha})(S_{\tilde{Y}_{i\alpha}}(x) - S_{\overline{\tilde{Y}}_\alpha}).$$

So that $\widehat{E\tilde{X}} = n\int_0^1 \int_{S^{n-1}} E(\widehat{S_{\tilde{X}_\alpha}(x)})\mu(dx)d\alpha$, $\widehat{Var\tilde{X}} = n\int_0^1 \int_{S^{n-1}} Var(\widehat{S_{\tilde{X}_\alpha}}(x)) \cdot \mu(dx)d\alpha$, $Cov\widehat{(\tilde{X}, \tilde{Y})} = n\int_0^1 \int_{S^{n-1}} Cov(S_{\tilde{X}_\alpha}\widehat{(x), S_{\tilde{Y}_\alpha}}(x))\mu(dx)d\alpha$.

3 A Simple Multivariate Linear Regression Model with frs

Now we consider a new two-variate linear model with frs's, i.e. the case where the response frs $\tilde{Y}$ can be approximately linearly expressed by two explanatory frs's $\tilde{x}_1, \tilde{x}_2$ (compare with the considered models in [1,9,8,14]),

$$\tilde{Y} = \tilde{a} \oplus \beta_1\tilde{x}_1 \oplus \beta_2\tilde{x}_2 \oplus \tilde{\varepsilon}, \tag{1}$$

where $\tilde{a}$ is a fuzzy number to be estimated, β_1, β_2 are real number- valued parameters to be estimated, $\tilde{\varepsilon}$ is a uncertain disturbance frs with unknown probability distribution, whose Aumann expectation is assumed to be $E(\tilde{\varepsilon}) = \tilde{0}$, which means that given the realization $\tilde{x}_1^0, \tilde{x}_2^0$ of $\tilde{x}_1, \tilde{x}_2$

$$E(\tilde{Y}|\tilde{x}_1^0, \tilde{x}_2^0) = \tilde{a} \oplus \beta_1\tilde{x}_1^0 \oplus \beta_2\tilde{x}_2^0 \oplus \tilde{0}. \tag{2}$$

We assume that for the model there exists Hukuhara difference $\tilde{Y} \ominus_H (\tilde{a} \oplus \beta_1\tilde{x}_1 \oplus \beta_2\tilde{x}_2)$ and frs $\tilde{\varepsilon}$ can be formally expressed as

$$\tilde{\varepsilon} = \tilde{Y} \ominus_H (\tilde{a} \oplus \beta_1\tilde{x}_1 \oplus \beta_2\tilde{x}_2). \tag{3}$$

such that $\tilde{Y} = (\tilde{a} \oplus \beta_1\tilde{x}_1 \oplus \beta_2\tilde{x}_2) \oplus (\tilde{Y} \ominus_H (\tilde{a} \oplus \beta_1\tilde{x}_1 \oplus \beta_2\tilde{x}_2)))$.

Assume that we have independent observation $\{\tilde{Y}_i\}$,$\{\tilde{x}_{1i}\}$,$\{\tilde{x}_{2i}\}$ on $\tilde{Y}, \tilde{x}_1, \tilde{x}_2$, respectively, equivalently we have three functional data sets $\{S_{\tilde{Y}_{i\alpha}}(x)\}$, $\{S_{\tilde{x}_{1i\alpha}}(x)\}$, $\{S_{\tilde{x}_{2i\alpha}}(x)\}$, $i = 1, \cdots, m$.

Theorem 3.1. The least squares problem

$$\min_{\tilde{a}\in F(\mathbb{R}^n), \beta_1,\beta_2\geqslant 0 \quad or \quad \beta_1,\beta_2\leqslant 0} \frac{1}{m}\sum_{i=1}^{m} \delta_2^2(\tilde{Y}_i, \tilde{a} \oplus \beta_1\tilde{x}_{1i} \oplus \beta_2\tilde{x}_{2i})$$

has solutions (1) when $\beta_1, \beta_2 \geqslant 0$,

$$\hat{\beta}_1 = \max\Big\{0, \frac{\widehat{Cov(\tilde{Y},\tilde{x}_1)}\widehat{Var\tilde{x}_2} - \widehat{Cov(\tilde{x}_1,\tilde{x}_2)}\widehat{Cov(\tilde{Y},\tilde{x}_2)}}{\widehat{Var\tilde{x}_1}\widehat{Var\tilde{x}_2} - [\widehat{Cov(\tilde{x}_1,\tilde{x}_2)}]^2}\Big\},$$

$$\hat{\beta}_2 = \max\Big\{0, \frac{\widehat{Cov(\tilde{Y},\tilde{x}_2)}\widehat{Var\tilde{x}_2} - \widehat{Cov(\tilde{x}_1,\tilde{x}_2)}\widehat{Cov(\tilde{Y},\tilde{x}_1)}}{\widehat{Var\tilde{x}_1}\widehat{Var\tilde{x}_2} - [\widehat{Cov(\tilde{x}_1,\tilde{x}_2)}]^2}\Big\},$$

$$\hat{\tilde{a}} = \overline{\tilde{Y}} \ominus_H (\hat{\beta}_1\overline{\tilde{x}}_1 \oplus \hat{\beta}_2\overline{\tilde{x}}_2).$$

(2)when $\beta_1, \beta_2 \leqslant 0$,

$$\hat{\beta}_1 = \min\Big\{0, -\frac{\widehat{Cov(\tilde{Y},-\tilde{x}_1)}\widehat{Var\tilde{x}_2} - \widehat{Cov(\tilde{x}_1,\tilde{x}_2)}\widehat{Cov(\tilde{Y},-\tilde{x}_2)}}{\widehat{Var\tilde{x}_1}\widehat{Var\tilde{x}_2} - [\widehat{Cov(\tilde{x}_1,\tilde{x}_2)}]^2}\Big\},$$

$$\hat{\beta}_2 = \min\Big\{0, -\frac{\widehat{Cov(\tilde{Y},-\tilde{x}_2)}\widehat{Var\tilde{x}_2} - \widehat{Cov(\tilde{x}_1,\tilde{x}_2)}\widehat{Cov(\tilde{Y},-\tilde{x}_1)}}{\widehat{Var\tilde{x}_1}\widehat{Var\tilde{x}_2} - [\widehat{Cov(\tilde{x}_1,\tilde{x}_2)}]^2}\Big\},$$

$$\hat{\tilde{a}} = \overline{\tilde{Y}} \ominus_H (\hat{\beta}_1(\overline{-\tilde{x}}_1) \oplus \hat{\beta}_2(\overline{-\tilde{x}}_2)).$$

Proof. (1) Based on Fréchet principle, we have $\frac{1}{m}\sum_{i=1}^{m} \delta_2^2(\tilde{Y}_i, \tilde{a}\oplus\beta_1\tilde{x}_{1i}\oplus\beta_2\tilde{x}_{2i}) =$
$n\int_0^1 \int_{S^{n-1}} \frac{1}{m}\sum_{i=1}^{m}(S_{\tilde{Y}_{i\alpha}}(t) - S_{(\tilde{a}\oplus\beta_1\tilde{x}_{1i}\oplus\beta_2\tilde{x}_{2i})_\alpha}(t))^2\mu(dt)d\alpha$
$=n\int_0^1 \int_{S^{n-1}} \frac{1}{m}\sum_{i=1}^{m}(S_{\tilde{Y}_{i\alpha}}(t) - S_{\tilde{a}_\alpha}(t) - \beta_1 S_{\tilde{x}_{1i\alpha}}(t) - \beta_2 S_{\tilde{x}_{2i\alpha}}(t))^2\mu(dt)d\alpha$
$\geqslant n\int_0^1 \int_{S^{n-1}} \frac{1}{m}\sum_{i=1}^{m}(S_{\tilde{Y}_{i\alpha}}(t) - \beta_1 S_{\tilde{x}_{1i\alpha}}(t) - \beta_2 S_{\tilde{x}_{2i\alpha}}(t) - (S_{\overline{\tilde{Y}}_\alpha}(t) - \beta_1 S_{\overline{\tilde{X}}_{1\alpha}}(t) -$

$\beta_2 S_{\overline{\tilde{X}}_{2\alpha}}(t)))^2\mu(dt)d\alpha$, which means $\tilde{a} = \overline{\tilde{Y}} \ominus_H (\beta_1\overline{\tilde{x}}_1 \oplus \beta_2\overline{\tilde{x}}_2)$ minimizes $\frac{1}{m}\sum_{i=1}^{m}\delta_2^2\cdot(\tilde{Y}_i, \tilde{a}\oplus\beta_1\tilde{x}_{1i}\oplus\beta_2\tilde{x}_{2i})$. Furthermore, set $f(\beta_1,\beta_2) := \frac{1}{m}\sum_{i=1}^{m}\delta_2^2(\tilde{Y}_i, (\overline{\tilde{Y}}\ominus_H(\beta_1\overline{\tilde{x}}_1 \oplus \beta_2\overline{\tilde{x}}_2)) \oplus (\beta_1\tilde{x}_{1i}\oplus\beta_2\tilde{x}_{2i})) = \widehat{Var\tilde{Y}} + \beta_1^2\widehat{Var\tilde{x}_1} + \beta_2^2\widehat{Var\tilde{x}_2} - 2\beta_1\widehat{Cov(\tilde{x}_1,\tilde{Y})} - 2\beta_2\widehat{Cov(\tilde{x}_2,\tilde{Y})} + 2\beta_1\beta_2\widehat{Cov(\tilde{x}_1,\tilde{x}_2)}$, solving the equations $\frac{\partial f}{\partial\beta_1} = 0$, $\frac{\partial f}{\partial\beta_2} = 0$, then we have the solutions $\hat{\beta}_1, \hat{\beta}_2$ of (1).

The proof of (2) is analogous to the proof of (1), but we should take $\beta_1\tilde{x}_1 = (-\beta_1)(-\tilde{x}_1), \beta_2\tilde{x}_2 = (-\beta_2)(-\tilde{x}_1)$. □

4 Simulation Example

Assume that the observed human's pulse, diastolic pressure and systolic pressure can be comprehensively expressed by $\tilde{Y} = (\mu_y, l_y)_L$, $\tilde{x}_1 = (\mu_1, l_1)_L$, $\tilde{x}_2 = (\mu_2, l_2)_L$, the symmetric triangular fuzzy numbers (see.[13]), respectively, as shown in Table 1.

Table 1. Data of human's pulse, diastolic pressure and systolic pressure

i	$(\mu_y, l_y)_L$	$(\mu_1, l_1)_L$	$(\mu_2, l_2)_L$
1	(74,16)	(145.5, 27.5)	(85.5, 19.5)
2	(57.5, 10.5)	(132.5, 28.5)	(94.5, 23.5)
3	(73, 41)	(158.5, 27.5)	(85.5, 27.5)
4	(85.5, 24.5)	(131, 26)	(90, 28)
5	(75.5, 13.5)	(149.5, 29.5)	(76.5, 17.5)
6	(91, 28)	(147.5, 46.5)	(82, 34)
7	(73, 22)	(141.5, 32.5)	(89.5, 29.5)
8	(63.5, 14.5)	(169, 41)	(100.5, 24.5)
9	(55,12)	(119.5, 25.5)	(75.5, 28.5)
10	(78.5,23.5)	(174.5, 26.5)	(109, 21)
11	(65,13)	(165.5, 46.5)	(70, 23)
12	(69.5,14.5)	(150, 28)	(89, 16)
13	(81,20)	(158, 31)	(99.5, 25.5)
14	(78.5,13.5)	(163, 50)	(82, 30)
15	(52,14)	(173, 32)	(101, 32)
16	(60.5,12.5)	(134, 35)	(81, 28)
17	(78.5,19.5)	(158.5, 32.5)	(79, 19)
18	(73,14)	(150,51)	(88, 33)
19	(65.5,16.5)	(154.5, 66.5)	(65.5, 28.5)
20	(62.5,14.5)	(148,35)	(70, 15)

We obtain

$$\hat{\beta}_1 = 0.0236, \hat{\beta}_2 = 0.0865, \hat{\mu}_a = 59.6572, \hat{l}_a = 14.8489.$$

Then the concerned linear regression equation is

$$\hat{\tilde{Y}} = (59.6572, 14.8489)_L \oplus 0.0236\tilde{x}_1 \oplus 0.0865\tilde{x}_2.$$

However, for the obtained estimators of the model, the Hukuhara difference based residuals $\hat{\tilde{\varepsilon}} = (\mu_\varepsilon, l_\varepsilon)$ may not exist for some data. The residuals computed with the Hukuhara difference formula [13] are shown in Table 2, where some residuals (fuzzy data) with negative spreads appeared.

Table 2. Data of the residual $\hat{\tilde{\varepsilon}}$

i	$(\mu_\varepsilon, l_\varepsilon)$	i	$(\mu_\varepsilon, l_\varepsilon)$	i	$(\mu_\varepsilon, l_\varepsilon)$
1	(3.7707, -1.1851)	8	(-8.8413, -3.4364)	15	(-20.479, -4.3728)
2	(-13.4607, -7.0548)	9	(-14.01, -5.9166)	16	(-9.3281, -5.5976)
3	(2.2043, 23.1227)	10	(5.2934, 6.2087)	17	(8.2666, 2.2401)
4	(14.964, 6.6149)	11	(-4.6199, -4.9364)	18	(2.1886, -4.9078)
5	(5.6954, -3.5593)	12	(-1.3979, -2.3941)	19	(-3.471, -2.3844)
6	(20.7667, 9.1119)	13	(9.0048, 2.2131)	20	(-6.7069, -2.4729)
7	(2.2595, 3.8317)	14	(7.9009, -5.1247)		

Thus, there are only 7 values of the Hukuhara difference based residuals for the observations of Table 1, that is,

$$B = \{(2.2043, 23.1227), (14.964, 6.6149), (20.7667, 9.1119),\\ (2.2595, 3.8317), (5.2934, 6.2087), (9.0048, 2.2131), (8.2666, 2.2401)\}.$$

In the following we give an example of distributional simulation for the disturbance term $\tilde{\varepsilon}$. Taking B as a bootstrap population [6] and randomly resampling times of 10000. Using SAS on the bootstrap data, we output the histograms for center variable and spread variable respctively. The hypotheses about the distributions for center variable and spread variable remain to be tested in our future research.

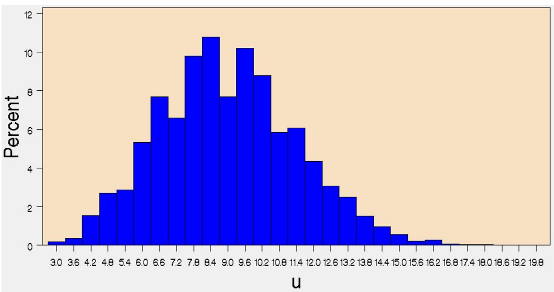

Fig. 1. The histogram for center variable of $\tilde{\varepsilon}$

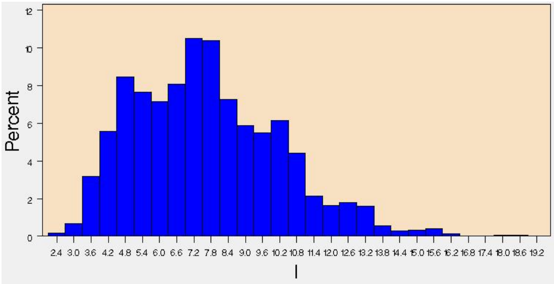

Fig. 2. The histograms histogram for spread variable of $\tilde{\varepsilon}$

Acknowledgment. The authors would like to thank the reviewers valuable comments. The research is supported by NSFC with grant number 11271096. Their financial support is gratefully acknowledged.

References

1. Sinova, B., Colubi, A., Gil, M.A., González-Rodríguez, G.: Interval arithmetic-based simple linear regression between interval data: Discussion and sensitivity analysis on choice of the metric. Information Sciences 199, 109–124 (2012)
2. Blanco-Fernández, A., Casals, M.R., Colubi, A., Corral, N., García-Bárzana, M., Gil, M.A., González-Rodríguez, G., López, M.T., De La Rosa De Sáa, S., Sinova, B.: Random fuzzy sets: A mathematical tool to develop statistical fuzzy data analysis. Iranian Journal of Fuzzy Systems 10(2), 1–28 (2013)
3. Cressie, N., Hulting, F.L.: A spatial statistical analysis of tumor growth. Journal of the American Statistical Association 87(418), 272–283 (1992)
4. Diamond, P., Kloeden, P.: Metric Spaces of Fuzzy Sets. World Scientific (1994)
5. Dimond, P., Körner, R.: Extended Fuzzy Linear Models and Least Squares Estimates. Computers and Mathematics with Applications 33, 15–32 (1997)
6. Efron and Tibshirani. An Introduction to the Bootstrap. Chapman & Hall (1993)
7. González-Rodríguez, G., Colubi, A., Gil, M.A.: Fuzzy data treated as functional data: A one-way ANOVA test approach. Computational Statistics and Data Analysis 56(4), 943–955 (2012)
8. Gil, M.A., Lubiano, M.A., Montenegro, M., López, M.T.: Least squares fitting of an affine function and strength of association for interval-valued data. Metrika 56, 97–111 (2002)
9. Krätschmer, V.: Least-squares estimation in linear regression models with vague concepts. Fuzzy Sets and Systems 157, 2579–2592 (2006)
10. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. 114, 409–422 (1986)
11. Tanaka, H., Uejima, S., Asai, K.: Linear regression analysis with fuzzy model. IEEE Trans. Systems Man Cybernet. 12, 903–907 (1982)
12. Näther, W.: Regression with fuzzy random data. Computational Statistics & Data Analysis 51, 235–252 (2006)
13. Näther, W.: Random fuzzy variables of second order and applications to statistical inference. Information Sciences 133, 69–88 (2001)
14. Wünsche, A., Näther, W.: Least -squares fuzzy regression with fuzzy random variables. Fuzzy Sets and Systems 130, 43–50 (2002)