Iranian Journal of Fuzzy Systems

Volume 21, Number 3, (2024), pp. 177-192

Original paper

University of Sistan and Baluchestan

Theil-Sen estimators for fuzzy regression model

Z. Behdani [1] and M. Darehmiraki [2]

[1,2] *Department of Mathematics, Behbahan Khatam Alanbia University of Technology, Khouzestan, Iran*

behdani@bkatu.ac.ir, darehmiraki@bkatu.ac.ir

Abstract

Both traditional and fuzzy regression analyses have demonstrated the significant characteristics of the least-squares methodology as a method for parameter estimation. The presence of outliers in the sample and/or minor variations in the dataset might impact the behaviour and characteristics of the least-squares estimators (LSE). In contrast, robust approaches provide estimators of the parameters that are resilient to the aforementioned unfavourable effects. This study aims to expand upon the Theil-Sen estimator in fuzzy regression analysis, with the objective of obtaining consistent findings even when outliers are present. We demonstrate the effectiveness of the suggested technique through simulation experiments and real-world examples, comparing it with commonly used fuzzy regression models. The applicative examples are based on hydrology and atmospheric environment datasets. We also show the sensitivity analysis of the estimated parameters using a Monte-Carlo simulation study, demonstrating the effectiveness of the suggested estimators in comparison to other established approaches in the field of fuzzy regression analysis. The results showed that the Theil-Sen estimator (TSE) is very effective in cases there are outliers, and the calculation error is smaller compared to other methods.

Keywords: Fuzzy outlier, regression model, Theil-Sen estimator, distance.

1 Introduction

Today, the role of regression models in data analysis, especially finding the relationship between dependent and independent data, is no longer a secret. Numerous research studies have investigated the applications of this model in various fields, such as the social sciences, engineering sciences, medicine, and so on [2, 27, 28, 32, 44, 45, 38]. Since we frequently encounter imprecise data in real applications, it is necessary to examine these models in a fuzzy environment that can display such data. The main goal of this article is to provide a fuzzy regression parameter estimation method that reduces the impact of outlier data on least squares estimators.

Fuzzy regression analysis can establish an acceptable relationship between a dependent variable and many independent variables. According to Chukhrova and Johannssen's comprehensive study [15], fuzzy regression analysis has witnessed significant development and has become the most important subject of research within the realm of fuzzy statistics. There are two distinct types of fuzzy regression models. Tanaka et al. [4] presented the possibilistic notion, which has since been expanded upon by various researchers to form the basis of the first group. In this method, regression parameters are estimated using linear programming techniques, which might be tricky throughout the estimation process. Increasing the number of data sets results in a corresponding increase in the restrictions imposed by the linear programming approach, as shown by Chang and Ayyub [10]. Because of this growth, there may be computational issues. The second kind utilizes the least squares method, which involves minimizing many distance measurements between fuzzy numbers to provide an optimal solution for the discrepancy between the estimated and observed fuzzy outputs. Therefore, this method incorporates the goodness-of-fit concept, allowing the produced residuals to probe the model's precision. This is the main benefit of using this method.

The least-squares approach is a frequently used technique for parameter estimation in the fuzzy regression model.

Corresponding Author: M. Darehmiraki

Received: April 2024; Revised: June 2024; Accepted: August 2024.

https://doi.org/10.22111/ijfs.2024.48443.8535

Unfortunately, this method is very susceptible to dataset outliers. When seeking a reliable estimating strategy, it is advantageous to use established, robust ways for identifying potentially important points and estimation methods [25, 37]. These techniques are significant because of the potential impact of a single observation on parameter estimations. Excluding such an observation from the dataset might result in substantially different estimates, emphasizing the importance of their inclusion [6]. Peters [36] examined the presence of outliers within Tanaka's possibilistic method [4] using crisp input-output data. This model was subsequently expanded by Chen [12] to include fuzzy output-crisp input data. In their study, the authors of [26] introduced an omission technique as an alternative to Tanaka's approach [4]. This approach enables the assessment of each observation's impact on the model's objective function upon its removal. In their research, Nasrabadi et al. [35] introduced a linear programming (LP) method for identifying outliers in the context of fuzzy regression analysis. In their study, Varga [40] introduced robust estimation methods for both fuzzy and non-fuzzy regression models. The authors of [34] proposed a resilient nonlinear fuzzy regression model that utilizes multi-layered feed forward neural networks. In their research work, the authors of [29] introduced a robust method for fuzzy regression analysis that utilizes the ranking of fuzzy sets. The least median squares-weighted least-squares estimation process is the basis of the resilient fuzzy linear regression model presented by [19]. The authors of [18] introduced the concepts of weighted least-squares and least median squares estimates as methods for conducting fuzzy linear regression analysis. The study by Ferraro and Giordani [23] focused on the concept of robustness within the domain of regression analysis, specifically in relation to the handling of ambiguous information via fuzzy sets. In their study, Chachi and Roozbeh [7] introduced a novel methodology for analysing bed-load transport data using a fuzzy robust regression technique. In their study, Chachi, Taheri, Fattahi, and colleagues [9] introduced and examined two robust fuzzy regression models for the purpose of predicting flaws in cotton yarn. A resilient fuzzy c-regression model was proposed by Leski and Kotas [30]. It includes an objective function that is based on Huber's M-estimators and Yager's ordered weighted averaging (OWA) operators. Chachi [6] introduced a weighted goal function to alleviate the drawbacks of the LS fuzzy regression algorithms in the presence of outliers. Using fuzzy outputs and fuzzy parameters, Arefi [3] explored a quantile fuzzy regression. An incomplete robust ridge regression model with fuzzy predictors and responses was studied by Akbari and Hesamian [1]. A Hesamian and Akbari ([24]) proposed a resilient variable coefficient method for the fuzzy multiple regression model. A gradient-boosting ensemble system using a fuzzy regression tree as its basis component for regression problems was suggested in [33]. The writers of [11] came up with a new fuzzy logistic regression method that does a better job of classifying things when there is imbalance and separation while still being easy to understand like traditional logistic regression. The robust intuitionistic fuzzy regression function methods proposed in [20] are immune to data set outliers. Instead of utilizing the standard least squares technique, the suggested methodology uses functions based on robust regression to estimate the parameters: Welsch, Bisquare, Talwar, Huber, Logistic, and Cauchy.

The least-squares technique is the go-to statistical approach for estimating the parameters of interest, and it holds true over a wide range of model complexity. The least-squares estimate uses criteria that are easy to understand and very intuitive. We evaluate the model's parameters using the least squares criteria, which minimizes the sum of squared differences between the predicted and observed points, to achieve the best possible agreement between model predictions and observed data [11]. In the simplest linear model, we assume that the real mean of the dependent variable W fluctuates at a constant rate when the value of the independent variable U increases or decreases. The Theil-Sen estimator's biggest strength is its resilience. Many writers have noted that the regular LSE, in contrast to the Theil-Sen estimator, is substantially less susceptible to outliers (for examples, see [13, 22] and the references therein). One can rationalize this result as follows: The larger a positive outlier is, the more it will raise the sample mean. In fact, there is no maximum allowable outlier impact on the sample mean. However, once an outlier becomes the biggest observation in the sample, it has no further impact on the median [41]. Examples of robust and non-robust estimators include the sample median and mean, respectively [14]. Recently, papers [8] and [5] examined robust fuzzy regression based on the M estimator.

The Theil-Sen estimator has several advantages over the least squares estimator. First, this method is not sensitive to outlier data, unlike the high sensitivity of the least squares estimator. Furthermore, even when the residuals lack a normal distribution, the Theil-Sen estimator remains reliable. This method is also effective when the data is skewed and skewed and heteroskedastic, or, in other words, lumpy. We also use this method when the variance is not constant. It has been called "the most popular non-parametric technique for estimating a linear trend" [21]. In order to compute the parameters in an effective manner, there are techniques that are quick. In general, the Theil-Sen estimator is competitive with alternative slope estimators in terms of MSE, simple to compute, and robust. In light of the above explanation, this study presents a robust estimate strategy for the fuzzy regression model by extending the Theil-Sen estimators to a fuzzy environment. The suggested technique involves estimating parameter values in the presence of outliers, taking into account the impact of these outliers on the calculated parameters. To our knowledge, this is the first time we have investigated the Theil-Sen estimator in a fuzzy environment. We then apply Theil-Sen estimators to

two simulated and real data examples and conclude by comparing the newly introduced method with some previously used methods.
We structure the remaining parts of the article as described below. In Section 2, we use the Theil-Sen estimators to develop a robust fuzzy multiple linear regression model for crisp input-fuzzy output data. We use the Theil-Sen estimators to fit the model to the data. In order to demonstrate the accuracy of the proposed model when there are outliers in the data, Section 3 presents the results of two simulated studies and two cases of real data with outliers. Section 4 concludes the paper with some final notes.

2 Fuzzy Theil-Sen estimator

The Theil (1950) [39] regression method is an additional non-parametric technique for doing linear regression analysis with a single predictor variable and a single responder variable. The Theil estimator is a reliable method for estimating linear regression models and identifying outliers in the response variable. When the estimator takes the form of a linear equation, the Ordinary Least Squares (OLS) estimate is equivalent to the mean and is not considered robust. A single data point has the potential to significantly influence the gradient of the linear relationship. The Theil-Sen estimator is a statistical method that provides a robust alternative to linear regression. The algorithm calculates the slopes of all the lines connecting each pair of points, then determines the median of these slopes. Sometimes, people refer to this operation as the Theil-Sen procedure.
We examine the fuzzy regression model to elucidate the functional correlation between variables with imperfect information, as formulated as follows:

$$W_i(U_i) = C_0 + C_1 u_{i1} + \cdots + C_p u_{ip}, \quad i = 1, \cdots, n, \tag{1}$$

where W_i and C_j, $i = 1, \cdots, n$, $j = 1, 2, \cdots, p$ are LR fuzzy numbers and u_{ij}, $i = 1, \cdots, n$, $j = 1, \cdots, p$ are crisp numbers. The main goal of fuzzy regression analysis is to find the regression coefficients that reduce the difference between the observed fuzzy numbers and the expected fuzzy numbers that can be derived from the data $\{(U_{i1}, \cdots, U_{ip}, W_i) : i = 1, \cdots, n\}$. A LR fuzzy number $(W = (w, l_w, r_w))$ has the following membership function [31].

$$W(x) = \begin{cases} L(\frac{w-x}{l_w}), & x \leq w, \\ R(\frac{x-w}{r_w}), & x > w, \end{cases}$$

The functions L and R are monotonically decreasing, with the initial conditions $L(0) = R(0) = 1$ and $L(1) = R(1) = 0$. In this context, the symbol w represents the mode of the fuzzy number A, whereas l_w and r_w represent the left and right spreads of the same fuzzy number W, respectively. When both functions $L(x)$ and $R(x)$ equal $1 - x$, we classify W as a triangular fuzzy number, which we can express as $(w, l_w, r_w)_T$. The LR-fuzzy number is denoted as $(w, s_w)_{LR}$, specifically when the fuzzy number exhibits symmetry, meaning that the left and right spreads are equal. The α-level set of the fuzzy number W may be defined as the closed interval with a mode of w, a left spread denoted as $(l_w L^{-1}(\alpha))$, and a right spread denoted as $(r_w R^{-1}(\alpha))$. As a result, we can represent the α-level set of the fuzzy number W as follows:

$$W(\alpha) = [w - l_w L^{-1}(\alpha), w + r_w R^{-1}(\alpha)].$$

2.1 Theil-Sen estimator

Fuzzy least-squares algorithms are susceptible to the impact of outliers, so the presence of outliers might skew the results. Hence, robust techniques have been developed to adapt least-squares approaches in order to minimize the impact of outliers on the final results. The following section will use the Theil-Sen estimator as the central concept, presenting a modified least-squares estimation in fuzzy regression. Theil-Sen estimator offers several benefits compared to ordinary least squares regression. It lacks sensitivity toward extreme values. Robust simple linear regression (least squares) is notably more precise than non-robust methods when dealing with skewed and heteroskedastic data. Even with properly distributed data, it outperforms least squares in terms of statistical power. People widely recognize it as the most common non-parametric method for predicting a linear trend. There are rapid methods for effectively calculating the parameters.
Let's assume that the documentation for n units looks like this.

$$(\tilde{w}_1, u_{11}, ..., u_{1p}), ..., (\tilde{w}_n, u_{n1}, ..., u_{np}), \tag{2}$$

where triangular fuzzy response variable as $\tilde{w}_i = (w_i, l_{w_i}, r_{w_i})_{LR}\,(i = 1, ..., n)$ and constitutes the crisp explanatory variables. If $C_j = (\beta_j, \sigma_j, \theta_j)$ is set, then equation (1) will be as follows [31]

$$(w_i, l_{w_i}, r_{w_i}) = (\beta_0 + \sum_{j=1}^{p} u_{ij}\beta_j, \sigma_0 + \sum_{j=1}^{p} u_{ij}\sigma_j, \theta_0 + \sum_{j=1}^{p} u_{ij}\theta_j), \quad i = 1, \cdots, n. \tag{3}$$

The optimization problem in this method is similar to the least squares method. In other words, here we are looking to minimize the sum of squared errors for different permutations of the sample. Here we perform the least squares method for each permutation of the sample and finally introduce the median of all the obtained estimates as the final estimate. That is, we will have a combination of resampling and least square methods. In the following, we explain the method in full. First, we will review the TSM in its specific mode. In this study, we will examine the TSM and begin by implementing a basic linear regression using a single predictor variable, denoted as $p = 1$. In the context of geometric estimation of the slope $C_1 = (\beta_1, \sigma_1, \theta_1)$, it is enough to look at two separate points (U_r, W_r) and (U_s, W_s), where U_r is not the same as U_s, to get an estimate of the slope *beta*, which is written as $C_{1_{rs}} = (W_r - W_s)/(U_r - U_s)$. A robust estimator $\hat{C}_1$ for the slope C_1 is obtained by taking the median of the least squares estimations [17].

$$\hat{C}_1 = (\beta_1, \sigma_1, \theta_1) = Med\{(W_r - W_s)/(U_r - U_s) : U_r \neq U_s, 1 \leq r < s \leq n\}. \tag{4}$$

Also, $C_{0_{rs}} = W_r - C_{1_{rs}}U_r$. The symbol $Med\{C_j : s \in \mathbb{S}\}$ represents the median of the set of numbers $\{C_s : s \in \mathbb{S}\}$. Let's examine a multiple linear regression with $p \geq 1$. To find an estimate for $C = (C_0, C^T)^T$, the technique outlined above involves solving a set of $p + 1$ equations.

$$W_i - C_0 - U_i^T C = 0, \quad i \in k_{p+1} = \{i_1, ..., i_{p+1}\}, \tag{5}$$

the $(p+1) \times (p+1)$ matrix $(U_k : k \in k_{p+1})$ is invertible, and k_{p+1} is a $(p+1)$ subset of $\{1, ..., n\}$. In order to emphasize the reliance on the $p + 1$ data, we represent this estimate as $\hat{C}_{k_{p+1}}$ [43].

$$\hat{C}_n = Mmed\{\hat{C}_{k_{p+1}} : \forall k_{p+1}\}.$$

It is important to observe that the value $\hat{C}_{k_{p+1}}$ is also the estimate that minimizes the sum of squared errors for C, using $p + 1$ observations $\{(U_i, w_i) : i \in k_{p+1}\}$.

Algorithm 1: Thiel Sen method for simple linear regression

Input : initial point $(u_1, \tilde{w}_1), \cdots, (u_n, \tilde{w}_n), \quad u_i \in R\, and \tilde{w}_i = (w_i, l_{w_i}, r_{w_i})_T$.

1 Initialize : $i = 1, 2, \cdots, n$.

2 **For**: $i = 1, 2, \cdots (n-1)$ **do For**: $j = i + 1, 2, \cdots n$ **do**

1. $\beta_{ij} = (w_i - w_j)/(u_i - u_j)$, $\sigma_{ij} = (l_{w_i} - l_{w_j})/(u_i - u_j)$ and $\theta_{ij} = (r_{w_i} - r_{w_j})/(u_i - u_j)$.
2. $\hat{\beta} = Median\{\beta_{ij} = (w_i - w_j)/(u_i - u_j) | \, u_i \neq u_j\}$.
3. $\hat{\sigma} = Median\{\sigma_{ij} = (l_{w_i} - l_{w_j})/(u_i - u_j) | \, u_i \neq u_j\}$.
4. $\hat{\theta} = Median\{\theta_{ij} = (r_{w_i} - r_{w_j})/(u_i - u_j) | \, u_i \neq u_j\}$.
5. Choose $\tilde{C}_1 = (\hat{\beta}, \hat{\sigma}, \hat{\theta})_T$.

For: $i = 1, 2, \cdots n$ **do**

1. $\beta_{0i} = w_i - \hat{\beta} * u_i$, $\sigma_{0i} = l_{w_i} - \hat{\sigma} * u_i$ and $\theta_{0i} = r_{w_i} - \hat{\theta} * u_i$
2. $\hat{\beta}_0 = Median\{\beta_{0i} | i = 1, \cdots, n\}$, $\hat{\sigma}_0 = Median\{\sigma_{0i} | i = 1, \cdots, n\}$ and $\hat{\theta}_0 = Median\{\theta_{0i} | i = 1, \cdots, n\}$
3. Set $\tilde{C}_0 = (\hat{\beta}_0, \hat{\sigma}_0, \hat{\theta}_0)_T$.

Algorithm 2: Thiel Sen method for multiple linear regression

Input : initial point $(u_{11}, u_{12}, \cdots, u_{1p}, \tilde{w}_1), \cdots, (u_{n1}, u_{n2}, \cdots, u_{np}, \tilde{w}_n), \quad u_{ij} \in R and \tilde{w}_i = (w_i, l_{w_i}, r_{w_i})_T$.

1 Initialize : $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, p$.

2 Set k equal to all possible permutations of choosing $p+1$ from n samples. **for**: $i = 1, 2, \cdots (k)$ **do**

1. Set X be a matrix with dimensions of $p+1$ in $p+1$
2. Set the first column of the X matrix equal to one and set the other columns as the first to pth independent variables corresponding to the ith permutation.
3. Set $W' = (w_{i1}, \cdots, w_{i(p+1)})$
4. Set $L' = (l_{w_{i1}}, \cdots, l_{w_{i(p+1)}})$
5. Set $R' = (r_{w_{i1}}, \cdots, r_{w_{i(p+1)}})$
6. $\beta_i = (X'X)^{-1}X'W$
7. $\sigma_i = (X'X)^{-1}X'L$
8. $\theta_i = (X'X)^{-1}X'R$

1. $\hat{\beta} = MultivariateMedian\{\beta_i,\ i = 1, 2, \cdots, k\}$
2. $\hat{\sigma} = MultivariateMedian\{\sigma_i,\ i = 1, 2, \cdots, k\}$.
3. $\hat{\theta} = MultivariateMedian\{\theta_i,\ i = 1, 2, \cdots, k\}$.

Set $\tilde{C} = (\hat{\beta}, \hat{\sigma}, \hat{\theta})_T$.

In this paper,we compare and evaluate the fitness of the competing models using three goodness-of-fit metrics: the mean of similarity measures (S) and the mean of absolute errors (E1 and E2)([6, 7, 8, 19].

$$\begin{aligned} S =& \frac{1}{n}\sum_{i=1}^{n}\frac{\int min\{\hat{w}_i(t), \tilde{\hat{w}}_i(t)\}dt}{\int max\{\hat{w}_i(t), \tilde{\hat{w}}_i(t)\}dt}, \\ E_1 =& \frac{1}{n}\sum_{i=1}^{n}\int |\hat{w}_i(t), \tilde{\hat{w}}_i(t)|dt, \\ E_2 =& \frac{1}{n}\sum_{i=1}^{n}\frac{\int |\hat{w}_i(t), \tilde{\hat{w}}_i(t)|dt}{\int \hat{w}_i(t)dt}. \end{aligned} \tag{6}$$

According to the definition of these criteria, it is clear that the closer S is to one, the greater the similarity between the fitted values and the real values, and the smaller the other two criteria, which are actually indicators for measuring error, the higher the accuracy.

3 Practical examples

In this section, we demonstrate the application of the article's method using real and simulated data. Furthermore, the obtained results, as well as their comparison with other proposed methods, will demonstrate the efficiency and superiority of this method over other methods.

Computation and simulation study

In this subsection, we describe the random sampling of a sub-population to calculate the estimator for a large sample size. A simulation is also conducted.

Example 3.1. *To investigate the behaviour of the proposed multiple Theil-Sen estimator (MTSE), three simulations are carried out for robustness. Samples are generated from the multiple regression model* $w_i = 1 + 5u_{1i} + 10u_{2i} + \epsilon_i$

where $u_{1i} \sim N(0,1)$, $u_{2i} \sim U(0,1)$ *also* $l_i = 0.2 + 0.8u_{1i} + 1.2u_{2i} + \epsilon_i'$, $r_i = 0.8 + 1.4u_{1i} + 1.8u_{2i} + \epsilon_i''$ *and the error* $\epsilon \overset{d}{=} \epsilon' \overset{d}{=} \epsilon'' \sim N(0, 0.5)$*'s are from different distributions for different purposes. In this example, we generate data with different volumes using the simulation method and estimate the parameters of the model using two methods: Theil-Sen and least squares. In each case, we select a number of data points from a different distribution to stand out from the rest, a process known as inserting outliers into the data. The results are recorded in Table 1. In the absence of outlier data, the results of the two methods are nearly identical. While there is a significant difference between these two methods for outlier data, As anticipated, outliers strongly influence the least squares method, causing it to deviate towards these data points. While the Theil-Sen method is stable and does not change much under the influence of outlier data, Particularly when the outlier data is small or increases, the estimates derived from this method will undergo minimal changes. Furthermore, as shown in the table (the first three lines for each method), the estimated value of the parameters becomes closer to the real value as the number of samples increases.So that for the sample size of 40, the values of parameters* $\tilde{C}_0 = (1.008, 0.21, 0.81)$, $\tilde{C}_1 = (5.14, 0.94, 1.54)$, $\tilde{C}_2 = (9.97, 1.17, 1.77)$ *, and* $\tilde{C}_0 = (1.026, 0.23, 0.83)$, $\tilde{C}_1 = (5.14, 0.94, 1.54)$, $\tilde{C}_2 = (9.94, 1.14, 1.74)$ *are estimated for the Theil-Sen method and the least squares, respectively, and in addition to being close to the real value, these two methods have performed almost the same. In a sample of 20 and with six outliers, the estimation of the model's parameters using the Theil-Sen method is equal to* $\tilde{C}_0 = (1.24, 0.34, 0.94)$, $\tilde{C}_1 = (4.65, 0.82, 1.43)$, $\tilde{C}_2 = (9.15, 0.9, 1.5)$*, whereas the least squares method provided values of* $\tilde{C}_0 = (0.45, 0.22, 0.82)$, $\tilde{C}_1 = (2.47, 0.82, 1.42)$, $\tilde{C}_2 = (7, 1.04, 1.64)$*. These values demonstrate the difference in accuracy between the two methods in the presence of outlier data, as well as the high efficiency of the Theil-Sen method in this case.*

Table 1: Estimation of parameters by ordinary least squares and Theil-Sen estimator for with outlier data from model $w_i = 1 + 5u_{1i} + 10u_{2i} + \epsilon_i$.

Sample	Theil-Sen		
size	$(\beta_0, \beta_1, \beta_2)$	$(\sigma_0, \sigma_1, \sigma_2)$	$(\theta_0, \theta_1, \theta_2)$
n=20	$(0.89, 4.81, 9.93)$	$(0.091, 0.61, 1.126)$	$(0.69, 1.21, 1.73)$
n=30	$(0.84, 5.15, 10.67)$	$(0.045, 0.95, 1.87)$	$(0.64, 1.55, 2.47)$
n=40	$(1.008, 5.14, 9.97)$	$(0.21, 0.94, 1.17)$	$(0.81, 1.54, 1.77)$
$n_1 = 16,\ n_2 = 4$	$(1.24, 4.65, 9.15)$	$(0.34, 0.82, 0.90)$	$(0.94, 1.43, 1.5)$
$n_1 = 15,\ n_2 = 5$	$(1.04, 4.71, 9.30)$	$(0.27, 0.58, 0.64)$	$(0.87, 1.18, 1.24)$
$n_1 = 14,\ n_2 = 6$	$(1.24, 4.65, 9.15)$	$(0.34, 0.82, 0.90)$	$(0.94, 1.43, 1.5)$
$n_1 = 13,\ n_2 = 7$	$(-3.1, 4.31, 13.39)$	$(-0.09, 0.9, 1.56)$	$(0.51, 1.50, 2.16)$
Sample	LSE		
size	$(\beta_0, \beta_1, \beta_2)$	$(\sigma_0, \sigma_1, \sigma_2)$	$(\theta_0, \theta_1, \theta_2)$
n=20	$(0.89, 4.81, 9.92)$	$(0.092, 0.61, 1.12)$	$(0.69, 1.21, 1.72)$
n=30	$(0.93, 5.13, 10.45)$	$(0.13, 0.93, 1.65)$	$(0.73, 1.53, 2.25)$
n=40	$(1.026, 5.14, 9.94)$	$(0.23, 0.94, 1.14)$	$(0.83, 1.54, 1.74)$
$n_1 = 16,\ n_2 = 4$	$(0.45, 2.47, 7)$	$(0.22, 0.82, 1.04)$	$(0.82, 1.42, 1.64)$
$n_1 = 15,\ n_2 = 5$	$(-1.19, 3.79, 11.5)$	$(0.32, 0.59, 0.71)$	$(0.92, 1.19, 1.31)$
$n_1 = 14,\ n_2 = 6$	$(0.45, 2.47, 7)$	$(0.22, 0.82, 1.04)$	$(0.82, 1.42, 1.64)$
$n_1 = 13,\ n_2 = 7$	$(-3.23, 3.64, 13.12)$	$(-0.09, 0.9, 1.57)$	$(0.51, 1.50, 2.17)$

Example 3.2. *This example presents a Monte-Carlo simulation analysis to validate the performance of the proposed TSM fuzzy regression model. In relation to this matter, a series of datasets consisting of 40 observations each, with varying percentages of outliers* $(P = 5\%, 10\%, 15\%)$*, are generated by simulation as shown below:*

$$\begin{aligned} u_i &\sim Uniform(1,4), \quad i = 1, 2, \cdots, 40 - 40P, \\ w_i &= 2 + 4u_i + N(0,1), \quad i = 1, 2, \cdots, 40 - 40P, \\ l_i &\sim Uniform(1,4), \quad i = 1, 2, \cdots, 40 - 40P, \\ r_i &\sim Uniform(1,4), \quad i = 1, 2, \cdots, 40 - 40P, \end{aligned}$$

We used a new scheme to increase the fraction of outliers in the input-output variables, resulting in the remaining 40 *injected outlier values We examine three fuzzy regression modelling simulation study approaches [8]. We design the*

generating scheme to produce observations that diverge from most of the $40 - 40P$ *sample values.*

Plan 1: *Using the formula below for* $i = 39$ *and* 40*, we can generate* 5% *of the data that deviates significantly from the model.*

$$\begin{aligned} &u_i \sim Uniform(3,4),\\ &w_i = u_i + N(0,1),\\ &l_i \sim Uniform(1,2),\\ &r_i \sim Uniform(1,2), \end{aligned}$$

Plan 2: *In this plan, we assume that ten percent of the data are outliers. For this purpose, we simulate the data from 37 to 40 as follows.*

$$\begin{aligned} &u_i \sim Uniform(4,6),\\ &w_i = -2 + u_i + N(0,1),\\ &l_i \sim Uniform(1,5),\\ &r_i \sim Uniform(1,5), \end{aligned}$$

Plan 3: *In the last part, we consider* 15% *of the data, far from the main data points and outliers, and as a result, we simulate and place the 35 data using the following relations and distributions:*

$$\begin{aligned} &u_i \sim Uniform(3,7),\\ &w_i = -7 + u_i + N(0,1),\\ &l_i \sim Uniform(3,6),\\ &r_i \sim Uniform(3,6), \end{aligned}$$

As we stated at the beginning, in this example, which is a quote from Chachi's article [8]. We simulated 1000 random samples in three designs with the presence of outlier data(Plan1, 2 and 3). In each repetition of estimations, we calculated the model's parameters using the TSM and also calculated the goodness of fit indices introduced in relation 6. Finally, we calculated the average of these indices after 1000 repetitions. Also, the calculated values were compared with the methods provided by Chachi [6], Xu and Li [42], Ferraro et al. [23] and the M-estimates method introduced by Chachi [8]. The results of this comparison are recorded in Table 2. The Theil-Sen method for the first plan shows a significant improvement over other methods, with the mean of the three goodness of fit criteria being $S = 0.8507$*,* $E_1 = 0.1687$*, and* $E_2 = 0.1129$*, respectively. We have obtained similar results for the mean of the three criteria and for the other two plans.*
Box plots for model evaluation criteria (left side) as well as comparison plots of real and fitted data (right side) for the designs presented in Fig. 1, Figs. 2 and 3, can be seen. For these methods, Chachi [8] has provided box plots of these indicators. These figures provide an overview of the statistical analysis of the goodness-of-fit criterion values. We conducted the analysis on 1000 datasets created for each outlier-generating technique. The box-plot data demonstrate that the suggested estimate approach offers superior performance criteria and efficiently eliminates outliers. These figures plot the observed and fitted values of the response variable for both the Theil-Sen estimator and least squares. This method observes the difference between the fitted lines. It is quite clear that adding the number of outliers to the regression line model using the least squares method deviates from the original data. We use the least squares and Theil-Sen estimator to draw the graph of the regression line that fits the data, focusing on the central points for enhanced clarity. We can draw a definitive conclusion based on the available evidence and logical reasoning.

- *The Theil-Sen fit demonstrates much superior tracking of the data compared to the other suggested fits.*
- *With the increase of outlier data, the accuracy of all methods decreases, but in comparison, the method introduced in this paper (Theil-Sen) has higher accuracy.*

Table 2 shows the goodness-of-fit measures of S, E1 and E2 for 1000 random samples from the given distributions with the assumed value of outlier data.

Table 2: Simulation result: A comparison was made between several models in a simulation research, considering varying percentages of outliers throughout 1000 simulation cycles.

		chachi	[8]	Ferraro et al.	[23]	Xu and Li	[42]	Theil-Sen	Theil-Sen
Outliers'	Plan	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Plan 1	S	0.429	0.428	0.344	0.344	0.347	0.347	0.8507	0.8512
	E_1	1.337	1.340	1.628	1.626	1.636	1.635	0.1687	0.1677
	E_2	0.905	0.907	1.101	1.101	1, 107	1.107	0.1129	0.1120
Plan 2	S	0.409	0.409	0.129	0.127	0.131	0.128	0.8201	0.8199
	E_1	1.570	1.563	2.705	2.711	2.768	2.775	0.2946	0.2931
	E_2	0.914	0, 910	1.659	1.663	1.695	1.700	0.1909	0.1873
Plan 3	S	0.256	0.223	0.122	0.120	0.124	0.122	0.7797	0.7800
	E_1	2.564	2.680	3.204	3.214	3.315	3.325	0.5616	0.5619
	E_2	1.301	1.456	1.699	1.704	1.757	1.760	0.3340	0.3305

Study real data

In this section, we analyse the introduced method and its results using real examples.

Example 3.3. *In this example, we use the data provided in Chachi [6] to provide a practical example. In Table 3, you can see the annual flow and suspended sediment concentration time series data for 12 watershed sites that were part of this study. Within each watershed region, there are different combinations of the following variables:*

- $\tilde{Load} = (Center, Spread)_T(\frac{T_{on}}{Year})$*: the suspended sediment load carried by a river in a year,*
- *Yield(Million* m^3*): the yearly flow of water in the river and/or watershed area,*
- *Area* (km^2)*: the area of the watershed,*
- *qs (* $\frac{T_{on}}{Year}\, km^2$*): the yearly amount of suspended sediment concentration in every one* km^2 *of the delta of the river, 5. qy (*$\frac{1000m^3}{km^2}$ *): the flow of water related for the yearly amount of suspended sediment concentration in every* km^2*,*

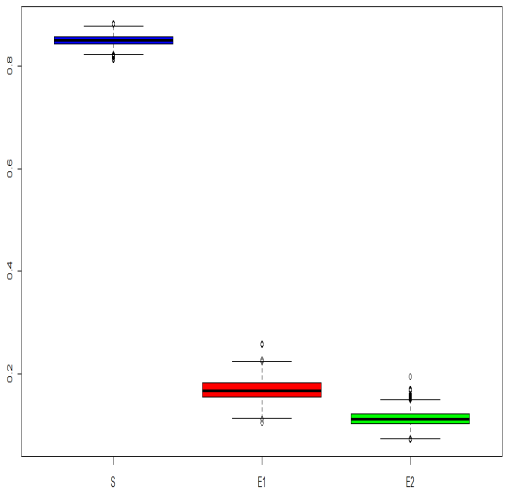
have been documented in order to determine the amount of suspended silt that a river carries throughout the following model:

$$\hat{\tilde{y}} = (\hat{\beta}_0 + \hat{\beta}_1 Yield + \hat{\beta}_2 Area + \hat{\beta}_3 qs + \hat{\beta}_4 qy, \exp\{\hat{\sigma}_0 + \hat{\sigma}_1 Yield + \hat{\sigma}_2 Area + \hat{\sigma}_3 qs + \hat{\sigma}_4 qy\})_T. \quad (7)$$

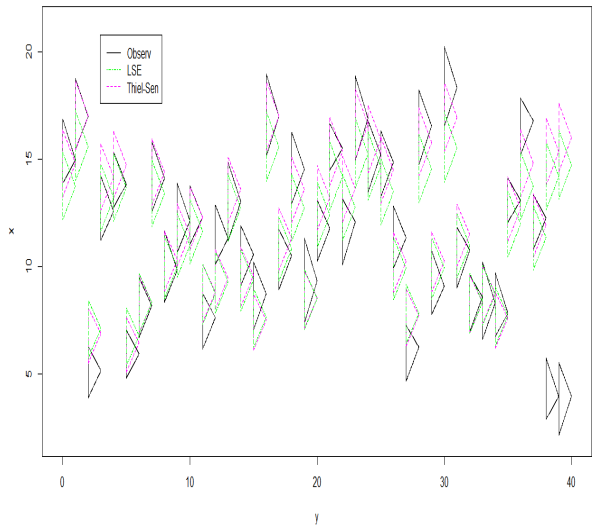
Table 3: The hydrological data's crisp inputs and fuzzy outputs

Location	$\tilde{Load} = (Center, Spread)_T$	Yield	Area	qs	qy	$\hat{\omega}_i$
1	$(163656, 27800)_T$	25.5	1117	150	22.85	1.000
2	$(245327, 22080)_T$	44.2	487	504	90.76	1.000
3	$(1061353, 116750)_T$	65.9	5925	179	11.12	0.145
4	$(162366, 24350)_T$	15.5	403	403	38.46	1.000
5	$(43133, 8650)_T$	15.8	207	208	76.33	1.000
6	$(480270, 28800)_T$	48.6	1256	382	38.69	1.000
7	$(410524, 28800)_T$	33.7	1355	303	24.87	1.000
8	$(248000, 19840)_T$	26.5	548	452	48.36	1.000
9	$(3877112, 77550)_T$	246.0	12037	322	20.44	1.000
10	$(5033425, 151000)_T$	299.9	17266	291	17.37	.176
11	$(9041480, 361660)_T$	445.0	27241	332	16.34	1.000
12	$(579670, 29000)_T$	22.0	884	656	6.78	1.000

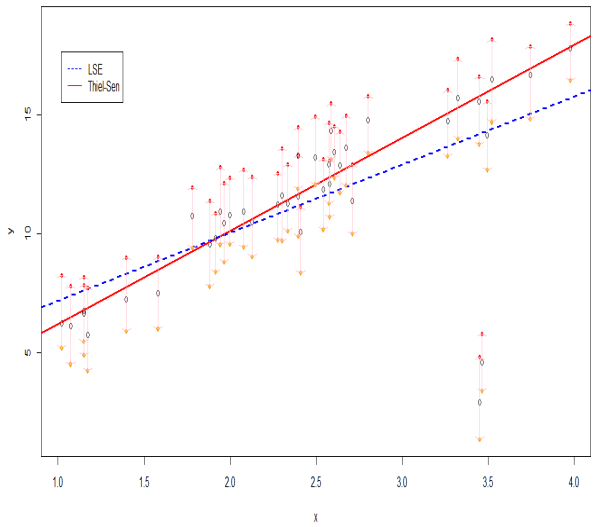
By introducing a new method using M-estimation, Chachi [8] made a comparison for ordinary least squares (Ferraro et al. [23]), weighted least squares (Chachi [6]) and M-estimation. We continue the comparison by adding the method introduced in this article while estimating the Theil-Sen parameters of model 7 for these data in Table 3. As is clear from the information in Table 4, the evaluation criteria confirm the superiority and goodness of this method compared



(a) The box plot in 1000 replications for evaluation criteria of equations 6 in Plan1.



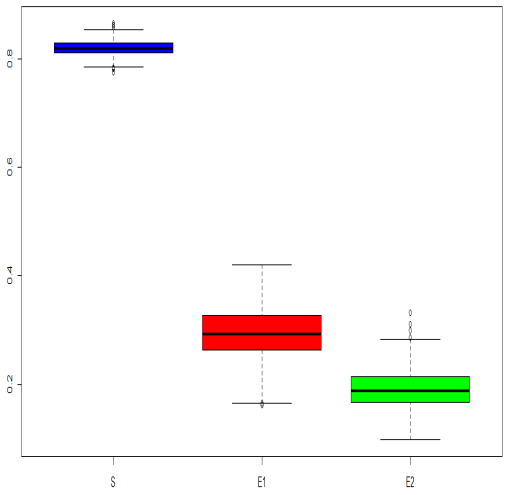
(b) The membership functions of $\tilde{y}$ (in black) and $\tilde{\hat{y}}$ for LSE regression (in green) and Theil-Sen estimator (in pink) in Plan1.



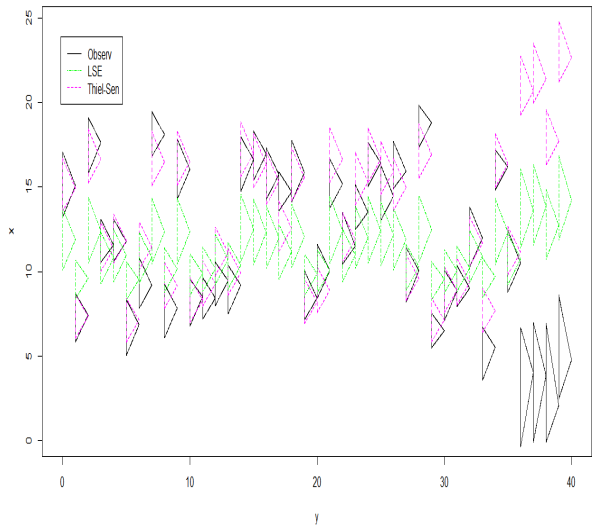
(c) Real data and the line estimated by the least square method (dashed line) and the line estimated by the Thiel-Sen method in Plan1.

Figure 1: Comparison of the accuracy of the models estimated by the Thiel-Sen and the least square in Plan 1, Example 3.2.

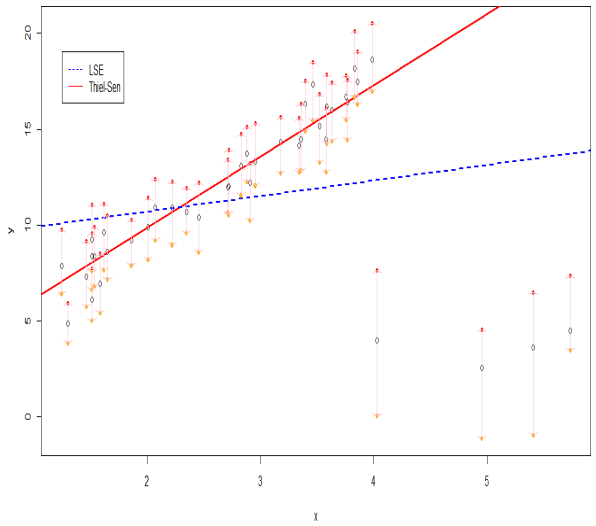
to all these methods. In the proposed method, the values of the goodness of fit criteria have been obtained as $S = 0.79$, $E_1 = 11695.65$, *and* $E_2 = 0.32$, *respectively, as it can be seen that the value of the similarity criterion* S *is significantly closer to one than other methods, and the other two criteria are much smaller than other methods. The predicted values for the parameters* β_j *and* σ_j, *where* j *ranges from 0 to 4, are shown in Table 5. We compute the model evaluation criteria S, E1, and E2 for the four estimated models to demonstrate the significant impact of outliers on the least-squares approach. The results show that the weighted method of the [6] estimator performs better than the ordinary least squares method. In the presence of outlier data, the M-estimation method performs exceptionally well. However, ultimately,*



(a) The box plot in 1000 replications for evaluation criteria of equations 6 in Plan 2



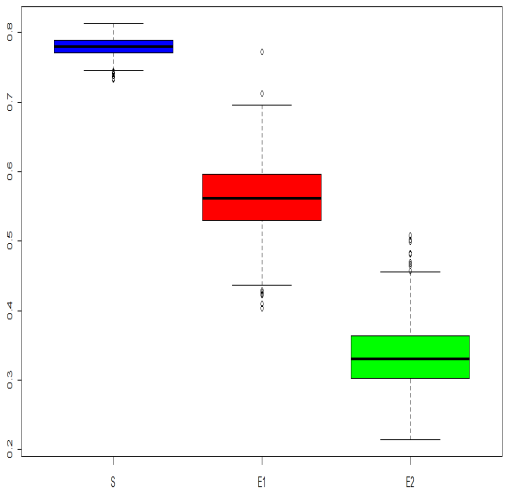
(b) The membership functions of $\tilde{y}$ (in black) and $\tilde{\hat{y}}$ for LSE regression (in green) and Theil-Sen estimator (in pink) in Plan 2



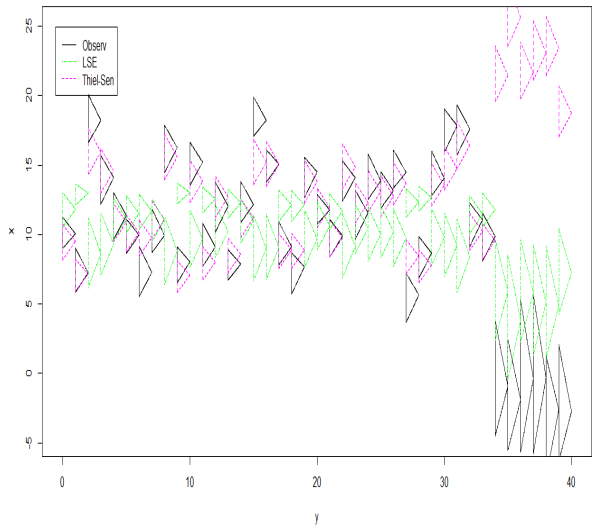
(c) Real data and the line estimated by the least square method (dashed line) and the line estimated by the Thiel-Sen method in Plan 2.

Figure 2: Comparison of the accuracy of the models estimated by the Thiel-Sen and the least square in Plan 2, Example 3.2.

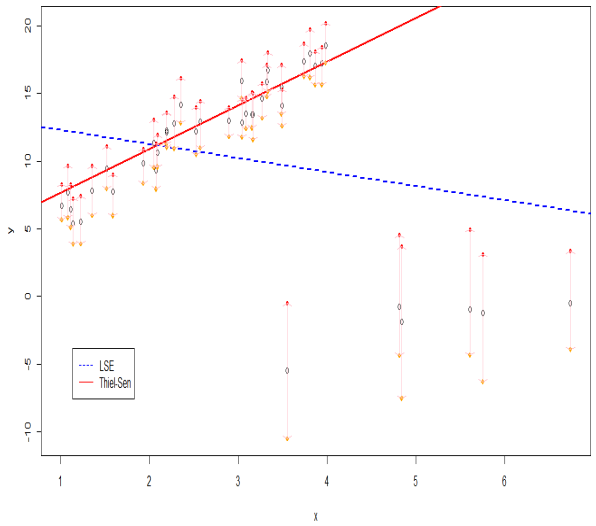
the Theil-Sen method outperforms all these methods, exhibiting higher accuracy and reduced error. So in the end, the Theil-Sen method has a much higher and more favourable performance than all the previously introduced methods. This method somewhat ignores the role of outlier data.



(a) The box plot in 1000 replications for evaluation criteria of equations 6 in Plan3.



(b) The membership functions of $\tilde{y}$ (in black) and $\tilde{\hat{y}}$ for LSE regression (in green) and Theil-Sen estimator (in pink) in Plan 3.



(c) Real data and the line estimated by the least square method (dashed line) and the line estimated by the Thiel-Sen method in Plan 3.

Figure 3: Comparison of the accuracy of the models estimated by the Thiel-Sen and the least square in Plan 3, Example 3.2.

Table 4: Examining four approaches to estimate the fuzzy regression model 7

Estimation Method	S	E_1	E_2
Approach proposed by [6]	0.14	118452.06	1.62
Approach proposed by [23]	0.06	123554.96	1.83
Approach proposed by [8]	0.2	80304.08	1.45
Theil-Sen	0.79	11695.65	0.32

Table 5: Model 7 parameters estimated using four methods

Technique	Intercept	Yield	Area	qs	qy
[6]	$(-257344.51, 10.34678)_T$	$(7066.46, 0.00146)_T$	$(175.05, 7E-5)_T$	$(585.97, 0.00078)_T$	$(-542.36, -0.01656)_T$
[23]	$(-650314.84, 10.43738)_T$	$(4226.51, -646E-5)_T$	$(257.39, 0.00020)_T$	$(1267.48, 22E-5)_T$	$(1730.62, -0.00965)_T$
[8]	$(-391972.63, 10.10876)_T$	$(-274.45, 0.00171)_T$	$(334.96, 7E-5)_T$	$(971.07, 0.00054)_T$	$(951.08, -0.00918)_T$
Theil-Sen	$(-385735.68, 10.2852)_T$	$(4000.52, -31E-4)_T$	$(264.88, 0.00015)_T$	$(988.85, -12E-6)_T$	$(188.42, -0.009897)_T$

Therefore, the results of our proposed model suggest the following model as a suitable fit for the data presented in Table 3.

$$\begin{aligned}\tilde{Load} &= (-385735.68, 10.2852)_T + (4000.52, -31E-4)_T Yield \\ &+ (264.88, 0.00015)_T Area + (988.85, -12E-6)_T qs + (188.42, -0.009897)_T qy\end{aligned}$$

As we said before (Table 4 results), it has less error and higher accuracy than other methods.

Example 3.4. *We use the standard dataset from Coppi et al. [16] to compare Theil-Sen's method again with the new ways of estimating the fuzzy regression model's parameters when there are outliers in the data. The dataset is connected to applied research and can be found in Table 6. This application aims to analyze the correlation between the atmospheric concentration of carbon monoxide (CO) (represented as a triangular fuzzy response variable) and a group of regression variables.*

- $x_1 =$ *temperature (C),*
- $x_2 =$ *relative humidity (%),*
- $x_3 =$ *atmospheric pressure (mbar),*
- $x_4 =$ *Rain (mm),*
- $x_5 =$ *radiation* $(\frac{mW}{cm^2})$,
- $x_6 =$ *wind speed* $(\frac{m}{s})$.

Table 6: The crisp input variables and fuzzy outcome variable in environmental research.

i	$\tilde{y}_i = (y_i, l_i, r_i)_T$	x_1	x_2	x_3	x_4	x_5	x_6	$\hat{\omega}_i$	$\hat{\omega}_i$[6]
1	$(1.15, 0.84, 0.60)_T$	19.04	82.27	1009.9	0.90	0.15	2.87	0.93	0.50
2	$(2.98, 2.83, 2.02)_T$	17.66	88.70	1017.5	0.90	0.16	1.12	1.00	2.92
3	$(3.92, 1.97, 1.97)_T$	18.15	82.51	1025.6	0.02	0.14	0.85	1.00	3.66
4	$(4.65, 1.89, 2.24)_T$	18.43	79.46	1032.1	0.00	0.16	0.45	1.00	2.46
5	$(3.98, 2.13, 2.13)_T$	20.67	68.85	1027.1	0.00	0.22	0.91	0.65	0.75
6	$(3.35, 2.63, 1.58)_T$	21.64	79.39	1020.0	0.02	0.29	1.07	1.00	1.74
7	$(3.13, 2.39, 1.71)_T$	18.85	88.87	1018.2	1.30	0.24	0.69	1.00	0.98
8	$(4.15, 2.41, 2.41)_T$	16.16	88.92	1020.9	0.09	0.16	0.40	1.00	1.43
9	$(3.96, 2.48, 2.10)_T$	17.02	83.52	1028.0	0.00	0.13	0.83	1.00	3.53
10	$(4.07, 2.01, 2.38)_T$	14.72	87.51	1025.3	0.05	0.04	2.09	0.25	0.29
11	$(3.30, 2.16, 1.83)_T$	15.89	86.04	1018.5	0.04	0.04	0.90	1.00	1.87
12	$(4.02, 2.35, 2.78)_T$	17.83	91.77	1015.6	0.93	0.06	1.22	0.31	0.22
13	$(2.06, 3.19, 1.59)_T$	20.96	83.62	1012.3	0.02	0.13	3.00	0.52	0.30
14	$(1.37, 0.66, 0.66)_T$	17.07	73.10	1020.4	0.02	0.13	1.75	0.27	0.51
15	$(3.35, 1.73, 1.73)_T$	14.71	80.38	1028.6	0.09	0.01	0.73	0.88	1.71
16	$(1.45, 1.15, 0.97)_T$	20.41	87.98	1026.4	0.24	0.11	1.87	0.17	0.17
17	$(2.74, 1.41, 1.41)_T$	19.04	82.27	1009.9	0.9	0.15	2.87	0.93	0.50
18	$(2.44, 1.95, 1.39)_T$	15.64	64.95	1022.4	0.00	0.08	1.25	1.00	1.42
19	$(2.79, 2.23, 1.59)_T$	13.22	80.16	1021.6	0.01	0.00	1.02	1.00	2.73
20	$(3.31, 1.73, 1.24)_T$	12.98	86.14	1023.4	0.44	0.00	0.70	1.00	2.04
21	$(4.02, 2.96, 2.11)_T$	13.10	89.12	1028.6	0.01	0.00	1.17	0.71	1.03

The estimation of parameters for the method introduced in this article and other methods are recorded in Table 7.

Table 7: Linear model parameters estimated using four methods

Estimation technique	Coefficient's	Estimate
Approach proposed by [6]	Intercept	$(-67.874, 15.785, -9.436)_T$
	x_1	$(0.065, -0.036, 0.035)_T$
	x_2	$(0.045, 0.011, 0.017)_T$
	x_3	$(0.065, -0.015, 0.008)_T$
	x_4	$(-0.441, -0.126, -0.136)_T$
	x_5	$(0.719, 1.532, -0.162)_T$
	x_6	$(-0.813, -0.242, -0.306)_T$
Approach proposed by [23]	Intercept	$(-33.98, 24.74, -0.12)_T$
	x_1	$(-0.02, 0.02, 0.04)_T$
	x_2	$(0.05, 0.02, 0.02)_T$
	x_3	$(0.03, -0.02, -0.0014)_T$
	x_4	$(-0.55, -0.36, -0.27)_T$
	x_5	$(1.92, -0.16, -0.79)_T$
	x_6	$(-0.76, -0.32, -0.35)_T$
Approach proposed by [8]	Intercept	$(-57.749, 18.513, -7)_T$
	x_1	$(0.077, 0.018, 0.065)_T$
	x_2	$(0.043, 0.013, 0.018)_T$
	x_3	$(0.055, -0.018, 0.005)_T$
	x_4	$(-0.465, -0.271, -0.231)_T$
	x_5	$(-0.027, 0.047, -1.073)_T$
	x_6	$(-0.766, -0.311, -0.357)_T$
Approach proposed	Intercept	$(-64.303, 17.27, -10.309)_T$
	x_1	$(0.043, -0.0356, 0.04278)_T$
	x_2	$(0.0426, 0.01246, 0.0142)_T$
	x_3	$(0.06293, -0.01628, 0.00915)_T$
	x_4	$(-0.4644, -0.2140, -0.4078)_T$
	x_5	$(0.9419, 1.3022, -0.9117)_T$
Proposed approach(Theil-Sen)	x_6	$(-0.5946, -0.2248, -0.2341)_T$

For a better comparison between different methods, different goodness-of-fit criteria have been calculated and displayed in Table 8. The results of this table are also similar to the previous ones, which explains the superiority of the introduced method over other methods.The goodness of fit criteria calculated for these data in the Theil-Sen method are $S = 0.7864$*,* $E_1 = 0.3874$*, and* $E_2 = 0.2089$*, respectively, which shows higher accuracy and less error than the values calculated for other methods. Furthermore, our proposed method estimates the model as follows:*

$$\begin{aligned}\tilde{y} &= (-64.303, 17.27, -10.309)_T + (0.043, -0.0356, 0.04278)_T x_1 + (0.0426, 0.01246, 0.0142)_T x_2 \\ &+ (0.06293, -0.01628, 0.00915)_T x_3 + (-0.4644, -0.2140, -0.4078)_T x_4 + (0.9419, 1.3022, -0.9117)_T x_5 \\ &+ (-0.5946, -0.2248, -0.2341)_T x_6\end{aligned}$$

Table 8: Comparative analysis of four methods for estimating a fuzzy linear regression model

Estimation technique	S	E_1	E_2
Approach proposed by [6]	0.6437	0.8538	0.5326
Approach proposed by [23]	0.6231	0.9022	0.5605
Approach proposed by [8]	0.6505	0.8359	0.5317
Proposed approach	0.7864	0.3874	0.2089

4 Conclusion

People often recognize the ordinary least squares approach as a prominent and long-standing statistical technique for parameter estimation. Outliers in the dataset compromise the features of ordinary least square estimators, including

their unbiasedness, lowest variance, and optimality as the best linear unbiased estimator. Hence, the presence of outliers has an influence on ordinary least squares estimators due to their heightened susceptibility to the influence of outliers. We use robust regression techniques when the dataset displays outliers. The use of robust regression analysis serves as a viable alternative to the conventional "least squares regression model" in situations where the essential assumptions of the data are not satisfied.
This study used the Theil-Sen approach to create a fuzzy regression model for data with crisp input and fuzzy output. The suggested technique, with its exceptional predictive capabilities and the need to adjust outlier identification, may find efficient use in many applications. We evaluated the suggested methodology against other established fuzzy regression approaches using four datasets that contained real-valued data. The article's goodness of fit criteria, which include similarity criteria and absolute error, indicate that the proposed method outperforms other well-known methods. Theil-Sen estimators, in fact, utilize the median of the least squares estimators for various sample permutations. Therefore, based on the characteristics of the median in relation to the influence of outlier data, these estimators accept the least influence from outlier data, making this method robust. Although this article considers fuzzy numbers to be triangular, the results are applicable to all fuzzy numbers. In general, we can apply this method, yet it outperforms other methods in situations where outliers or incomplete data exist. When the data are symmetrical or devoid of outliers, the mean and median nearly match, ensuring that employing this method won't harm us overall. The extension of the proposed model to Intuitionistic and hesitant fuzzy responses is a topic that can be considered in future research. It is possible to extend the proposed technique to include common variable selection methods in high-dimensional scenarios. This would be another potential area for investigation in future research.

Conflict of interest

The authors declare that they have no conflict of interest.

Data Availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study

Funding

This study has not received any external funding.

References

[1] M. G. Akbari, G. Hesamian, *A partial-robust-ridge-based regression model with fuzzy predictors-responses*, Journal of Computational and Applied Mathematics, **351** (2019), 290-301. https://doi.org/10.1016/j.cam.2018.11.006

[2] M. G. Amiri, A. R. Zarei, J. Abedi-Koupai, S. Eslamian, *The performance of fuzzy regression method for estimating of reference evapotranspiration under controlled environment*, International Journal of Hydrology Science and Technology, **9**(1) (2019), 28-38. https://doi.org/10.1504/IJHST.2019.096791

[3] M. Arefi, *Quantile fuzzy regression based on fuzzy outputs and fuzzy parameters*, Soft Computing, **24**(1) (2020), 311-320. https://doi.org/10.1007/s00500-019-04424-2

[4] H. T. S. U. K. Asai, S. Tanaka, K. Uegima, *Linear regression analysis with fuzzy model*, IEEE Transactions Systems, Man, and Cybernetics, **12** (1982), 903-907. https://doi.org/10.1109/TSMC.1982.4308925

[5] E. Bas, *Robust fuzzy regression functions approaches*, Information Sciences, **613** (2022), 419-434. https://doi.org/10.1016/j.ins.2022.09.047

[6] J. Chachi, *A weighted least squares fuzzy regression for crisp input-fuzzy output data*, IEEE Transactions on Fuzzy Systems, **27**(4) (2017), 739-748. https://doi.org/10.1109/TFUZZ.2018.2868554

[7] J. Chachi, M. Roozbeh, *A fuzzy robust regression approach applied to bedload transport data*, Communications in Statistics-Simulation and Computation, **46**(3) (2017), 1703-1714. https://doi.org/10.1080/03610918.2015.1010002

[8] J. Chachi, S. M. Taheri, P. D'Urso, *Fuzzy regression analysis based on M-estimates*, Expert Systems with Applications, **187** (2022), 115891. https://doi.org/10.1016/j.eswa.2021.115891

[9] J. Chachi, S. M. Taheri, S. Fattahi, S. A. Hosseini Ravandi, *Two robust fuzzy regression models and their applications in predicting imperfections of cotton yarn*, Journal of Textiles and Polymers, **4**(2) (2016), 60-68.

[10] Y. H. O. Chang, B. M. Ayyub, *Fuzzy regression methods-a comparative assessment*, Fuzzy Sets and Systems, **119**(2) (2001) 187-203. https://doi.org/10.1016/S0165-0114(99)00091-3

[11] G. Charizanos, H. Demirhan, D. İçen, *A Monte Carlo fuzzy logistic regression framework against imbalance and separation*, Information Sciences, **655** (2024), 119893. https://doi.org/10.1016/j.ins.2023.119893

[12] Y. S. Chen, *Outliers detection and confidence interval modification in fuzzy regression*, Fuzzy Sets and Systems, **119**(2) (2001), 259-272. https://doi.org/10.1016/S0165-0114(99)00049-4

[13] H. Chervenkov, K. Slavov, *Theil-sen estimator for the parameters of the generalized extreme value distributions: Demonstration for meteorological applications*, Comptes rendus de I'Académie bulgare des Sciences, **70**(12) (2017), 1701-1708.

[14] H. Chervenkov, K. Slavov, *Theil-Sen estimator vs. ordinary least squares-trend analysis for selected ETCCDI climate indices*, Comptes Rendus de L'Académie Bulgare des Sciences, **72** (2019), 47-54. https://doi.org/10.7546/CRABS.2019.01.06

[15] N. Chukhrova, A. Johannssen, *Fuzzy regression analysis: Systematic review and bibliography*, Applied Soft Computing, **84** (2019), 105708. https://doi.org/10.1016/j.asoc.2019.105708

[16] R. Coppi, P. D'Urso, P. Giordani, A. Santoro, *Least squares estimation of a linear regression model with LR fuzzy response*, Computational Statistics and Data Analysis, **51** (2006), 267-286. https://doi.org/10.1016/j.csda.2006.04.036

[17] X. Dang, H. Peng, X. Wang, H. Zhang, *Theil-sen estimators in a multiple linear regression model*, Olemiss Edu, The University of Mississippi, 2008.

[18] P. D'Urso, R. Massari, *Weighted least squares and least median squares estimation for the fuzzy linear regression analysis*, Metron, **71** (2013), 279-306. https://doi.org/10.1007/s40300-013-0025-9

[19] P. D'Urso, R. Massari, A. Santoro, *Robust fuzzy regression analysis*, Information Sciences, **181**(19) (2011), 4154-4174. https://doi.org/10.1016/j.ins.2011.04.031

[20] E. Egrioglu, E. Bas, *Robust intuitionistic fuzzy regression functions approaches*, Information Sciences, **638** (2023), 118992. https://doi.org/10.1016/j.ins.2023.118992

[21] A. H. El-Shaarawi, W. W. Piegorsch, *Encyclopedia of environmetrics*, Volume 1, John Wiley and Sons, p. 19, ISBN 978-0-471-89997-6, 2001.

[22] R. Fernandes, S. G. Leblanc, *Parametric (modified least squares) and non-parametric (Theil-Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors*, Remote Sensing of Environment, **95**(3) (2005), 303-316. https://doi.org/10.1016/j.rse.2005.01.005

[23] M. B. Ferraro, P. Giordani, *A proposal of robust regression for random fuzzy sets. In synergies of soft computing and statistics for intelligent data analysis*, Springer Berlin Heidelberg, 2013. https://doi.org/10.1007/978-3-642-33042-1_13

[24] G. Hesamian, M. G. Akbari, *A robust multiple regression model based on fuzzy random variables*, Journal of Computational and Applied Mathematics, **388** (2021), 113270. https://doi.org/10.1016/j.cam.2020.113270

[25] P. J. Huber, *Robust statistics*, John Wiley and Sons, 2004.

[26] W. L. Hung, M. S. Yang, *An omission approach for detecting outliers in fuzzy regression models*, Fuzzy Sets and Systems, **157**(23) (2006), 3109-3122. https://doi.org/10.1016/j.fss.2006.08.004

[27] R. Jiang, Y. Xin, Z. Chen, Y. Zhang, *A medical big data access control model based on fuzzy trust prediction and regression analysis*, Applied Soft Computing, **117** (2022), 108423. https://doi.org/10.1016/j.asoc.2022.108423

[28] L. Kokkinen, *Studying social determinants of health using fuzzy-set qualitative comparative analysis: A worked example*, Social Science and Medicine, **309** (2022), 115241. https://doi.org/10.1016/j.socscimed.2022.115241

[29] K. S. Kula, A. Apaydin, *Fuzzy robust regression analysis based on the ranking of fuzzy sets*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **16**(05) (2008), 663-681. https://doi.org/10.1142/S0218488508005558

[30] J. M. Leski, M. Kotas, *On robust fuzzy c-regression models*, Fuzzy Sets and Systems, **279** (2015), 112-129. https://doi.org/10.1016/j.fss.2014.12.004

[31] Y. Li, X. He, X. Liu, *Fuzzy multiple linear least squares regression analysis*, Fuzzy Sets and Systems, **459** (2023), 118-143. https://doi.org/10.1016/j.fss.2022.06.012

[32] F. Maturo, S. Hošková-Mayerová, *Fuzzy regression models and alternative operations for economic and social sciences*, Recent Trends in Social Systems: Quantitative Theories and Quantitative Models, (2017), 235-247. https://doi.org/10.1007/978-3-319-40585-8_21

[33] Z. Mei, T. Zhao, X. Xie, *Hierarchical fuzzy regression tree: A new gradient boosting approach to design a TSK fuzzy model*, Information Sciences, **652** (2024), 119740. https://doi.org/10.1016/j.ins.2023.119740

[34] E. Nasrabadi, S. M. Hashemi, *Robust fuzzy regression analysis using neural networks*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **16**(04) (2008), 579-598. https://doi.org/10.1142/S021848850800542X

[35] E. Nasrabadi, S. M. Hashemi, M. Ghatee, *An LP-based approach to outliers detection in fuzzy regression analysis*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, **15**(04) (2007), 441-456. https://doi.org/10.1142/S0218488507004789

[36] G. Peters, *Fuzzy linear regression with fuzzy intervals*, Fuzzy Sets and Systems, **63**(1) (1994), 45-55. https://doi.org/10.1016/0165-0114(94)90144-9

[37] P. J. Rousseeuw, A. M. Leroy, *Robust regression and outlier detection*, John Wiley and Sons, 2005.

[38] M. Salman, X. Long, G. Wang, D. Zha, *Paris climate agreement and global environmental efficiency: New evidence from fuzzy regression discontinuity design*, Energy Policy, **168** (2022), 113128. https://doi.org/10.1016/j.enpol.2022.113128

[39] H. Theil, *A rank-invariant method of linear and polynomial regression analysis*, Proceedings of the Royal Netherlands Academy of Sciences, **53** (1950), 368-392 (Part I), 521-525 (Part II), 1397-1412 (Part III).

[40] S. Varga, *Robust estimations in classical regression models versus robust estimations in fuzzy regression models*, Kybernetika, **43**(4) (2007), 503-508.

[41] H. Von Storch, F. W. Zwiers, *Statistical analysis in climate research*, Cambridge University Press, 2022.

[42] R. Xu, C. Li, *Multidimensional least-squares fitting with a fuzzy model*, Fuzzy Sets and Systems, **119** (2001), 215-223. https://doi.org/10.1016/S0165-0114(98)00350-9

[43] T. Zaman, K. Alakuş, *Integrating Jackknife into the Theil-Sen estimator in multiple linear regression model*, Revstat-Statistical Journal, **21**(1) (2023), 97-114. https://doi.org/10.57805/revstat.v21i1.398

[44] Y. A. Zelenkov, E. V. Lashkevich, *Fuzzy regression model of the impact of technology on living standards*, Business Informatics, **14**(3) (2023), 67-81. https://doi.org/10.17323/2587-814X.2020.3.67.81

[45] J. Zhou, H. Zhang, Y. Gu, A. A. Pantelous, *Affordable levels of house prices using fuzzy linear regression analysis: the case of Shanghai*, Soft Computing, **22** (2018), 5407-5418. https://doi.org/10.1007/s00500-018-3090-4