FUZZY OPTIMIZATION AND MODELLING 1(2) (2020) 60-68

E-ISNN: 2676-7007

Contents lists available at FOMJ

Fuzzy Optimization and Modelling

Journal homepage: http://fomj.qaemiau.ac.ir/

Paper Type: Original Article

A Parameterized Approach for Linear Regression of Interval Data: Suggested Modifications

Abdullah Al-Qudaimi*

CSE Department, Hodeidah University, Hodeidah, Yemen

A R T I C L E I N F O

Article history:
Received 30 April 2020
Revised 21 May 2020
Accepted 2 June 2020
Available online 10 June 2020

Keywords:
Interval linear regression fuzzy
Symbolic data analysis
Interval parameterization

A B S T R A C T

Souza et al. (Knowledge-Based Systems, 131 (2017), pp. 149-159) pointed out that although several approaches have been proposed in the literature for fitting interval linear regression models (linear regression models its parameters are represented as intervals). However, as there are flaws in all the existing approaches, it is scientifically incorrect to use these approaches in real life problems. To resolve the flaws of the existing approaches, Souza et al. proposed a new approach for fitting interval linear regression models. After a deep study, it is observed that in the approach, proposed by Souza et al., some mathematical incorrect assumptions have been considered and hence, it is scientifically incorrect to use the Souza et al.'s approach, in real life problems. In this paper the mathematical incorrect assumption, considered by Souza et al, is pointed out and suggested modifications are provided as well as a new approach is proposed as for fitting the interval linear regression models. The proposed model guarantee mathematical coherent such that the predicted values of the model are intervals with lower bound less than or equal upper bound. Furthermore, the proposed has been illustrated with the help of a numerical example.

1. Introduction

Souza et al. [13, Section 1, pp. 149] pointed out that in all the existing approaches [1-13], proposed for fitting interval regression models, certain fixed reference points are used to transform an interval linear regression model into a crisp linear regression model whereas, the better results can be obtained by varying the reference points according to the collected data. Therefore, assuming some certain fixed points for the data of all the problems is mathematically incorrect. To resolve this problem, Souza et al. [13, Section 3.2, pp. 152] proposed a parameterized approach to fit an interval linear regression model. After a deep study, it is observed in Step 2 of the approach, proposed by Souza et al. [13] that the multiplication of two intervals is replaced by multiplication of an interval with a real number i.e., it is assumed that to find the multiplication of two intervals is equivalent to find the multiplication of an interval with real number, which is mathematically incorrect. Hence, it is scientifically incorrect to use the approach, proposed by Souza et al. [13], for fitting the interval linear regression models. Therefore, in this

* Corresponding author
E-mail Address: aalqudaimi@yahoo.com (Abdullah Al-Qudaimi)

paper, to modify Souza et al. [13] after discussing the existing approach [13] and pointing out incorrect assumptions, a new multiplication for two intervals is proposed and a new approach is proposed to construct linear regression model to fit interval valued data. Furthermore, the proposed is illustrated with the help of a numerical example.

2. Preliminary

This section is devoted for necessary basic definition and properties of bounded closed real intervals including arithmetic operations of mathematical intervals.

Definition 1: A bounded real interval A is a finite set of all real numbers contained between any two endpoints a^l and a^u called lower and upper bounds, respectively. If $\underline{a}$ and $\bar{a}$ are included the real bounded interval is called closed and is terminologically denoted $[\underline{a}\,,\ \bar{a}]$, i.e., $A = [\underline{a}\,,\ \bar{a}] = \{x \in R \mid \underline{a} \le x \le \bar{a}, \underline{a} = \inf(A)\,,\ \bar{a} = \sup(A)\}$. An interval $A = [a, a] = \{a\}$ is called degenerate interval which is a special cases of bounded closed interval. Henceforth, the concept of interval denotes mathematical bounded closed real interval.

Definition 2: An interval A may be expressed in the term of centric-point and range as follows:

$A = (a^c; a^r)$ where $a^c = \frac{\underline{a} + \bar{a}}{2}$ is the centric-point of A and $a^r = \underline{a} - \bar{a}$ is the range ($a^r \ge 0$) of A such that$\underline{a} = a^c - \frac{a^r}{2}$ and $\bar{a} = a^c + \frac{a^r}{2}$.

Definition 3 (Moore 1966): Let $A = [\underline{a}\,,\ \bar{a}]$ and $B = [\underline{b}\,,\ \bar{b}]$ are two intervals, then we have

- $A + B = [\underline{a} + \underline{b},\ \bar{a} + \bar{b}]$
- $A - B = [\underline{a} - \underline{b},\ \bar{a} - \bar{b}]$
- $\gamma \cdot A = \begin{cases} [\gamma\underline{a}, \gamma\bar{a}] & \gamma \ge 0 \\ [\gamma\bar{a}, \gamma\underline{a}] & \gamma < 0 \end{cases}$
- $A \cdot B = [\min\{\underline{a}\underline{b}, \underline{a}\bar{b}, \bar{a}\underline{b}, \bar{a}\bar{b}\}, \max\{\underline{a}\underline{b}, \underline{a}\bar{b}, \bar{a}\underline{b}, \bar{a}\bar{b}\}]$

Definition 4: An interval $A = [\underline{a}\,,\ \bar{a}]$ is said mathematically coherent interval iff $\underline{a} \le \bar{a}$, i.e., the lower bound of A is less than or equal the upper bound. Otherwise A is mathematically incoherent.

3. Existing parameterized approach

To point out the mathematical incorrect assumptions, considered by Souza et al. [13, Section 3.2, pp. 152] in their proposed approach, there is need to discuss this approach. Therefore, the same is discussed in this section.

Souza et al. [1, Section 3.2, pp. 152] proposed the following approach to fit the interval linear regression model (1).

$$[\underline{y}\,,\ \bar{y}] = [\underline{\beta}_0,\ \bar{\beta}_0] + \sum_{j=1}^{p} [\underline{\beta}_j,\ \bar{\beta}_j]\, [\underline{x}_j\,,\ \bar{x}_j] + [\epsilon^l,\ \epsilon^u] \tag{1}$$

Step 1: Using the existing relation [14], $[\underline{x}_j\,,\ \bar{x}_j] = \{(1-\lambda)\underline{x}_j + \lambda\bar{x}_j : 0 \le \lambda \le 1\}$, the interval linear regression model (1) can be transformed into its equivalent interval linear regression model (2).

$$[\underline{y}\,,\ \bar{y}] = [\underline{\beta}_0,\ \bar{\beta}_0] + \sum_{j=1}^{p} [\underline{\beta}_j,\ \bar{\beta}_j]\, \{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\} + [\epsilon^l,\ \epsilon^u] \tag{2}$$

Step 2: Using the existing relation [14], $\lambda[a, b] = [\lambda a, \lambda b], \lambda \ge 0$, the interval linear regression model (2) can be transformed into its equivalent interval linear regression model (3).

$$[\underline{y}\,, \bar{y}] = [\underline{\beta}_0, \bar{\beta}_0] + \sum_{j=1}^{p} [\underline{\beta}_j\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\}\,,\ \bar{\beta}_j\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\}] + [\epsilon^l,\ \epsilon^u] \tag{3}$$

Step 3: Assuming, $\alpha_j^l = \underline{\beta}_j(1-\lambda_j)$, $\omega_j^l = \underline{\beta}_j\lambda_j$, $\alpha_j^u = \bar{\beta}_j(1-\lambda_j)$ and $\omega_j^u = \bar{\beta}_j\lambda_j$, the interval linear regression models (3) can be transformed into its equivalent interval linear regression model (4).

$$[\underline{y}\,,\ \bar{y}] = [\underline{\beta}_0,\ \bar{\beta}_0] + \sum_{j=1}^{p} [\alpha_j^l\underline{x}_j + \omega_j^l\bar{x}_j\,,\ \alpha_j^u\underline{x}_j + \omega_j^u\bar{x}_j] + [\epsilon^l,\ \epsilon^u] \tag{4}$$

Step 4: Using the existing relations [14], $[a\ ,\ b] + \sum_{i=1}^{n}[a_i\ ,\ b_i] = [a + \sum_{i=1}^{n} a_i\ ,\ \ b + \sum_{i=1}^{n} b_i]$ and $[a\ ,\ b] = [c\ , d] \Rightarrow a = b, c = d$ the interval linear regression model (4) can be transformed into its equivalent crisp linear regression models (5) and (6).

$$\underline{y} = \underline{\beta}_0 + \sum_{j=1}^{p} \alpha_j^l \underline{x}_j + \sum_{j=1}^{p} \omega_j^l \bar{x}_j + \epsilon^l \tag{5}$$

$$\bar{y} = \bar{\beta}_0 + \sum_{j=1}^{p} \alpha_j^u \underline{x}_j + \sum_{j=1}^{p} \omega_j^u \bar{x}_j + \epsilon^u \tag{6}$$

Step 5: If $\underline{y}_i$ and $\bar{y}_i$ are the known values of $\underline{y}$ and $\bar{y}$ corresponding to $\underline{x}_{ji}$ and $\bar{x}_{ji}$, $i = 1,2, \dots, n$; $j = 1,2, \dots, p$ then using the equations (5) and (6), the sum of squares of errors in the predicted values of $\underline{y}_i$ and $\bar{y}_i$ can be obtained by equations (7) and (8) respectively.

$$\sum_{i=1}^{n} \left(\epsilon_i^l\right)^2 = \sum_{i=1}^{n} \left(\underline{y}_i - \underline{\beta}_0 - \sum_{j=1}^{p} \alpha_j^l \underline{x}_{ji} - \sum_{j=1}^{p} \omega_j^l \bar{x}_{ji}\right)^2 \tag{7}$$

$$\sum_{i=1}^{n} (\epsilon_i^u)^2 = \sum_{i=1}^{n} \left(\bar{y}_i - \bar{\beta}_0 - \sum_{j=1}^{p} \alpha_j^u \underline{x}_{ji} - \sum_{j=1}^{p} \omega_j^u \bar{x}_{ji}\right)^2 \tag{8}$$

Step 6: According to the principle of least squares [15], the values of β_0^l, α_j^l and ω_j^l, corresponding to which the sum of squares of errors $\sum_{i=1}^{n} \left(\epsilon_i^l\right)^2$ will be minimum, can be obtained by partially differentiating Eq. (7) with respect to $\underline{\beta}_0$, α_j^l and ω_j^l. Similarly, the values of $\bar{\beta}_0$, α_j^u and ω_j^u, corresponding to which the sum of squares of errors $\sum_{i=1}^{n} (\epsilon_i^u)^2$ will be minimum, can be obtained by partially differentiating Eq. (8) with respect to $\bar{\beta}_0$, α_j^u and ω_j^u.

On partially differentiating Eq. (7) with respect to $\underline{\beta}_0$, α_j^l and ω_j^l, the crisp system of linear equations (9), (10) and (11) respectively are obtained and on partially differentiating Eq. (8) with respect to $\bar{\beta}_0$, α_j^u and ω_j^u, the crisp system of linear equations (12), (13) and (14) respectively are obtained.

$$\sum_{i=1}^{n} \underline{y}_i = n\underline{\beta}_0 + \sum_{j=1}^{p} \alpha_j^l \left(\sum_{i=1}^{n} \underline{x}_{ji}\right) + \sum_{j=1}^{p} \omega_j^l \left(\sum_{i=1}^{n} \bar{x}_{ji}\right) \tag{9}$$

$$\sum_{i=1}^{n} \underline{x}_{ki}\, \underline{y}_i = \underline{\beta}_0 \sum_{i=1}^{n} \underline{x}_{ki} + \sum_{j=1}^{p} \alpha_j^l \left(\sum_{i=1}^{n} \underline{x}_{ji}\, \underline{x}_{ki}\right) + \sum_{j=1}^{p} \omega_j^l \left(\sum_{i=1}^{n} \bar{x}_{ji} \underline{x}_{ki}\right),\ k = 1,2, \dots, p. \tag{10}$$

$$\sum_{i=1}^{n} \bar{x}_{ki}\, \underline{y}_i = \underline{\beta}_0 \sum_{i=1}^{n} \bar{x}_{ki} + \sum_{j=1}^{p} \alpha_j^l \left(\sum_{i=1}^{n} \underline{x}_{ji} \bar{x}_{ki}\right) + \sum_{j=1}^{p} \omega_j^l \left(\sum_{i=1}^{n} \bar{x}_{ji}\, \bar{x}_{ki}\right),\ k = 1,2, \dots, p. \tag{11}$$

$$\sum_{i=1}^{n} \bar{y}_i = n\bar{\beta}_0 + \sum_{j=1}^{p} \alpha_j^u \left(\sum_{i=1}^{n} \underline{x}_{ji}\right) + \sum_{j=1}^{p} \omega_j^u \left(\sum_{i=1}^{n} \bar{x}_{ji}\right) \tag{12}$$

$$\sum_{i=1}^{n} \underline{x}_{ki}\, \bar{y}_i = \bar{\beta}_0 \sum_{i=1}^{n} \underline{x}_{ki} + \sum_{j=1}^{p} \alpha_j^u \left(\sum_{i=1}^{n} \underline{x}_{ji}\, \underline{x}_{ki}\right) + \sum_{j=1}^{p} \omega_j^u \left(\sum_{i=1}^{n} \bar{x}_{ji} \underline{x}_{ki}\right),\ k = 1,2, \dots, p. \tag{13}$$

$$\sum_{i=1}^{n} \bar{x}_{ki}\, \bar{y}_i = \bar{\beta}_0 \sum_{i=1}^{n} \underline{x}_{ki} + \sum_{j=1}^{p} \alpha_j^u \left(\sum_{i=1}^{n} \underline{x}_{ji} \bar{x}_{ki}\right) + \sum_{j=1}^{p} \omega_j^u \left(\sum_{i=1}^{n} \bar{x}_{ji}\, \bar{x}_{ki}\right),\ k = 1,2, \dots, p. \tag{14}$$

Step 7: Find the values of $\underline{\beta}_0$, α_j^l and ω_j^l, $j = 1,2, \dots, p$ by solving the crisp system of linear equations (9) to (11) and find the values of $\bar{\beta}_0$, α_j^u and $\omega_j^u, j = 1,2, \dots, p$ by solving the crisp system of linear equations (12) to (14).

Step 8: Find the interval linear regression model by putting the values of $\underline{\beta}_0$, $\bar{\beta}_0$, α_j^l, α_j^u, ω_j^l and ω_j^u, obtained in Step 7, in the interval linear regression model (4).

4. Mathematical incorrect assumptions considered in the existing parameterized approach

In the existing parameterized approach [13, Section 3.2, pp. 152], discussed in Section 3, the following mathematical incorrect assumptions have been considered.

In Step 1 of the existing approach [13], discussed in the Section 2, the interval linear regression model (1) has been transformed into the interval linear regression model (2) by considering the relation (15)

$$\left[\underline{\beta}_j, \bar{\beta}_j\right] \left[\underline{x}_j\ , \bar{x}_j\right] = \left[\underline{\beta}_j, \bar{\beta}_j\right] \left\{\left(1 - \lambda_j\right)\underline{x}_j + \lambda_j \bar{x}_j\right\}$$

However, the following example clearly indicates that it is mathematically incorrect to assume the same.

Let $\left[\underline{\beta}_j,\ \bar{\beta}_j\right] = [1\ ,\ 2]$, $\left[\underline{x}_j\ ,\ \bar{x}_j\right] = [3\ ,4]$, then

$$\left[\underline{\beta_j},\ \bar{\beta}_j\right]\left[\underline{x}_j\ ,\ \bar{x}_j\right] = [1,2][3,4] = [3,8] \tag{16}$$

$$\left[\underline{\beta_j},\ \bar{\beta}_j\right]\left\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\right\} = [1,2]\left\{3(1-\lambda_j) + 4\lambda_j\right\} = [3+\lambda_j, 6+2\lambda_j] \tag{17}$$

Now, using the relation (15) for (16) & (17),

$[3,8] = [3+\lambda_j, 6+2\lambda_j] \Rightarrow\ 3 = 3+\lambda_j$ and $8 = 6+2\lambda_j \Rightarrow \lambda_j = 0$ and $\lambda_j = 1$.

5. The effect of the mathematical incorrect assumptions

To point out the effect of mathematical incorrect assumptions, discussed in Section 3, the existing parameterized approach [13, Section 3.2, pp. 152], discussed in Section 3, is applied to obtain an interval linear regression model and shown that the obtained model is not valid. It is obvious that there does not exist any value of λ_j for which the relation (15) will be satisfied. Hence, it is mathematically incorrect to transform interval linear regression model (1) into interval linear regression model (2) by considering the relation (15).

5.1 Interval linear regression model

Let $[\underline{x}_{11}\ ,\ \bar{x}_{11}] = [1\ ,\ 2]$, $[\underline{x}_{12}\ ,\ \bar{x}_{12}] = [3\ ,\ 4$, $\left[\underline{y}_1\ ,\ \bar{y}_1\right] = [12\ ,\ 26]$ and $\left[\underline{y}_2\ ,\ \bar{y}_2\right] = [4\ ,30]$. Then, using the existing parameterized approach [1, Section 3.2, pp. 152], discussed in Section 3, the interval linear regression model (18), for the considered data can be obtained as follows:

$$\left[\underline{y}\ ,\ \bar{y}\right] = \left[\underline{\beta_0},\ \bar{\beta}_0\right] + \left[\underline{\beta_1},\ \bar{\beta}_1\right]\left[\underline{x_1}\ ,\ \bar{x}_1\right] + [\varepsilon^l, \varepsilon^u] \tag{18}$$

Step 1: Using the existing relation [14], $\left[\underline{x_1}\ ,\ \bar{x}_1\right] = \left\{(1-\lambda_1)\underline{x_1} + \lambda_1\bar{x}_1 : 0 \le \lambda_1 \le 1\right\}$, the interval linear regression model (18) can be transformed into its equivalent interval linear regression model (19).

$$\left[\underline{y}\ ,\ \bar{y}\right] = \left[\underline{\beta_0},\ \bar{\beta}_0\right] + \left[\underline{\beta_1},\ \bar{\beta}_1\right]\left((1-\lambda_1)\underline{x_1} + \lambda_1\bar{x}_1\right) + [\varepsilon^l, \varepsilon^u] \tag{19}$$

Step 2: Using the existing relation [14], $\lambda[a,b] = [\lambda a, \lambda b], \lambda \ge 0$, the interval linear regression model (19) can be transformed into its equivalent interval linear regression model (20).

$$\left[\underline{y}\ ,\ \bar{y}\right] = \left[\underline{\beta_0},\ \bar{\beta}_0\right] + \left[\underline{\beta_1}(1-\lambda_1)\underline{x_1} + \underline{\beta_1}\lambda_1\bar{x}_1,\ \bar{\beta}_1(1-\lambda_1)\underline{x_1} + \bar{\beta}_1\lambda_1\bar{x}_1\right] + [\varepsilon^l, \varepsilon^u] \tag{20}$$

Step 3: Assuming $\alpha_1^l = \beta_1^l(1-\lambda_1)$, $\omega_1^l = \beta_1^l\lambda_1$, $\alpha_1^u = \beta_1^u(1-\lambda_1)$ and $\omega_1^u = \beta_1^u\lambda_1$, the interval linear regression model (20) can be transformed into its equivalent interval linear regression model (21).

$$\left[\underline{y}\ ,\ \bar{y}\right] = \left[\underline{\beta_0},\ \bar{\beta}_0\right] + \left[\alpha_1^l\underline{x_1} + \omega_1^l\bar{x}_1,\ \alpha_1^u\underline{x_1} + \omega_1^u\bar{x}_1\right] + [\varepsilon^l, \varepsilon^u] \tag{21}$$

Step 4: Using the existing relations [14], $\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}[a_i,\ b_i] = [\sum_{i=1}^{n} a_i\ ,\ \sum_{i=1}^{n} b_i]$ and $[a,\ b] = [c,\ d] \Rightarrow a = c, b = d$, the interval linear regression model (21) can be transformed into the equivalent crisp linear regression models (22) and (23).

$$\underline{y} = \underline{\beta_0} + \alpha_1^l\underline{x_1} + \omega_1^l\bar{x}_1 + \varepsilon^l \tag{22}$$

$$\bar{y} = \bar{\beta}_0 + \alpha_1^u\underline{x_1} + \omega_1^u\bar{x}_1 + \varepsilon^u \tag{23}$$

Step 5: Using Eq. (22) the sum of the squares of errors in the predicted values of $\underline{y_1}$ and $\underline{y_2}$ can be obtained by Eq. (24) and using Eq. (23) the sum of the squares of errors in the predicted values of $\bar{y}_1$ and $\bar{y}_2$ can be obtained by Eq. (25).

$$\left(\varepsilon_1^l\right)^2 + \left(\varepsilon_2^l\right)^2 = \left(\underline{y_1} - \underline{\beta_0} - \alpha_1^l\underline{x_{11}} - \omega_1^l\bar{x}_{11}\right)^2 + \left(\underline{y_2} - \underline{\beta_0} - \alpha_1^l\underline{x_{12}} - \omega_1^l\bar{x}_{12}\right)^2 \tag{24}$$

$$(\varepsilon_1^u)^2 + (\varepsilon_2^u)^2 = \left(\bar{y}_1 - \bar{\beta}_0 - \alpha_1^u\underline{x_{11}} - \omega_1^u\bar{x}_{11}\right)^2 + \left(\bar{y}_2 - \bar{\beta}_0 - \alpha_1^u\underline{x_{12}} - \omega_1^u\bar{x}_{12}\right)^2 \tag{25}$$

Step 6: Partially differentiating Eq. (24) with respect to $\underline{\beta_0}$, α_1^l and ω_1^l, the crisp system of linear equations (26) to (28) is obtained and partially differentiating Eq. (25) with respect to $\bar{\beta}_0$, α_1^u and ω_1^u the crisp system of linear equations (28) to (31) is obtained .

$$\underline{y}_1 + \underline{y}_2 = 2\underline{\beta}_0 + \alpha_1^l(\underline{x}_{11} + \underline{x}_{12}) + \omega_1^l(\bar{x}_{11} + \bar{x}_{12}) \quad (26)$$
$$\underline{x}_{11}\underline{y}_1 + \underline{x}_{12}\underline{y}_2 = \underline{\beta}_0(\underline{x}_{11} + \underline{x}_{12}) + \alpha_1^l(\underline{x}_{11}{}^2 + \underline{x}_{12}{}^2) + \omega_1^l(\underline{x}_{11}\bar{x}_{11} + \underline{x}_{12}\bar{x}_{12}) \quad (27)$$
$$\bar{x}_{11}\underline{y}_1 + \bar{x}_{12}\underline{y}_2 = \underline{\beta}_0(\bar{x}_{11} + \bar{x}_{12}) + \alpha_1^l(\underline{x}_{11}\bar{x}_{11} + \underline{x}_{12}\bar{x}_{12}) + \omega_1^l(\bar{x}_{11}{}^2 + \bar{x}_{12}{}^2) \quad (28)$$
$$\bar{y}_1 + \bar{y}_2 = 2\bar{\beta}_0 + \alpha_1^u(\underline{x}_{11} + \underline{x}_{12}) + \omega_1^u(\bar{x}_{11} + \bar{x}_{12}) \quad (29)$$
$$\underline{x}_{11}\bar{y}_1 + \underline{x}_{12}\bar{y}_2 = \bar{\beta}_0(\underline{x}_{11} + \underline{x}_{12}) + \alpha_1^u(\underline{x}_{11}{}^2 + \underline{x}_{12}{}^2) + \omega_1^u(\underline{x}_{11}\bar{x}_{11} + \underline{x}_{12}\bar{x}_{12}) \quad (30)$$
$$\underline{x}_{11}\bar{y}_1 + \underline{x}_{12}\bar{y}_2 = \bar{\beta}_0(\bar{x}_{11} + \bar{x}_{12}) + \alpha_1^u(\underline{x}_{11}\bar{x}_{11} + \underline{x}_{12}\bar{x}_{12}) + \omega_1^u(\bar{x}_{11}{}^2 + \bar{x}_{12}{}^2) \quad (31)$$

Since, in the considered problem $\underline{x}_{11} = 1$, $\bar{x}_{11} = 2$, $\underline{x}_{12} = 3$, $\bar{x}_{12} = 4$, $\underline{y}_1 = 12$, $\bar{y}_1 = 26$, $\underline{y}_2 = 4$ and $\bar{y}_2 = 30$. So, putting these values in equations (26) to (31) these equations are transformed into equations (32) to (37) respectively.

$$32 = 4\underline{\beta}_0 + 8\alpha_1^l + 12\omega_1^l \quad (32)$$
$$48 = 8\underline{\beta}_0 + 20\alpha_1^l + 28\omega_1^l \quad (33)$$
$$80 = 12\underline{\beta}_0 + 28\alpha_1^l + 40\omega_1^l \quad (34)$$
$$112 = 4\bar{\beta}_0 + 8\alpha_1^u + 12\omega_1^u \quad (35)$$
$$232 = 8\bar{\beta}_0 + 20\alpha_1^u + 28\omega_1^u \quad (36)$$
$$344 = 12\bar{\beta}_0 + 28\alpha_1^u + 40\omega_1^u \quad (37)$$

Step 7: On solving the crisp linear equations (32) to (34), the obtained values of $\underline{\beta}_0$, α_1^l and ω_1^l are $12, -8$ and 4 respectively and on solving the crisp linear equations (35) to (37), the obtained values of $\bar{\beta}_0$, α_1^u and ω_1^u are $\frac{46}{3}, -\frac{20}{3}$ and $\frac{26}{3}$ respectively.

Step 8: Putting the values of $\underline{\beta}_0$, α_1^l, ω_1^l, $\bar{\beta}_0$, α_1^u and ω_1^u, obtained in Step 7, in the interval linear regression model (18), the interval linear regression model (38) is obtained for the considered data.

$$[\underline{y}\,,\ \bar{y}] = \left[\frac{5}{2},\ \frac{9}{2}\right] + \left[\frac{1}{2}\underline{x}_1,\ \frac{1}{2}\bar{x}_1\right] + [\varepsilon^l, \varepsilon^u]. \quad (38)$$

5.2 Chen and Nien's approach

Since for the values of f $\underline{\beta}_0$, α_1^l, ω_1^l, $\bar{\beta}_0$, α_1^u and ω_1^u obtained in Step 7 of Section 4.1, the conditions $\underline{\beta}_0 \leq \underline{\beta}_0$, $\alpha_1^l \leq \alpha_1^u$, $\omega_1^l \leq \omega_1^u$ are satisfying so, according to the existing parameterized approach [13, Section 3.2, pp. 152], discussed in Section 3, the obtained interval linear regression model (35) is valid. However, in actual case the interval linear regression model (32) is not valid due to the following reasons.

1. Since, the interval linear regression model $[\underline{y}\,,\ \bar{y}] = \left[\frac{5}{2},\ \frac{9}{2}\right] + \left[\frac{1}{2}\underline{x}_1\,,\ \frac{1}{2}\bar{x}_1\right] + [\varepsilon^l, \varepsilon^u]$ has been assumed by considering the data, $[\underline{x}_{11}\,,\ \bar{x}_{11}] = [1,2]$, $[\underline{x}_{12}\,,\ \bar{x}_{12}] = [3,5]$, $[\underline{y}_1,\ \bar{y}_1] = [3,5]$ and $[\underline{y}_2,\ \bar{y}_2] = [4,6]$. Therefore, on putting $\underline{x}_{12} = 3$ and $\bar{x}_{12} = 5$ in the left hand side of the obtained model, interval $[\underline{y}_1,\ \bar{y}_1] = [3,5]$should be obtained. However, it can be easily verified that on putting $\underline{x}_{12} = 3$ and $\bar{x}_{12} = 5$ in the obtained model, the interval $[\underline{y}_2,\ \bar{y}_2] = [4,6]$ is obtained. It is clearly indicates that the assumed model is not valid.

2. $\alpha_1^l = (1-\lambda_1)\,\underline{\beta}_1and\omega_1^l = \lambda_1\,\underline{\beta}_1 \Rightarrow \frac{\alpha_1^l}{\omega_1^l} = \frac{1-\lambda_1}{\lambda_1} \Rightarrow \frac{\alpha_{1\lambda_1}^l}{\omega_1^l} = 1 - \lambda_1 \Rightarrow \lambda_1\left(\frac{\alpha_1^l+\omega_1^l}{\omega_1^l}\right) = 1 \Rightarrow \lambda_1 = \frac{\omega_1^l}{\alpha_1^l+\omega_1^l}$.

Since, for the value of λ_1 the condition $0 \leq \lambda_1 \leq 1$ i.e., $0 \leq \frac{\omega_1^l}{\alpha_1^l+\omega_1^l} \leq 1$ should always be satisfied. Therefore, the values of α_1^l and ω_1^l obtained in Step 7 of Section 4.1, and hence, the interval linear regression model (35) will be valid only if for the obtained values of α_1^l and ω_1^l, the condition $0 \le \frac{\omega_1^l}{\alpha_1^l+\omega_1^l} \le 1$ will be satisfied. However, it can be easily verified that for the values of α_1^l and ω_1^l, the condition $0 \le \frac{\omega_1^l}{\alpha_1^l+\omega_1^l} \le 1$ is not satisfying.

Similarly, for $\lambda_1 = \frac{\omega_1^u}{\alpha_1^u+\omega_1^u}$, obtained by using the relations $\alpha_1^u = (1-\lambda_1)\bar{\beta}_0$ and $\omega_1^u = \lambda_1\bar{\beta}_0$, the condition $0 \le \lambda_1 \le 1$ i.e., $0 \le \frac{\omega_1^u}{\alpha_1^u+\omega_1^u} \le 1$ should always be satisfied. However, it can be easily verified that for the values of α_1^u and ω_1^u the condition $0 \le \frac{\omega_1^u}{\alpha_1^u+\omega_1^u} \le 1$ is not satisfying. Therefore, the values of α_1^l, ω_1^l, α_1^u, ω_1^u, obtained in Step 7 of Section 4.1, and hence the obtained interval linear regression model (35), obtained in Step 8 of Section 4.1, are not valid.

6. Limitations of Souza et al. parameterized approach

The following clearly indicates that Souza et al. approach has been proposed to fit the interval linear regression model for such problems in which for the known values$[\underline{x}_{ij}\ ,\ \bar{x}_{ij}]$, the condition $\underline{x}_{ij} \ge 0$ will be satisfied.

In Souza et al. approach, discussed in Section 2, the interval linear regression model (2) has been transformed into the interval linear regression model (3) by considering the $[\underline{\beta}_j,\ \bar{\beta}_j]\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\} = [\underline{\beta}_j\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\}$, $\bar{\beta}_j\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\}]$.

It is pertinent to mention that this relation will be valid only if $(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j$ is a negative real number. Furthermore, as the minimum value of $(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j$ will be $\underline{x}_j$ and the maximum value will be $\overline{x}_j$. Therefore, $(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j$ will be negative real number only if $\underline{x}_j$ is a negative real number i.e., for the known values of $[\underline{x}_j, \bar{x}_j]$, the condition $\underline{x}_j \ge 0$ will be satisfied.

The following example clearly indicates that if for the known values of $[\underline{x}_j, \bar{x}_j]$, the condition $\underline{x}_j \ge 0$ will not be satisfied then $[\underline{\beta}_j,\ \bar{\beta}_j]\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\} \ne [\underline{\beta}_j\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\},\ \bar{\beta}_j\{(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j\}]$.

If $[\underline{x}_j, \bar{x}_j] = [-3,-2]$ then $(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j = (1-\lambda_j)(-3) + \lambda_j(-2) = -3 - 5\lambda_j$.

Since $0 \le \lambda_j \le 1$, $(1-\lambda_j)\underline{x}_j + \lambda_j\bar{x}_j$ will be a negative real number for all values of λ_j, it is incorrect to use relation (19).

7. Proposed multiplication of intervals

The objective of this paper is to propose a new approach for fitting interval linear regression models. To achieve this objective, there is need to use the multiplication of a known interval with an unknown interval. Although it is well known fact that if $[a_1, a_2]$ and $[b_1, b_2]$ are two intervals then the multiplication of these intervals is defined as $[a_1, a_2][b_1, b_2] = [min\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}, \max\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}]$. But, more computational efforts are required to apply this multiplication directly. Therefore, using this multiplication, new multiplication has been defined to find some type of multiplication, with the help of the existing multiplication. , $a_1 \ge 0$.

Let $[a_1, a_2]$ be an unknown interval and $[b_1, b_2]$ be a known interval then,

$$[a_1, a_2][b_1, b_2] = \begin{cases} [\min\{a_1b_1, a_1b_2\}, \max\{a_2b_1, a_2b_2\}] & b_1 \ge 0 \\ [\min\{a_1\mathrm{b}_2, a_2\mathrm{b}_1\}, \max\{a_1\mathrm{b}_1, a_2\mathrm{b}_2\}] & b_1 < 0,\ b_2 \ge 0. \\ [\min\{a_2b_1, a_2b_2\}, \max\{a_1b_1, a_1b_2\}] & b_2 < 0 \end{cases}$$

8. Proposed approach

In this section, a new approach proposed to fit an interval linear regression model.

The steps of the proposed approach are as follows:

Step 1: Assuming $\left[\underline{\beta}_j,\ \bar{\beta}_j\right]\left[\underline{x}_{ij}\ ,\ \bar{x}_{ij}\right] = \left[Y_j^L, Y_j^U\right]$ where $\left[Y_j^L, Y_j^U\right]$ will be obtained according to the proposed multiplication defined in Section 5, the interval linear regression model (1) can be transformed into the interval linear regression (36) is

$$\left[\underline{y}\ ,\bar{y}\right] = \left[\underline{\beta}_0,\ \bar{\beta}_0\right] + \sum_{j=1}^{p}\left[Y_j^L, Y_j^U\right] + \left[\varepsilon^l, \varepsilon^u\right] \tag{36}$$

Step 2: Using the existing relation $\sum_{i=1}^{n}[a_i, b_i] = \left[\sum_{i=1}^{n} a_i\ , \sum_{i=1}^{n} b_i\right]$, the interval linear regression model (36) can be transformed into its equivalent interval linear regression model (37)

$$\left[\underline{y}\ ,\bar{y}\right] = \left[\underline{\beta}_0,\ \bar{\beta}_0\right] + \left[\sum_{j=1}^{p} Y_j^L\ ,\sum_{j=1}^{p} Y_j^U\right] + \left[\varepsilon^l, \varepsilon^u\right] \tag{37}$$

Step 3: Using the relation $[a_1, a_2] + [b_1, b_2] = [a_1 + b_1, a_2 + b_2]$, the interval linear regression model (37) can be transformed into the interval linear regression model (38).

$$\left[\underline{y}\ ,\bar{y}\right] = \left[\underline{\beta}_0 + \sum_{j=1}^{p} Y_j^L\ ,\bar{\beta}_0 + \sum_{j=1}^{p} Y_j^U\right] + \left[\varepsilon^l, \varepsilon^u\right] \tag{38}$$

Step 4: Using the relation, $[a_1, b_1] = [a_2, b_2] \Longrightarrow a_1 = a_2, b_1 = b_2$, the interval linear regression model (38) can be transformed into the crisp linear regression models (39) and (40).

$$\underline{y} = \underline{\beta}_0 + \sum_{j=1}^{p} Y_j^L + \varepsilon^l \tag{39}$$

$$\bar{y} = \bar{\beta}_0 + \sum_{j=1}^{p} Y_j^U + \varepsilon^u \tag{40}$$

Step 5: If $\underline{y_i}$ and $\bar{y}_i$ are the known values of $\underline{y}$ and $\bar{y}$ corresponding to $\underline{x}_{ij}$ and $\bar{x}_{ij}$, $i = 1,2, \dots, n$ and $j = 1, 2, \dots, p$ then with the help of equations (5) and (6) the sum of the absolute values of errors in the predicted values of $\underline{y_i}$ and $\bar{y}_i$ can be obtained by using the equations (41) and (42) respectively.

$$\sum_{i=1}^{n}\left|\varepsilon_i^l\right| = \sum_{i=1}^{n}\left|\underline{y_i} - \underline{\beta}_0 - \sum_{j=1}^{p} Y_j^L\right| \tag{41}$$

$$\sum_{i=1}^{n}\left|\varepsilon_i^u\right| = \sum_{i=1}^{n}\left|\bar{y}_i - \bar{\beta}_0 - \sum_{j=1}^{p} Y_j^U\right| \tag{42}$$

Step 6: Find those values of β_j^l and β_j^u, $j = 0, 1, 2, \dots, p$ corresponding to which

$\sum_{i=1}^{n}\left|\underline{y_i} - \underline{\beta}_0 - \sum_{j=1}^{p} Y_j^L\right| + \sum_{i=1}^{n}\left|\bar{y}_i - \bar{\beta}_0 - \sum_{j=1}^{p} Y_j^U\right|$ will be minimum. i.e., find the optimal solution of the mathematical programming

$$minimize\ \sum_{i=1}^{n}\left|\underline{y_i} - \underline{\beta}_0 - \sum_{j=1}^{p} Y_j^L\right| + \sum_{i=1}^{n}\left|\bar{y}_i - \bar{\beta}_0 - \sum_{j=1}^{p} Y_j^U\right|$$

Step 7: Find the interval linear regression model by putting the values of $\underline{\beta_j}$ and $\bar{\beta}_j$, $j = 0, 1, 2, \dots, p$, obtained in Step 6, in the interval linear regression model (4).

9. Illustrative example

It is shown that on applying the existing parameterized approach [1, Section 3.2, pp. 152], discussed in Section 3, for the data $\left[\underline{x}_{11}\ \bar{x}_{11}\right] = [1\ 2]$, $\left[\underline{x}_{12}\ \bar{x}_{12}\right] = [3\ 5]$, $\left[\underline{y}_1\ \bar{y}_1\right] = [3\ 5]$ and $\left[\underline{y}_2\ \bar{y}_2\right] = [4\ 6]$. It is not possible to fit a correct interval linear regression model. In this section, on applying the existing parameterized approach [13, Section 3.2, pp. 152] with the suggested modifications for this considered data, an exact interval linear regression model is obtained.

Using the existing parameterized approach [1, Section 3.2, pp. 152], discussed in Section 3, with the modifications, suggested in Section 4, the exact interval linear regression model for considered data can be obtained as follows:

Putting $\underline{x}_{11} = 1$, $\bar{x}_{11} = 2$, $\underline{x}_{12} = 3$, $\bar{x}_{12} = 5$, $\underline{y}_1 = 3$, $\bar{y}_1 = 5$, $\underline{y}_2 = 4$ and $\bar{y}_2 = 6$ in the mathematical programming problem (P1), it is transformed into the mathematical programming (P2).

Minimize $[0\underline{\beta}_0 + 0\bar{\beta}_0 + 0\alpha_1^l + 0\alpha_1^u + 0\omega_1^l + 0\omega_1^u]$

Subject to

$2\underline{\beta}_0 + 4\alpha_1^l + 7\omega_1^l = 7$, $4\underline{\beta}_0 + 10\alpha_1^l + 17\omega_1^l = 15$, $7\beta_0^l + 17\alpha_1^l + 29\omega_1^l = 26$, $2\bar{\beta}_0 + 4\alpha_1^u + 7\omega_1^u = 11$, $4\bar{\beta}_0 + 10\alpha_1^u + 17\omega_1^u = 23$, $7\bar{\beta}_0 + 17\alpha_1^u + 29\omega_1^u = 40$,

$\underline{\beta}_0 \le \bar{\beta}_0$, $\alpha_1^l \le \alpha_1^u$, $\omega_1^l \le \omega_1^u$, $\frac{\alpha_1^l+\omega_1^l}{2} - \frac{1}{2}\left|\alpha_1^l + \omega_1^l\right| \le \omega_1^l$, $\omega_1^l \le \frac{\alpha_1^l+\omega_1^l}{2} + \frac{1}{2}\left|\alpha_1^l + \omega_1^l\right|$, $\frac{\alpha_1^u+\omega_1^u}{2} - \frac{1}{2}|\alpha_1^u + \omega_1^u| \le \omega_1^u$, $\omega_1^u \le \frac{\alpha_1^u+\omega_1^u}{2} + \frac{1}{2}|\alpha_1^u + \omega_1^u|$. (P2)

On solving the mathematical programming problem (P2), the obtained solution is $\underline{\beta}_0 = \frac{5}{2}$, $\bar{\beta}_0 = \frac{9}{2}$, $\alpha_1^l = \frac{1}{2}$, $\alpha_1^u = \frac{1}{2}$, $\omega_1^l = 0$, $\omega_1^u = 0$.

Putting these values in $\lambda_1 = \frac{\omega_1^l}{(\alpha_1^l+\omega_1^l)}$ as well as in $\lambda_1 = \frac{\omega_1^u}{(\alpha_1^u+\omega_1^u)}$, the obtained values of λ_1 is 0. Since for the obtained value of λ_1, the condition $0 \le \lambda_1 \le 1$ is satisfying. Therefore, these obtained values of $\underline{\beta}_0$, α_1^l, ω_1^l, $\bar{\beta}_0$, α_1^u and ω_1^u are valid.

Furthermore, putting these values in the interval linear regression model (18), $\left[\underline{y}\,, \bar{y}\right] = \left[\underline{\beta}_0\,, \bar{\beta}_0\right] + \left[\alpha_j^l \underline{x}_j + \omega_j^l \bar{x}_j \;\; \alpha_1^u \underline{x}_j + \omega_j^u \bar{x}_j\right] + [\epsilon^l,\ \epsilon^u]$, the vaild interval linear regression model for the considered data is as follows

$$\left[\underline{y}\,, \bar{y}\right] = \left[\frac{5}{2}\,, \frac{9}{2}\right] + \left[\frac{1}{2}\,\underline{x}_1\,, \frac{1}{2}\bar{x}_1\right] + [\epsilon^l, \quad \epsilon^u].$$

10. Conclusion

It is shown that the existing parameterized approach [1, Section 3.2, pp. 152] has considered incorrect mathematical assumptions so, the required modifications are suggested. Also, based on least absolute deviations minimization, a new approach is proposed. The silent feature of this approach is the multiplication. The illustrative example shows that the proposed interval-valued linear regression is sound and robust.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Billard, L., & Diday, E. (2000). Regression analysis for interval-valued data. In *Data Analysis, Classification, and Related Methods* (pp. 369-374). Springer, Berlin, Heidelberg.
2. De Souza, R. M., Queiroz, D. C., & Cysneiros, F. J. A. (2011). Logistic regression-based pattern classifiers for symbolic interval data. *Pattern Analysis and Applications*, *14*(3), 273.
3. Domingues, M. A., de Souza, R. M., & Cysneiros, F. J. A. (2010). A robust method for linear regression of symbolic interval data. *Pattern Recognition Letters*, *31*(13), 1991-1996.
4. Fagundes, R. A., De Souza, R. M., & Cysneiros, F. J. A. (2013). Robust regression with application to symbolic interval data. *Engineering Applications of Artificial Intelligence*, *26*(1), 564-573.
5. Fagundes, R. A., De Souza, R. M., & Cysneiros, F. J. A. (2014). Interval kernel regression. *Neurocomputing*, *128*, 371-388.
6. Giordani, P. (2015). Lasso-constrained regression analysis for interval-valued data. *Advances in Data Analysis and Classification*, *9*(1), 5-19.
7. Moore, R. E., Kearfott, R. B., & Cloud, M. J. (2009). *Introduction to interval analysis*. Society for Industrial and Applied Mathematics.
8. Neto, E. D. A. L., & de Carvalho, F. D. A. (2008). Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis*, *52*(3), 1500-1515.
9. Neto, E. D. A. L., & de Carvalho, F. D. A. (2010). Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics & Data Analysis*, *54*(2), 333-347.
10. Neto, E. D. A. L., & de Carvalho, F. D. A. (2017). Nonlinear regression applied to interval-valued data. *Pattern Analysis and Applications*, *20*(3), 809-824.

11. Neto, E. D. A. L., Cordeiro, G. M., & de Carvalho, F. D. A. (2011). Bivariate symbolic regression models for interval-valued variables. *Journal of Statistical Computation and Simulation*, *81*(11), 1727-1744.
12. Seber, G. A., & Lee, A. J. (2012). *Linear regression analysis* (Vol. 329). John Wiley & Sons.
13. Souza, L. C., Souza, R. M., Amaral, G. J., & Silva Filho, T. M. (2017). A parametrized approach for linear regression of interval data. *Knowledge-Based Systems*, *131*, 149-159.
14. Wang, H., Guan, R., & Wu, J. (2012). Linear regression of interval-valued data based on complete information in hypercubes. *Journal of Systems Science and Systems Engineering*, *21*(4), 422-442.