Commun. Korean Math. Soc. **31** (2016), No. 1, pp. 185–198
http://dx.doi.org/10.4134/CKMS.2016.31.1.185

ON THEIL'S METHOD IN FUZZY LINEAR REGRESSION MODELS

Seung Hoe Choi, Hye-Young Jung, Woo-Joo Lee, and Jin Hee Yoon

Abstract. Regression analysis is an analyzing method of regression model to explain the statistical relationship between explanatory variable and response variables. This paper propose a fuzzy regression analysis applying Theils method which is not sensitive to outliers. This method use medians of rate of increment based on randomly chosen pairs of each components of α-level sets of fuzzy data in order to estimate the coefficients of fuzzy regression model. An example and two simulation results are given to show fuzzy Theils estimator is more robust than the fuzzy least squares estimator.

One of the central objectives of mathematics is to interpret natural or social phenomena with mathematical tools including numbers, signs, and axioms. Uncertainties may occur during the process of transforming a natural or social phenomenon into a mathematical problem. These uncertainties involve two distinctive types: stochastic uncertainty whose uncertainty can be naturally resolved as time passes, and fuzzy uncertainty whose uncertainty cannot be resolved even with the passage of time. Of the two types of uncertainties, the study on stochastic uncertainty is particularly active that it is now applied in numerous fields. Zadeh introduced fuzzy theory in explaining fuzzy uncertainty with respect to ambiguity and vagueness. He further applied this theory to establish a necessary system for handling information expressed in such ambiguous or vague manners ([18], [19]).

Tanaka established fuzzy regression model as an attempt to explicate the relationship among variables that are ambiguously or vaguely presented ([15], [16]). Fuzzy regression model can be classified into two categories depending on the response function which may be known or unknown. In this case, the former and latter are called parametric model and non-parametric model, respectively. In addition, the method to estimate the fuzzy regression model can be classified into two kinds methods in terms of minimization. One is the numerical method which minimize the sum of spreads of estimated fuzzy numbers, the

Received March 19, 2015.
2010 *Mathematics Subject Classification.* Primary 62A86, 60A86.
Key words and phrases. fuzzy regression model, Theil's method, fuzzy outlier.

other is the statistical method which minimize the sum of distance between observed and estimated data. The fuzzy least squares estimation (LSE) ([3], [9]) is one of the popular statistical method and it has a weakness that is sensitive to outlier which is an observation point that is distant from other observations. Outliers are frequently observed in fuzzy regression due to ambiguity and vagueness. Therefore some alternative robust methods are needed to make up for this weak point. Theil first introduced his method in 1950, since Theil's method based on median is not sensitive to outliers. Afterward, various kinds of Theil's methods have been suggested in many studies ([10], [13], [17]), such as regression, ANOVA and financial studies. In this paper we estimate parametric fuzzy regression model through applying Theils method using median in accordance with mode and end points of the alpha-level sets. We propose an example using a data which has been frequently used in many studies in order to compare the fuzzy regression model derived from the suggested method and the least squares method. In addition we propose two simulation studies which show that fuzzy Theil's method is more robust than least squares method ([1], [4], [9], [14]) when dataset includes fuzzy outliers.

1. Fuzzy regression model

Tanaka first introduced the fuzzy regression model ([15], [16]):

$$Y(X_i) = A_0 \oplus A_1 \otimes X_{i1} \oplus \cdots \oplus A_p \otimes X_{ip}, \tag{1}$$

where X_{ij} is the j-th observation of i-th explanatory variable, A_i is the fuzzy regression coefficient, and $Y(X_i)$ is the response variable ($i = 1, \ldots, n$). And $\oplus$ and $\otimes$ which are given in (1) are the addition and multiplication of fuzzy numbers, respectively. Several fuzzy regression models have been introduced. More popular models are:

$$Y(x_i) = A_0 \oplus A_1 \otimes x_{i1} \oplus \cdots \oplus A_p \otimes x_{ip}$$

which has crisp independent variables, and

$$Y(X_i) = a_0 \oplus a_1 \otimes X_{i1} \oplus \cdots \oplus a_p \otimes X_{ip}$$

which has crisp regression coefficients. The fuzzy regression model (1) which has fuzzy coefficients and fuzzy dependent variables is not frequently mentioned due to its complexity in analyzing. In this paper, the model (1) is considered based on the operations α-level sets and Theil's method.

One of main purpose of fuzzy regression analysis is to minimize the errors between fuzzy observations and fuzzy predicted values based on fuzzy observations $\{(X_{i1}, \ldots, X_{ip}, Y_i) : i = 1, \ldots, n\}$ ([3], [5], [10]). Fuzzy least squares estimation using the distance between fuzzy observations to minimize these errors is the popular method in fuzzy regression analysis. Fuzzy least squares estimation is the popular method in fuzzy regression model. However, the least squares estimation has a weakness that is sensitive to outliers. In addition, outliers are frequently observed in fuzzy regression due to ambiguity and vagueness. Therefore some alternative robust methods are needed to make up for this weak point.

In addition, in order to estimate the variables which are expressed vaguely, it is important to determine the membership functions for the forecasted values. Hence, determining membership functions based on analysing the ambiguity and vagueness is significant in fuzzy regression model.

Zadeh suggested the extension principle to define the membership functions of a function of fuzzy numbers ([18], [19]). Furthermore, the extension principle enables us to define all operations among fuzzy numbers. If X and Y are subsets of real number, and f is a function from X to Y, $f : X \to Y$, and A is a fuzzy set on X. Then the membership function of $f(A)$ can be defined

$$\mu_{f(A)}(y) = \begin{cases} \sup_{y=f(x)} \mu_A(x) & \text{if } f^{-1}(y) \neq \phi, \\ 0 & \text{if } f^{-1}(y) = \phi, \end{cases}$$

based on extension principle.

In addition, Zadeh suggested another important principle, the resolution identity. The membership function $\mu_A(x)$ of a fuzzy set A can be derived by α-level set A_α and characteristic function $I_{A_\alpha}(x)$ based on the resolution identity theorem. If the characteristic function $I_{A_\alpha}(x)$ of α-level set A_α is defined by

$$I_{A_\alpha}(x) = \begin{cases} 1, & \text{if } x \in A_\alpha, \\ 0, & \text{if } x \notin A_\alpha, \end{cases}$$

the membership function of a fuzzy set A can be derived:

$$\mu_A(x) = \sup\{\alpha I_{A_\alpha}(x) : \alpha \in [0, 1]\}$$

using the resolution identity theorem. This explains that forecasting fuzzy data can be implemented by estimating α-level set of observed fuzzy sets.

In this paper, the forecasted values are obtained through estimating the membership functions using alpha-level sets and fuzzy regression models based on given observed fuzzy data.

2. Fuzzy Theil's method

One of main purpose of fuzzy regression analysis is to minimize the errors between fuzzy observations and fuzzy predicted values based on fuzzy observations $\{(X_{i1}, \ldots, X_{ip}, Y_i) : i = 1, \ldots, n\}$ ([3], [5], [10]). Fuzzy least squares estimation using the distance between fuzzy observations to minimize these errors is the popular method in fuzzy regression analysis. However, the least squares estimation has a weakness that is sensitive to outliers. In addition, outliers are frequently observed in fuzzy regression due to ambiguity and vagueness. Therefore some alternative robust methods are needed to make up for this weak point. Theil introduced a method which is not significantly affected by some outliers to find a best fitted line of data points in a plane ([17]). Hussain and Sprent applied Theil's method to classical regression model after they transformed the explanatory variables to orthogonal sets [6]. In this paper, we propose a fuzzy Theil's method to analyze a fuzzy regression model (1), using α-level sets

$$\{(X_{i1}(\alpha), \ldots, X_{ip}(\alpha), Y_i(\alpha)) : i = 1, \ldots, n\} \tag{2}$$

of fuzzy observations. From the resolution identity, the multiple regression model of α-level sets of left end points, $\{(l_{X_{i1}}(\alpha), \ldots, l_{X_{ip}}(\alpha), l_{Y_i}(\alpha)) : i = 1, \ldots, n\}$, can be obtained by

$$l_{Y_i}(\alpha) = \sum_{k=0}^{p} l_{X_k}(\alpha) l_{A_{ik}}(\alpha), \tag{3}$$

where $l_{X_{i0}}(\alpha) = 0$, and $l_{Y_i}(\alpha)$, $l_{X_{ik}}(\alpha)$, $l_{A_k}(\alpha)$ are the left points of Y_i, X_{ik}, and A_k respectively. After estimating the regression coefficients $A_k(\alpha)$ ($k = 0, 1, \ldots, p$) of model (3), the left endpoint $l_{Y_i}(\alpha)$ of the α-level set of the dependent variable Y_i can be estimated using following steps:

Step 1. Use Gram-Schmidts process to transform the data set $\{l_{X_1}(\alpha), \ldots, l_{X_p}(\alpha)\}$ to orthogonal data set $\{Z_1(\alpha), \ldots, Z_p(\alpha)\}$ using following formular:

$$Z_k(\alpha) = \begin{cases} l_{X_k}(\alpha) & \text{if} \quad k = 1, \\ l_{X_k}(\alpha) - \displaystyle\sum_{m=1}^{k-1} Proj_{Z_m(\alpha)}(l_{X_k}(\alpha)) & \text{if} \quad 1 < k \leq p, \end{cases} \tag{4}$$

where

$$Proj_{Z_m(\alpha)}(l_{X_k}(\alpha)) = \frac{\langle l_{X_k}(\alpha), Z_m(\alpha)\rangle}{\langle Z_m(\alpha), Z_m(\alpha)\rangle} Z_m(\alpha),$$
$$l_{X_k}(\alpha) = (l_{X_{1k}}(\alpha), \ldots, l_{X_{nk}}(\alpha))$$

and

$$Z_m(\alpha) = (Z_{1m}(\alpha), \ldots, Z_{nm}(\alpha)).$$

Step 2. Set a new regression model based on $\{Z_1(\alpha), \ldots, Z_p(\alpha)\}$ with new coefficients $\theta_k(\alpha)$.

$$l_{Y_i}(\alpha) = \theta_0(\alpha) + \theta_1(\alpha) Z_{i1}(\alpha) + \cdots + \theta_p(\alpha) Z_{ip}(\alpha). \tag{5}$$

The estimation of coefficients $\theta_k(\alpha)$ ($k = 1, \ldots, p$) can be obtained from following procedure.

2-1. Set $\theta_k^{(0)}(\alpha) = 0$ ($k = 1, \ldots, p$) and $l_{Y_i}^{(0)}(\alpha) = l_{Y_i}(\alpha)$.

2-2. Set $\delta\theta_k^{(0)}(\alpha)$ to be the median of

$$\left\{ \frac{d^{(0)}_{Y_{ij}(\alpha)}}{d_{Z_{ijk}(\alpha)}} : Z_{ik}(\alpha) < Z_{jk}(\alpha),\ 1 \leq i < j \leq n \right\}, \tag{6}$$

where

$$d^{(0)}_{Y_{ij}(\alpha)} = l^{(0)}_{Y_j}(\alpha) - l^{(0)}_{Y_i}(\alpha)$$

and

$$d_{Z_{ijk}(\alpha)} = Z_{jk}(\alpha) - Z_{ik}(\alpha).$$

2-3. Let

$$\theta_k^{(1)}(\alpha) = \theta_k^{(0)}(\alpha) + \delta\theta_k^{(0)}(\alpha)$$

and

$$l_{Y_i}^{(1)}(\alpha) = l_{Y_i}^{(0)}(\alpha) - \delta\theta_k^{(0)}(\alpha) Z_{ki}(\alpha)$$

respectively.

Generally,

$$\theta_k^{(j)}(\alpha) = \theta_k^{(j-1)}(\alpha) + \delta\theta_k^{(j-1)}(\alpha)$$

and

$$l_{Y_i}^{(j)}(\alpha) = l_{Y_i}^{(j-1)}(\alpha) - \delta\theta_k^{(j-1)}(\alpha) Z_{ki}(\alpha)$$

respectively.

2-4. Iterate above procedure until $\theta_k^{(j)}(\alpha)$ converges to a constant $h_k(\alpha)$, let the estimated coefficient $\hat{\theta}_k(\alpha)$ be

$$\hat{\theta}_k(\alpha) = h_k(\alpha), \qquad k = 1, \ldots, p.$$

That is,

$$\delta\theta_k^{(n-1)}(\alpha) = \left\{ \frac{d_{Y_{ij}}^{(0)}(\alpha)}{d_{Z_{ijk}}(\alpha)} - \sum_{s=0}^{n-2} \delta\theta_k^{(s)}(\alpha) : \right.$$

$$\left. Z_{ik}(\alpha) < Z_{jk}(\alpha), \quad i = 1, \ldots, n-1,\, j = i+1, \ldots, n \right\}, \tag{7}$$

$$l_{Y_i}^{(n)}(\alpha) = l_{y_i}^{(0)}(\alpha) - \sum_{s=0}^{n-1} \delta\theta_k^{(s)}(\alpha) Z_{ki}(\alpha), \tag{8}$$

$$\theta_k^{(n)}(\alpha) = \sum_{s=0}^{n-1} \delta\theta_k^{(s)}(\alpha). \tag{9}$$

Step 3. Use Gram-Schmidts process again to estimate $l_{A_k}(\alpha)$ based on $\hat{\theta}_k(\alpha)$ $(k = 1, \ldots, p)$

$$\hat{l}_{A_{p-k}} = \begin{cases} \hat{\theta}_p(\alpha) & \text{if } \ k = 0, \\ \hat{\theta}_{p-k}(\alpha) - \sum_{m=0}^{k-1} \frac{\langle l_{X_{p-m}}(\alpha), Z_{p-k}(\alpha) \rangle}{\langle Z_{p-k}(\alpha), Z_{p-k}(\alpha) \rangle}\, \hat{\theta}_{p-m}(\alpha) & \text{if } \ 1 \le k \le p-1. \end{cases} \tag{10}$$

The constant term of the fuzzy regression model (5) consisted of the α-level sets of the independent and dependent variables can be estimated using the mean or median of all estimated values.

And $\hat{l}_{A_0}(\alpha)$ is the median of

$$\left\{ l_{Y_i}(\alpha) - \sum_{k=1}^{p} \hat{l}_{A_k}(\alpha) l_{X_{ik}}(\alpha) : i = 1, \ldots, n \right\}$$

or

$$\hat{l}_{A_0}(\alpha) = \bar{l}_{Y_i}(\alpha) - \sum_{k=1}^{p} \hat{l}_{A_k}(\alpha)\bar{l}_{X_{ik}}(\alpha).$$

Step 4. Obtain the estimated output $\bar{l}_{Y_i}(\alpha)$ based on the model

$$\bar{l}_{Y_i}(\alpha) = \sum_{k=0}^{p} \hat{l}_{A_k}(\alpha) l_{X_{ik}}(\alpha).$$

Step 5. Obtain the intermediate estimated values $\bar{l}_{Y_i}(\alpha)$ and $\bar{r}_{Y_i}(\alpha)$ of $l_{Y_i}(\alpha)$ and $r_{Y_i}(\alpha)$ following from Step 1 to Step 4 based on the observations

$$\{(l_{X_{i1}}(\alpha), \ldots, l_{X_{ip}}(\alpha), l_{Y_i}(\alpha)) : i = 1, \ldots, n\}$$

and

$$\{(r_{X_{i1}}(\alpha), \ldots, r_{X_{ip}}(\alpha), r_{Y_i}(\alpha)) : i = 1, \ldots, n\}.$$

Step 6. Estimated values of $l_{Y_i}(\alpha)$ and $r_{Y_i}(\alpha)$ are defined as following to satisfy properties of fuzzy numbers:

$$\hat{l}_{Y_i}(\alpha) = \min\{\bar{l}_{Y_i}(\alpha),\ \hat{l}_{Y_i}(0)\}$$

and

$$\hat{r}_{Y_i}(\alpha) = \max\{\bar{r}_{Y_i}(\alpha),\ \hat{r}_{Y_i}(0)\}.$$

Step 7. Estimate the membership functions of $\hat{Y}_i$ using

$$\left\{\left(\hat{l}_{Y_i}(\alpha_k), \alpha_k\right) : k = 1, \ldots, s\right\}$$

and

$$\left\{\left(\hat{r}_{Y_i}(\alpha_k), \alpha_k\right) : k = 1, \ldots, s\right\}$$

obtained from Step 1 to Step 6. For these, we apply the fuzzy Theil's method once more to estimate the membership functions.

Especially, $l_{X_k}(\alpha)$ is equal to $Z_k(\alpha)$ in the simple fuzzy linear regression model

$$Y_i(X_i) = A_0 \oplus A_1 \otimes X_i.$$

This means the regression coefficient $\hat{\theta}_1(\alpha)$ coincides with $\hat{l}_{A_1}(\alpha)$. Therefore the estimate $\hat{l}_{A_1}(\alpha)$ of $l_{A_1}(\alpha)$ is derived from the median of

$$\left\{ \frac{l_{Y_i}(\alpha) - l_{Y_j}(\alpha)}{|l_{X_i}(\alpha) - l_{X_j}(\alpha)|} : 1 \leq i < j \leq n \right\}.$$

In addition, $\hat{l}_{A_1}(\alpha^*)$ can be derived by

$$\hat{l}_{A_1}(\alpha^*) = \min\{\hat{l}_{A_1}(\alpha),\ \bar{l}_{A_1}(\alpha^*)\} \quad \text{if} \quad \alpha^* \leq \alpha,$$

and

$$\hat{l}_{A_1}(\alpha^*) = \max\{\hat{l}_{A_1}(\alpha),\ \bar{l}_{A_1}(\alpha^*)\} \quad \text{if} \quad \alpha^* > \alpha,$$

for α^* which is not equal to previous α. Here, $\bar{l}_{A_1}(\alpha)$ is the median of

$$\left\{ \frac{l_{Y_i}(\alpha^*) - l_{Y_j}(\alpha^*)}{|l_{X_i}(\alpha^*) - l_{X_j}(\alpha^*)|} : 1 \leq i < j \leq n \right\}.$$

From the same procedure, the estimate $\hat{r}_{A_1}(\alpha)$ of the right end point $r_{A_1}(\alpha)$ for α-level set of A_1 can be derived. Therefore the estimate $\hat{A}_1$ of A_1 can be obtained by above 7 steps after deriving $\hat{l}_{A_1}(\alpha)$ and $\hat{r}_{A_1}(\alpha)$.

3. Numerical example and simulation study

In this chapter, we propose an example using simulated data. To compare the efficiencies with Least Squares Estimation (LSE), we propose a new the performance measure which is modified based on a measure introduced by Kim and Bishu ([2], [8]). Kim and Bishu used an integration of the membership functions to compare the accuracy of the developed fuzzy regression model. They considered the difference between the membership values of the observed fuzzy number Y_i and the estimated fuzzy number $\hat{Y}_i$. But, this measure of performance has a weakness when Y_i and $\hat{Y}_i$ are not overlapped. Because the value of the measure of performance will be the same regardless of between two. In order to overcome this problem, we propose a new PMFD (Performance Measure based on Fuzzy Distance), denoted by $d(Y_i, \hat{Y}_i)$, as follows:

$$\frac{\int_{-\infty}^{\infty} \left|\mu_{Y_i}(x) - \mu_{\hat{Y}_i}(x)\right| dx}{\int_{-\infty}^{\infty} \left|\mu_{Y_i}(x)\right| dx} + h_d(Y_i(0), \hat{Y}_i(0)), \tag{11}$$

where Y_i is the observed fuzzy output and $\hat{Y}_i$ is the estimated fuzzy output, and $h_d(A, B) = \inf_{b \in B} \inf_{a \in A} |a - b|$.

The smaller the distance between two fuzzy numbers is, the closer the value of (11) is to zero. Thus, we can define the measure of performance based on fuzzy distance (MPFD) for the estimated fuzzy regression model as follows:

$$M(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^{n} d(Y_i, \hat{Y}_i). \tag{12}$$

Hence, we can assume that when the value of the performance measure in (12) is the smallest, the accuracy of the method is the highest.

To compare the efficiency of the fuzzy regression model using Theil's method, we propose following example using the data which has been frequently used in many studies. Two simulations studies are proposed using crisp independent, and fuzzy independent variables respectively.

Example 1. Diamond [4] proposed following data to illustrate his approach for dealing with the problem of fuzzy input-output data.

TABLE 1. Data for Example 1.

	$X_i = (x_i, l_{x_i}, r_{x_i})_T$			$Y_i = (y_i, l_{y_i}, r_{y_i})_T$		
i	x	l_x	r_x	y	l_y	r_y
1	21.00	4.20	2.10	4.0	0.6	0.8
2	15.00	2.25	2.25	3.0	0.3	0.3
3	15.00	1.50	2.25	3.5	0.35	0.35
4	9.00	1.35	1.35	2.0	0.4	0.4
5	12.00	1.20	1.20	3.0	0.3	0.45
6	18.00	3.60	1.80	3.5	0.53	0.7
7	6.00	0.60	1.20	2.5	0.25	0.38
8	12.00	1.80	2.40	2.5	0.5	0.5

The fuzzy regression model based on the least squares estimation method is as follows:

$$\hat{Y}_i^{LS} = (1.375, 0.346, 0.060) \oplus (0.120, 0, 0.014) \otimes X_i.$$

The fuzzy regression models using Theil's method based on

$$\hat{l}_{A_0}(\alpha) = Med\{l_{Y_i}(\alpha) - \sum_{k=1}^{p} \hat{l}_{A_k}(\alpha) l_{X_{ik}}(\alpha) : i = 1, \ldots, n\}$$

and

$$\hat{l}_{A_0}(\alpha) = \bar{l}_{Y_i}(\alpha) - \sum_{k=1}^{p} \hat{l}_{A_k}(\alpha) \bar{l}_{X_{ik}}(\alpha)$$

are as follows:

$$\hat{Y}_i^{T1} = (0.5, 0, 0.4) \oplus (0.167, 0.006, 0) \otimes X_i,$$
$$\hat{Y}_i^{T2} = (0.75, 0, 0.545) \oplus (0.167, 0.006, 0) \otimes X_i,$$

$\hat{Y}_i^{T1}$ and $\hat{Y}_i^{T2}$ can be obtained until Y_i^{T1} and Y_i^{T2} get converged. Above results are obtained after 4 iterations.

Fig. 1 shows the results of Theil's methods $\hat{Y}^{T2}$ when the constant term is estimated by median.

Table 2 shows the errors of the fuzzy regression models based on the least squares estimation and Theil's method.

TABLE 2. Performance measure for Example 1.

	$M(Y, \hat{Y}^{LS})$	$M(Y, \hat{Y}^{T1})$	$M(Y, \hat{Y}^{T2})$
Total error	1.153	1.018	1.591

Following simulation studies show that proposed fuzzy Theil's method is more efficient than LSE when there is an outlier in data set. When we compare the measure of performance, it is common to find the error between the

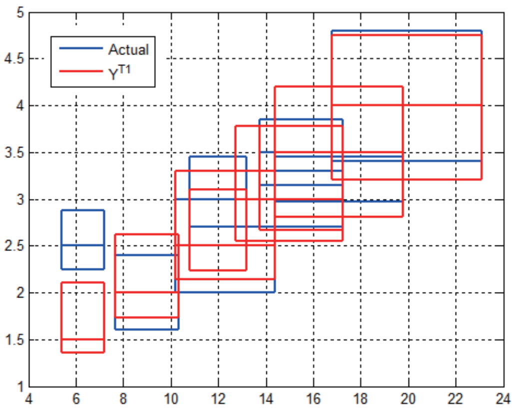

FIGURE 1. Fitted line for Example 1.

estimated value and the actual value. But we propose another comparison here that we compare the error between the estimated parameter and the true parameter, which is more basic comparison. The comparison of the estimated value and the actual value is to show how much the estimated model is close to the true model, which can be different from the selection of data. However, if we use comparison proposed in following examples, it is much more reliable because it is not dependent on the choice of data.

Example 2. 10 crisp input and fuzzy output are generated to set a fuzzy regression model as follows:

$$\begin{aligned} Y_i &= A_0 \oplus A_1 \otimes x_i \oplus E_i \\ &= (1.2, 2.1)_T \oplus (2.7, 1.3)_T \otimes x_i \oplus E_i, \quad i = 1, \dots, 10. \end{aligned}$$

The mode and spreads of E_i are generated from a log normal distribution $LN(0,1)$ and a Weibull distribution $W(1,1.5)$ respectively, which are heavy tailed distribution, We added some outliers from these heavy tail distributions. Table 3 shows the data obtained using lognormal distribution and Weibull distribution.

Based on proposed fuzzy regression model using Theil's method, the coefficients were converged after 4 times of iterations, which result in following models:

(a) $\hat{Y}_i^{LS} = \tilde{A}_0 \oplus \tilde{A}_1 \otimes x_i = (1.717, 3.997)_T \oplus (3.068, 1.239)_T \otimes x_i.$

(b) $\hat{Y}_i^{T1} = \hat{A}_0^1 \oplus \hat{A}_1^1 \otimes x_i = (1.727, 1.695)_T \oplus (2.876, 1.863)_T \otimes x_i.$

(c) $\hat{Y}_i^{T2} = \hat{A}_0^2 \oplus \hat{A}_1^2 \otimes x_i = (1.634, 2.344)_T \oplus (2.876, 1.863)_T \otimes x_i.$

Fig. 2 shows the data in Example 2 and the results $\hat{Y}^{T1}$ and $\hat{Y}^{T2}$ of Theil's method when the constant terms are estimated by median and mean.

TABLE 3. Data for Example 2.

x_i	$Y_i = (y_i, l_{y_i}, r_{y_i})_T$		
x	y_i	l_{y_i}	r_{y_i}
1	4.7	3.41	3.41
2	7.12	7.04	7.04
3	10.97	6.03	6.03
4	13.23	10.36	10.36
5	15.39	13.96	13.96
6	20.03	11.44	11.44
7	26.79	12.17	12.17
8	31.68	12.64	12.64
9	27.25	15.62	15.62
10	28.75	15.46	15.46

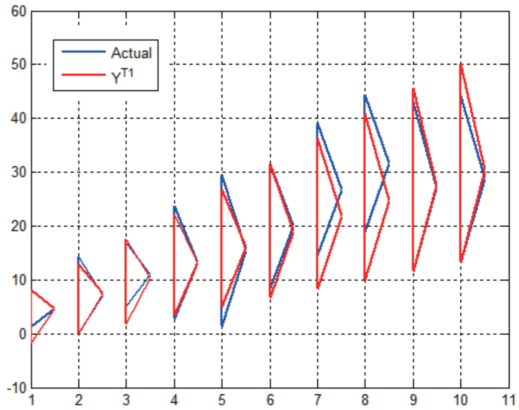

FIGURE 2. Fitted line for Example 2.

Here we obtained the errors PMFD between true coefficients and estimated coefficients as follows:

TABLE 4. PMFD for Example 2

k	$d(A_k, \tilde{A}_k)$	$d(A_k, \hat{A}_k^1)$	$d(A_k, \hat{A}_k^2)$
0	0.978	0.943	0.462
1	1.205	1.01	1.01

Example 3. In this example we consider the following model

$$Y_i = A_0 \oplus A_1 \otimes X_i \oplus E_i$$

$$= (5.2, 3)_T \oplus (2, 1.7)_T \otimes X_i \oplus E_i, \quad i = 1, \ldots, 10.$$

For the observations $X_i = (x_i, s_{x_i})$, we generate the mode x_i from uniform distribution and the spread s_{x_i} from triangular distribution, respectively. For error term $E_i = (m_i, e_i)_T$, the modes m_k are generated from Weibull distribution and the spreads e_i are generated from t-distribution with degree of freedom 1 ($i = 1, \ldots, 10$).

Weibull distributions and t-distribution are heavy tailed distribution, so that E_i includes outliers. Table 5 is the data for the fuzzy regression model obtained by using a Weibull distribution and t-distribution.

TABLE 5. Data for Example 3.

	$X_i = (x_i, s_{x_i})_T$		$Y_i = (y_i, s_{y_i})$	
i	x	s_x	y	s_y
1	2	1.6	11	6.97
2	8	2.1	22.04	10.78
3	6	4.0	17.95	11.31
4	3	2.4	11.25	8.23
5	11	3.4	27.47	15.68
6	7	4.3	19.38	13.83
7	8	3.5	22.04	16.89
8	9	2.0	23.37	12.07
9	5	3.7	15.51	9.97
10	13	2.8	31.47	8.1

The estimated coefficients $\tilde{A}_k$, and $\hat{A}_k^1$ and $\hat{A}_k^2$ ($k = 0, 1$) using LSE and proposed fuzzy Theil's method are followings:

(a) $\hat{Y}_i^{LS} = \tilde{A}_0 \oplus \tilde{A}_1 \otimes X_i = (6.231, 3.126)_T \oplus (1.933, 0.259)_T \otimes X_i$.

(b) $\hat{Y}_i^{T1} = \hat{A}_0^1 \oplus \hat{A}_1^1 \otimes X_i = (5.539, 1.533)_T \oplus (1.994, 0.346)_T \otimes X_i$.

(c) $\hat{Y}_i^{T2} = \hat{A}_0^2 \oplus \hat{A}_1^2 \otimes X_i = (5.680, 2.399)_T \oplus (1.994, 0.346)_T \otimes X_i$.

Fig. 3 shows the data obtained from Weibull distribution and t-distribution and the result $\hat{Y}^{T1}$ of Theil's method when the constant term is estimated by median.

To compare the efficiencies of LSE and proposed fuzzy Theil's method, we use the performance measure (12). Table 5 shows proposed method is robust, that is not sensitive to outliers even if the dataset includes fuzzy outliers.

4. Conclusions

In this paper, we propose fuzzy Theil's method to obtain a robust fuzzy regression model and a new performance measure, PMFD (Performance measure based on fuzzy distance) to compare the efficiencies which consider distance

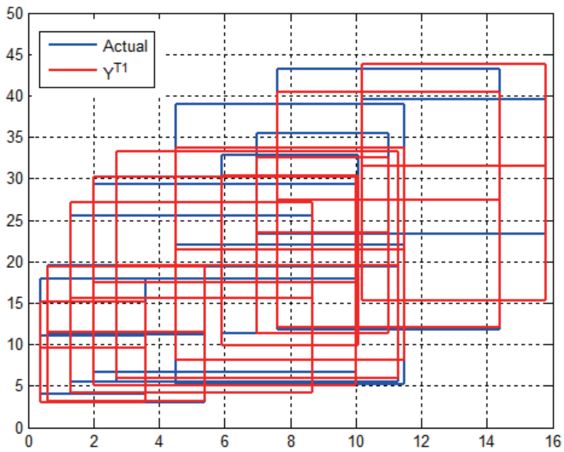

FIGURE 3. Fitting lines for Example 3

TABLE 6. Errors of estimated fuzzy regression coefficients

k	$d(A_k, \tilde{A}_k)$	$d(A_k, \hat{A}_k^1)$	$d(A_k, \hat{A}_k^2)$
0	0.648	0.321	0.528
1	0.825	0.575	1.575

when there is no overlapped area between two fuzzy numbers. In order to find the fuzzy regression model, we take the median of estimated value from arbitrary pairs of observations in terms of α-level sets, based on the procedures provided in the paper. An example using a data which has been frequently used in many studies is given to compare the fuzzy regression model derived from the suggested method and the least squares method. Two simulation studies shows that proposed fuzzy Theil's method is robust when dataset includes fuzzy outliers.

References

[1] L. H. Chen and C. C. Hsueh, *Fuzzy regression models using the least-squares method based on the concept of distance*, IEEE Transactions on Fuzzy Systems **17** (2009), 1259–1272.

[2] S. H. Choi and J. J. Buckley, *Fuzzy regression using least absolute deviation estimators*, Soft Computing **12** (2008), 257–263.

[3] S. H. Choi and J. H. Yoon, *General fuzzy regression using least squares method*, Int. J. Sys. Sci. **41** (2010), 477–485.

[4] P. Diamond, *Fuzzy least squares*, Inform. Sci. **46** (1988), no. 3, 141–157.

[5] P. Diamond and H. Tanaka, *Fuzzy regression analysis*, Fuzzy sets in decision analysis, operations research and statistics, 349–387, Handb. Fuzzy Sets Ser., 1, Kluwer Acad. Publ., Boston, MA, 1998.

[6] S. S. Hussain and P. Sprent, *Nonparametric regression*, J. Roy. Statist. Soc. **146** (1983), no. 2, 182–191.
[7] C. Kao and C. Chyu, *Least-squares estimates in fuzzy regression analysis*, European J. Oper. Ress **148** (2003), no. 2, 426–435.
[8] B. Kim and R. R. Bishu, *Evaluation of fuzzy linear regression models by comparing membership functions*, Fuzzy Sets and Systems **100** (1998), 343–352.
[9] H. K. Kim, J. H. Yoon, and Y. Li, *Asymptotic properties of least squares estimation with fuzzy observations*, Inform. Sci. **178** (2008), 439–451.
[10] J. H. Yoon and S. H. Choi, *Componentwise fuzzy linear regression using least squares estimation*, J. Muli.-Valued Logic **15** (2009), 137–153.
[11] ______, *Fuzzy least squares estimation with new fuzzy operations*, Advances in Intelligent Systems and Computing **190** (2013), 193–202.
[12] J. H. Yoon, S. H. Choi, and H. K. Kim, *Asymptotic consistency of least squares estimators in fuzzy regression model*, Commun. Korean Statis. Soc. **11** (2008), 799–813.
[13] J. H. Yoon, W. J. Lee, and S. H. Choi, *Fuzzy Theil regression model*, J. Korean Instit. Intelligent Systems **23** (2013), 336–340.
[14] D. Savic and W. Pedryzc, *Evaluation of fuzzy linear regression models*, Fuzzy Sets and Systems **39** (1991), no. 1, 51–63.
[15] H. Tanaka, I. Hayashi, and J. Watada, *Possibilistic linear regression analysis for fuzzy data*, European J. Oper. Res. **40** (1989), no. 3, 389–396.
[16] H. Tanaka, S. Uejima, and K. Asai, *Linear regression analysis with fuzzy model*, IEEE Trans. Syst. Man Cybernet. **12** (1982), 903–907.
[17] H. Theil, *A rank invariant method of linear and polynomial regression analysis. I*, Pro. Kon. Ned. Akad. Wetensch. A **53** (1950), 386–392.
[18] L. A. Zadeh, *Fuzzy sets*, Information and Control **8** (1965), 338–353.
[19] ______, *The concept of linguistic variable and its application to approximate reasoning I*, Inform. Sci. **8** (1975), 199–249.

Seung Hoe Choi
School of Liberal Arts and Science,
Korea Aerospace University
Koyang 411, Korea
E-mail address: shchoi@kau.ac.kr

Hye-Young Jung
Department of Statistics
Seoul National University
Seoul 151-742, Korea
E-mail address: hyjunglove@snu.ac.kr

Woo-Joo Lee
Department of Mathematics
Yonsei University
Seoul 120-749, Korea
E-mail address: bobspace@yonsei.ac.kr

Jin Hee Yoon
School of Mathematics and Statistics
Sejong University
Seoul 143-749, Korea
E-mail address: jin9135@yonsei.ac.kr