1. Introduction

1.1. Related Works and Motivations. Regression analysis is widely used in engineering science, social science, economy, and finance [1–4] since it is a significant and comprehensive method to analyze the dependence between dependent variables and one or more variables. Generally, in a regression model, deviations (errors) between the estimated and the observed values are deemed to be because of random variations and/or measurement errors. To this end, some statistical analysis techniques are usually developed for model determination. Nevertheless, in practice, the deviations are sometimes caused by the indefiniteness of the system structure or the incomplete measurable data. In such a case, the uncertainties are ordinarily called fuzziness but not randomness. Then, to deal with the fuzziness, the regression analysis based on fuzzy data or fuzzy numbers called fuzzy regression analysis was proposed and developed in [5–7].

Different from conventional regression techniques which are based on the nonstatistical method, fuzzy regression methodology uses possibility theory and fuzzy set theory [8, 9] for modeling and analysis. Thus, it is much appropriate to manage uncertainties. Zadeh [9] has introduced two types of fuzzy sets to describe the uncertainties [10–12]. The first type of fuzzy set called the type-1 fuzzy set has several typical extended forms, such as interval fuzzy set, triangular fuzzy set, trapezoidal fuzzy set, pentagonal fuzzy set, and intuitionistic fuzzy set [13]. Because of some variations in dealing with uncertainties, such as the inaccurate description of language uncertainties and the parameter perturbation in the uncertainties, the type-2 fuzzy set was presented to tackle this problem. Its degree of ambiguity is characterized by two subordinate membership functions. This is actually an extension of the type-1 one and leads to wider applicability for complex systems. But, in return, it makes the calculation of the fuzzy regression complicated when using type-2 fuzzy sets. For simplicity of calculation, Mendel et al. [14] specialized the type-2 fuzzy sets as a kind of interval type-2 fuzzy sets, which can also describe the uncertainties better over the type-1 fuzzy sets. In many research and applications of the fuzzy regression model, the fuzzy numbers based on triangular fuzzy sets are most frequently applied due to its simplicity. However, such fuzzy numbers manifest some limitations in observation, particularly when the complete observed output of the predicted model is required [15].
In many practical studies, the complete description of regression problems greatly depends on the property of input-output data [16]. Different types of input-output data lead to dissimilar analysis results of the fuzzy regression. There are typically four cases of the fuzzy regression analysis. These are the cases based on the crisp-input crisp-output (CICO) [17], crisp-input fuzzy-output (CIFO) [15, 18], fuzzy-input crisp-output (FICO) [19], and fuzzy-input fuzzy-output (FIFO) [19–21] observations in the literature. Most commonly, the case of CIFO data is studied in practice. In this paper, the interested fuzzy regression is analyzed based on CIFO data. Specifically, the case where the predictor variable is crisp but the parameters (coefficients) are fuzzy is considered. Therefore, the observed responses, in this case, are naturally fuzzy.

To sum up, there are three main methods of the fuzzy regression analysis, such as least squares fitting criterion, minimum fuzzy criterion, and interval regression analysis approach [22]. For example, in [23], mathematical programming methods were used for estimating the parameters of a fuzzy regression model in terms of the trapezoidal case and triangular case. The work in [24] developed a fuzzy regression model and used the least square method to estimate the coefficients in the sense of distance. The authors [15] presented a modified fuzzy linear model, based on which all the observed data can be enveloped by the identified model output. A tolerance approach was introduced in [25] to the construction of fuzzy regression coefficients based on a possibilistic linear regression model with fuzzy data. In this paper, the least square method will be used to calculate the estimation error of the fuzzy regression values.

1.2. Contributions of This Work. Actually, the kernel interval in a trapezoidal fuzzy number is limited to a single point equal to the midpoint of the support interval. As an extension of the triangular fuzzy numbers, the trapezoidal fuzzy numbers can fill those gaps. That is why we develop the fuzzy regression model using the trapezoidal interval type-2 fuzzy numbers and even use a more general case quadrilateral interval type-2 fuzzy numbers considered in this paper. Meanwhile, the model structure is assumed to be linear, which is commonly used in the literature. Consequently, the corresponding fuzzy regression problem becomes a parameter estimation problem of the regression model.

In another research field, data information of all individuals often fails to be collected by experimenters due to measurement methods, preservation methods, and human factors. It results in incomplete observation values of some data indicators in the samples, which are common in clinical trials, socioeconomic statistics, environmental ecology, and other researches. In fact, under normal circumstances, samples with missing values cannot fully reflect the real characteristics of interested systems and the internal relationship between variables. Thus, improper treatment may even lead to large deviations in the results. Therefore, how to deal with the missing data and extract information from data correctly and effectively is an important issue in statistical inference.

However, it is often unavoidable to mix a certain proportion of outliers or strong influence points into the actual data due to the interference of many factors, such as negligence error and rounding error. Once the outliers are mixed, these fuzzy regression methods will become unpractical and even severe challenges can even lead to wrong conclusions. It results in incomplete observation values of some data indicators in the samples, which are common in clinical trials, socioeconomic statistics, environmental ecology, and other researches [26–28]. Hence, the influence analysis of outliers on models is an important part of the statistical diagnosis. To this end, this paper will investigate the fuzzy regression analysis based on the type-2 trapezoidal fuzzy numbers regarding the data outlier detection. The contributions of the paper can be summarized as follows:

(1) Firstly, the definition of interval trapezoidal type-2 fuzzy numbers is provided. The parameter estimation laws of the fuzzy linear regression model based on trapezoidal fuzzy numbers are designed in the sense of Euclidean distance.

(2) Then, some parameter estimation laws in terms of the data outlier detection are synthesized for the trapezoidal fuzzy regression model.

(3) Moreover, based on the trapezoidal fuzzy linear regression model, the impact evaluation rule is established by using the root mean squared error method.

The rest of the paper is organized as follows. Section 2 describes the fuzzy linear regression model and the quadrilateral type-2 fuzzy numbers. Sections 3 and 4 present parameter estimation laws of the fuzzy regression model. Section 5 provides the impact evaluation rule. Section 7 summarizes the paper.

2. Model Description and Preliminaries

\((x_i, y_i)\) is denoted as a set of pairs of observation data, where \(x_i \in R\) is the predictor variable (input) and \(y_i \in R\) is the observed response variable (output), and \(i = 1, 2, \ldots, n\). For each observation data \((x_i, y_i)\), the functional form of the linear regression model is formulated as

\[ y_i = \alpha + \beta x_i, \]

where \(\alpha \in R\) and \(\beta \in R\) are the unknown parameters to be estimated.

Some definitions used in developing the theoretical results are presented in Appendix.

The uncertainty of a type-2 fuzzy sets \(F\) can be described by a bounded region, that is, the projection area of the fuzzy sets \(F\) on the plane \((x, u)\), which is called the footprint of uncertainty, expressed by \(\mu_F(x, u)\). The upper-bound membership function \(\mu_U^F(x, u)\) and the lower-bound membership function \(\mu_L^F(x, u)\) of the interval type-2 fuzzy numbers are actually corresponding to type-1 fuzzy sets, respectively.

An illustrative description and comparison of the type-2 membership functions under different cases is shown in
from the defined Euclidean distance, for the estimation of the $\bar{a}$ and $\bar{b}$ in (3). Furthermore, the parameter estimation in terms of the data outlier will be also discussed subsequently.

3. Parameter Estimation of the Type-2 Fuzzy Regression Model

For the considered CIFO observation data, we write the interval type-2 fuzzy parameters $\bar{a}$ and $\bar{b}$ in (3) as follows:

$$\bar{a} = \left( \left[ a^U_1, a^L_1, a^U_3, a^L_3 ; h^U_a, h^L_a \right], \left[ a^U_2, a^L_2, a^U_4, a^L_4 ; h^U_a, h^L_a \right] \right),$$

$$\bar{b} = \left( \left[ b^U_1, b^L_1, b^U_3, b^L_3 ; h^U_b, h^L_b \right], \left[ b^U_2, b^L_2, b^U_4, b^L_4 ; h^U_b, h^L_b \right] \right).$$

According to (3), the fuzzy-observed response $\tilde{y}_i$ can be represented subject to the positive or negative $x_i$. If $x_i > 0$, then we can write the type-2 fuzzy set membership function as

$$\mu^U_i(x, h^U) = \left\{ \begin{array}{ll}
\frac{h^U \cdot (x - a^U_i)}{a^U_i - a^L_i}, & a^L_i < x < a^U_i, \\
h^U, & a^L_i \leq x \leq a^U_i, \\
\frac{h^U \cdot (a^U_i - x)}{a^U_i - a^L_i}, & a^L_i < x < a^U_i, \\
0, & \text{else},
\end{array} \right.$$

$$\mu^L_i(x, h^L) = \left\{ \begin{array}{ll}
\frac{h^L \cdot (x - a^L_i)}{a^U_i - a^L_i}, & a^L_i < x < a^U_i, \\
h^L, & a^L_i \leq x \leq a^U_i, \\
\frac{h^L \cdot (a^U_i - x)}{a^U_i - a^L_i}, & a^L_i < x < a^U_i, \\
0, & \text{else},
\end{array} \right.$$\

where $a^U_1 \leq a^L_1 \leq a^U_3 \leq a^L_3$ and $a^L_1 \leq a^L_2 \leq a^L_4 \leq a^U_4$ are nonnegative real numbers, and $0 \leq h^L \leq h^U \leq 1$ denote the heights of the two trapezoids. Considering $h_i \neq h_i$, we give the following definition of general quadrilateral interval type-2 fuzzy numbers.

Based on this discussion above, we will analyze the regression model (3) for the general case when the quadrilateral type-2 fuzzy numbers are used. The CIFO-based interval type-2 fuzzy regression model is formulated as follows:

$$\tilde{y}_i = \bar{a} + \bar{b}x_i,$$

where $\bar{a}$ and $\bar{b}$ denote the two fuzzy numbers, which are quadrilateral interval type-2 fuzzy numbers considered in this paper. $\tilde{y}_i$ is the observed response $(i = 1, 2, \ldots, n)$. Before proceeding further, we give the following definition of the Euclidean distance of two quadrilateral interval type-2 fuzzy numbers.

**Remark 1.** The coefficients $a$ and $b$ used in (A.6) can be adjusted as needed. If $a = b = 1/2$, then the Euclidean distance $d(F_1, F_2)$ is called the centralized Euclidean distance.

In the following section, the fuzzy regression is analyzed in the sense of the defined Euclidean distance, for the estimation of the $\bar{a}$ and $\bar{b}$ in (3). Furthermore, the parameter estimation in terms of the data outlier will be also discussed subsequently.

**Theorem 1.** Consider the fuzzy regression model $\tilde{y}_i = \bar{a} + \bar{b}x_i$, with a set of observation data $(x_i, \tilde{y}_i)$, $i = 1, 2, \ldots, n$. For the quadrilateral type-2 fuzzy parameters $\bar{a}$ and $\bar{b}$ in (4)-(5), if the observed data $\tilde{y}_i$, $i = 1, 2, \ldots, n$, are a set of quadrilateral type-2 fuzzy numbers $\tilde{y}_i = ((y^U_1, y^L_1), (y^U_2, y^L_2), \ldots, (y^U_n, y^L_n))$, then the
estimates of $\bar{\alpha}$ and $\bar{\beta}$ in the Euclidean distance sense are designed as follows:

$$
\hat{\alpha}_1^\kappa = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\beta}_1^\kappa = \frac{\sum_{i=1}^{n} x_i y_i^\kappa - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\alpha}_2^\kappa = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\beta}_2^\kappa = \frac{\sum_{i=1}^{n} x_i y_i^\kappa - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\alpha}_3^\kappa = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\beta}_3^\kappa = \frac{\sum_{i=1}^{n} x_i y_i^\kappa - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\alpha}_4^\kappa = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

$$
\hat{\beta}_4^\kappa = \frac{\sum_{i=1}^{n} x_i y_i^\kappa - \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^\kappa}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2},
$$

where $\kappa \in \{U, L\}$.

**Proof.** According to Definition 4, for $n$ observation pairs $(x_i, \bar{y}_i)$, the sum $s(\bar{\alpha}, \bar{\beta})$ of the squared Euclidean distance between $\bar{y}_i$ and $\bar{\alpha} + \bar{\beta}x_i$, subject to the fuzzy numbers $\bar{\alpha}$ and $\bar{\beta}$ can be obtained as follows:

$$
s(\bar{\alpha}, \bar{\beta}) = \sum_{i=1}^{n} d_{E}(\bar{y}_i, \bar{\alpha} + \bar{\beta}x_i) = \sum_{i=1}^{n} \sqrt{a \cdot d^U(y_i^U, \bar{\alpha}^U + \bar{\beta}^U x_i)}
$$

$$
+ \sum_{i=1}^{n} b \cdot d^L(y_i^L, \bar{\alpha}^L + \bar{\beta}^L x_i)
$$

$$
= a \cdot \left\{ \frac{1}{4} \sum_{i=1}^{n} \left( [y_i^U - (\alpha_i^U + \beta_i^U x_i)]^2 
+ [y_i^L - (h_i^U \alpha_i^U + h_i^L \beta_i^L x_i)]^2 
+ [y_i^U - (h_i^L \alpha_i^L + h_i^U \beta_i^U x_i)]^2 
+ [y_i^L - (\alpha_i^L + \beta_i^L x_i)]^2 \right) \right\}^{1/2}
$$

$$
+ b \cdot \left\{ \frac{1}{4} \sum_{i=1}^{n} \left( [y_i^U - (\alpha_i^U + \beta_i^U x_i)]^2 
+ [y_i^L - (h_i^U \alpha_i^U + h_i^L \beta_i^L x_i)]^2 
+ [y_i^U - (h_i^L \alpha_i^L + h_i^U \beta_i^U x_i)]^2 
+ [y_i^L - (\alpha_i^L + \beta_i^L x_i)]^2 \right) \right\}^{1/2}.
$$

(9)

Then, we take the partial derivative of $s(\bar{\alpha}, \bar{\beta})$ with respect to $\bar{\alpha}_\tau^\kappa$ and $\bar{\beta}_\tau^\kappa$ ($\tau = 1, 2, 3, 4$ and $\kappa \in \{U, L\}$) and receptively obtain the following equations for $\kappa \in \{U, L\}$:
\[
\frac{\partial s}{\partial \alpha_i} = \sum_{i=1}^{n} \left[ y_{i1} - (\alpha_i^x + \beta_i^x x_i) \right] = 0,
\]
\[
\frac{\partial s}{\partial \alpha_2} = \sum_{i=1}^{n} \left[ y_{i2} - \left( h_{a_2}^x \alpha_2^x + h_{\beta_2}^x \beta_2^x x_i \right) \right] = 0,
\]
\[
\frac{\partial s}{\partial \alpha_3} = \sum_{i=1}^{n} \left[ y_{i3} - \left( h_{a_3}^x \alpha_3^x + h_{\beta_3}^x \beta_3^x x_i \right) \right] = 0,
\]
\[
\frac{\partial s}{\partial \alpha_4} = \sum_{i=1}^{n} \left[ y_{i4} - (\alpha_4^x + \beta_4^x x_i) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_1} = \sum_{i=1}^{n} \left[ y_{i1} - (\alpha_1^x + \beta_1^x x_i) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_2} = \sum_{i=1}^{n} \left[ y_{i2} - \left( h_{a_2}^x \alpha_2^x + h_{\beta_2}^x \beta_2^x x_i \right) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_3} = \sum_{i=1}^{n} \left[ y_{i3} - \left( h_{a_3}^x \alpha_3^x + h_{\beta_3}^x \beta_3^x x_i \right) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_4} = \sum_{i=1}^{n} \left[ y_{i4} - (\alpha_4^x + \beta_4^x x_i) \right] x_i = 0.
\]

Let
\[
\tilde{\alpha} = \left( \tilde{\alpha}_1^U, \tilde{\alpha}_2^U, \tilde{\alpha}_3^U, \tilde{\alpha}_4^U, h_{\alpha_1}^U, h_{\alpha_2}^U, \tilde{\alpha}_1^L, \tilde{\alpha}_2^L, \tilde{\alpha}_3^L, h_{\alpha_4}^L, h_{\alpha_3}^L \right),
\]
\[
\tilde{\beta} = \left( \tilde{\beta}_1^U, \tilde{\beta}_2^U, \tilde{\beta}_3^U, \tilde{\beta}_4^U, h_{\beta_1}^U, h_{\beta_2}^U, \tilde{\beta}_1^L, \tilde{\beta}_2^L, \tilde{\beta}_3^L, h_{\beta_4}^L, h_{\beta_3}^L \right).
\]

be the estimates of \( \tilde{\alpha} \) and \( \tilde{\beta} \), respectively. Then, solving the algebraic equation sets obtained above, we get the estimates of the bounds of \( \tilde{\alpha} \) and \( \tilde{\beta} \) as exactly expressed in Theorem 1. This completes the proof.

\[
\frac{\partial s}{\partial \alpha_i} = \sum_{i=1}^{n} \left[ y_{i1} - (\alpha_i^x + \beta_i^x x_i) \right] = 0,
\]
\[
\frac{\partial s}{\partial \beta_1} = \sum_{i=1}^{n} \left[ y_{i1} - (\alpha_1^x + \beta_1^x x_i) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_2} = \sum_{i=1}^{n} \left[ y_{i2} - \left( h_{a_2}^x \alpha_2^x + h_{\beta_2}^x \beta_2^x x_i \right) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_3} = \sum_{i=1}^{n} \left[ y_{i3} - \left( h_{a_3}^x \alpha_3^x + h_{\beta_3}^x \beta_3^x x_i \right) \right] x_i = 0,
\]
\[
\frac{\partial s}{\partial \beta_4} = \sum_{i=1}^{n} \left[ y_{i4} - (\alpha_4^x + \beta_4^x x_i) \right] x_i = 0.
\]

4. Parameter Estimation against the Data Deletion Fuzzy Regression Model

For the sake of the evaluation of the impact of the j-th data \((x_j, \tilde{y}_j)\) in regression analysis based on the regression model (3), we can delete the j-th data \((x_j, \tilde{y}_j)\) and detect if these data are an outlier or a strong influence factor, by comparing the changes in statistical inference results. The regression model (1) when the j-th data \((x_j, \tilde{y}_j)\) are deleted is called the data deletion-based regression model, which is represented as

\[
\tilde{y}_{ij} = \tilde{\alpha}_{ij} + \tilde{\beta}_{ij} x_i, \quad j \in \{1, 2, \ldots, n\},
\]

where \(\tilde{\alpha}_{ij}\) and \(\tilde{\beta}_{ij}\) are the two quadrilateral interval type-2 fuzzy parameters.

By the parameter estimation method in Theorem 1, the following results can be drawn for the parameter estimation against the j-th data deleted.

**Theorem 2.** Consider the trapezoidal fuzzy regression model (12) with a set of observation data \((x_i, \tilde{y}_i), i = 1, 2, \ldots, n\), and \(\tilde{y}_i = ((y_{i1}^L, y_{i1}^U), y_{i2}^{(u)}, y_{i2}^{(l)}, y_{i3}^{(u)}, y_{i3}^{(l)}, y_{i4}^{(u)}, y_{i4}^{(l)})\) is a class of quadrilateral interval type-2 fuzzy number. If the j-th data point \((x_j, \tilde{y}_j)\) is deleted, then the following estimates of \(\tilde{\alpha}_{ij}\) and \(\tilde{\beta}_{ij}\) in the Euclidean distance are designed for \(i, j = 1, 2, \ldots, n\).

\[
\tilde{\alpha}_{i1j} = \frac{\sum_{i \neq j} x_i y_{i1j}^L - \sum_{i \neq j} x_i y_{i1j}^U}{(\sum_{i \neq j} x_i^2)^2 - (n-1)\sum_{i \neq j} x_i^2},
\]
\[
\tilde{\alpha}_{i2j} = \frac{\sum_{i \neq j} x_i y_{i2j}^L - \sum_{i \neq j} x_i y_{i2j}^U}{h_{\alpha_2}^L \left( \left( \sum_{i \neq j} x_i^2 \right)^2 - (n-1)\sum_{i \neq j} x_i^2 \right)},
\]
\[
\tilde{\alpha}_{i3j} = \frac{\sum_{i \neq j} x_i y_{i3j}^L - \sum_{i \neq j} x_i y_{i3j}^U}{h_{\alpha_3}^L \left( \left( \sum_{i \neq j} x_i^2 \right)^2 - (n-1)\sum_{i \neq j} x_i^2 \right)},
\]
\[
\tilde{\alpha}_{i4j} = \frac{\sum_{i \neq j} x_i y_{i4j}^L - \sum_{i \neq j} x_i y_{i4j}^U}{h_{\alpha_4}^L \left( \left( \sum_{i \neq j} x_i^2 \right)^2 - (n-1)\sum_{i \neq j} x_i^2 \right)},
\]
\[
\tilde{\beta}_{i1j} = \frac{\sum_{i \neq j} x_i y_{i1j}^L - \sum_{i \neq j} x_i y_{i1j}^U}{(\sum_{i \neq j} x_i^2)^2 - (n-1)\sum_{i \neq j} x_i^2},
\]
\[
\tilde{\beta}_{i2j} = \frac{\sum_{i \neq j} x_i y_{i2j}^L - \sum_{i \neq j} x_i y_{i2j}^U}{h_{\beta_2}^L \left( \left( \sum_{i \neq j} x_i^2 \right)^2 - (n-1)\sum_{i \neq j} x_i^2 \right)},
\]
\[
\tilde{\beta}_{i3j} = \frac{\sum_{i \neq j} x_i y_{i3j}^L - \sum_{i \neq j} x_i y_{i3j}^U}{h_{\beta_3}^L \left( \left( \sum_{i \neq j} x_i^2 \right)^2 - (n-1)\sum_{i \neq j} x_i^2 \right)},
\]
\[
\tilde{\beta}_{i4j} = \frac{\sum_{i \neq j} x_i y_{i4j}^L - \sum_{i \neq j} x_i y_{i4j}^U}{h_{\beta_4}^L \left( \left( \sum_{i \neq j} x_i^2 \right)^2 - (n-1)\sum_{i \neq j} x_i^2 \right)},
\]

where \(\kappa \in \{U, L\}\).

**Proof.** Similar to the case of \(s(\tilde{\alpha}, \tilde{\beta})\) in Theorem 1, after the j-th data \((x_j, \tilde{y}_j)\) are deleted, the sum \(s(\tilde{\alpha}_{ij}, \tilde{\beta}_{ij})\) of the squared Euclidean distance between \(\tilde{y}_i\) and \(\tilde{\alpha}_{ij} + \tilde{\beta}_{ij} x_i\) results in the following, for \(i, j = 1, 2, \ldots, n\):
5. Impact Evaluation Rule for the Data

Outlier Detection

Since \( \bar{y}_{(j)} \) and \( \bar{y} \) introduced above are two type-2 fuzzy numbers, it is inconvenient to compare their difference. For this reason, a suitable statistical measure is usually suggested in order to compare the impact quantitatively. In this paper, we introduce the standard deviation of the regression equation as the statistical measure to analyze the impact of the data deletion. Let us define the standard deviation of the regression equation (3) as Definition 5 in Appendix.

From Definition 5, we know that the standard deviation of the regression equation is actually the average deviation between the observed value and the estimated value. Apparently, the smaller the standard deviation is, the closer the estimated value is to the observed value, as well as the closer the observation points are clustered around the fuzzy regression model.

When calculating the standard error in (A.8), we should firstly obtain the parameters \( \bar{a} \) and \( \bar{\beta} \) by solving the extremum-value problem in the statistical analysis and then estimate the regression value. It will use two rounds of statistical calculations, and thus, two degrees of freedom are taken. Therefore, the denominator in (A.8) uses \( n - 2 \) and not \( n \) in the statistical analysis.

According to the data deletion-based type-2 fuzzy regression model (12), when the \( j \)-th data point is deleted, the corresponding standard deviation is

\[
\sigma_{j} = \left( \frac{1}{n - 3} \sum_{i \neq j} d_{i}^{2}(\bar{y}_{(i,j)}, \bar{y}_{(i,j)}) \right)^{1/2}.
\]

Specifically, the square of \( \sigma_{j} \) for the type-2 fuzzy regression model (12) with fuzzy \( \bar{a} \) and fuzzy \( \bar{\beta} \) can be calculated by

\[
s(\bar{a}_{(j)}, \bar{\beta}_{(j)}) = \frac{a}{n - 3} \cdot \frac{1}{4} \sum_{i \neq j} \left[ \left( \left( y_{(i,j)}^{U} - (a_{(i,j)}^{U} + \beta_{(i,j)}^{U} x_{i}) \right)^{2} \right)^{1/2} \right]^{1/2} \]

(14)

Then, following the steps in the proof of Theorem 1, the results in Theorem 2 can be obtained accordingly. We omit the specific proof for saving space. \( \Box \)

### Table 1: A set of interval type-2 trapezoidal fuzzy numbers.

| \( x_{i} \) | \( \bar{y}_{i} = ((y_{i}^{L}, y_{i}^{U}), y_{i}^{L}, y_{i}^{U}) \) |
|---|---|
| 1 | (1, 38.09, 36.43, 5), (1, 43.09, 41.43, 5) |
| 2 | (1, 62.10, 26.50, 6), (1, 67.10, 31.50, 6) |
| 3 | (1, 63.76, 44.71, 7), (1, 68.76, 49.71, 7) |
| 4 | (1, 74.52, 38.09, 8), (1, 79.52, 43.09, 8) |
| 5 | (1, 75.38, 41.40, 7), (1, 80.38, 46.40, 7) |
| 6 | (2, 52.99, 26.49, 4), (2, 57.99, 31.49, 4) |
| 7 | (2, 62.93, 26.49, 5), (2, 68.93, 31.49, 5) |
| 8 | (2, 72.04, 33.12, 6), (2, 77.04, 38.12, 6) |
| 9 | (2, 76.12, 43.06, 7), (2, 81.12, 47.06, 7) |
| 10 | (2, 90.26, 42.64, 7), (2, 95.26, 47.64, 7) |
| 11 | (3, 85.70, 31.33, 6), (3, 90.70, 36.33, 6) |
| 12 | (3, 95.27, 27.64, 6), (3, 100.27, 32.64, 6) |
| 13 | (3, 105.98, 27.64, 6), (3, 110.98, 32.64, 6) |
| 14 | (3, 79.25, 66.81, 6), (3, 84.25, 71.81, 6) |
| 15 | (3, 120.5, 32.25, 6), (3, 125.5, 37.25, 6) |
where $\hat{\alpha}_j^k$ and $\hat{\beta}_j^k (r = 1, 2, 3, 4 \text{ and } k \in \{U, L\})$ are defined in Theorem 2.

For the data deletion-based fuzzy linear regression model in (12), let $\sigma_j$ be the metric of the impact on the regression model (12). Evidently, if $\sigma_j^2$ increases after deleting the $j$-th data point, then it indicates that the impact is greater and this data point may be an outlier; otherwise, the $j$-th data are normal.

The derived results can be reduced to the case of trapezoidal type-2 fuzzy regression model when $h_1^k = h_2^k = 1$ ($r \in \{a, b\}, \ k \in \{U, L\}$). Besides, it becomes the normal case when using $h_1^k = h_2^k = 1$. Therefore, one can consider $\hat{\alpha}_j^k = \alpha^k$ and $\hat{\beta}_j^k = \beta^k (k \in \{U, L\})$ when dealing with the triangular type-2 fuzzy numbers in practice.

### 6. Simulation Example

In this part, we provide an example to validate the presented fuzzy regression model and the designed impact evaluation rule for the data outlier detection. We borrow a set of data; these are the estimation errors produced in Table 9 from [30] but considered as some type-2 trapezoidal fuzzy numbers. Table 1 gives the considered interval type-2 trapezoidal fuzzy numbers as the observed data.

Based on this set of observed data, we will detect if some of them is an outlier or a strong impact point by using the designed impact evaluation rule. For simplicity, we use the normal trapezoidal type-2 fuzzy number $h_1^k = h_2^k = 1$ ($r \in \{a, b\}, \ k \in \{U, L\}$) for the type-2 fuzzy parameters $\bar{\alpha}$ and $\bar{\beta}$. Considering $\hat{y}_j = \bar{\alpha} + \bar{\beta}x_j$, according to Theorem 1, by setting $a = b = 0.5$, we obtain $\bar{\alpha}$ and $\bar{\beta}$ of the resulting type-2 fuzzy regression equation as follows:

$$\hat{y}_j = ((0.5714, 45.7767, 33.3210, 6.1619), (0.5714, 50.8719, 38.2936, 6.1619)) + ((0.1786, 3.8984, 0.3680, 0.0036), (0.1786, 3.902, 0.3725, 0.0036))x_j.$$

\[ \text{(17)} \]

Then, in the following steps, we can use this fuzzy regression equation to calculate the standard deviation of the regression value after deleting the $j$-th data point. Based on the fuzzy regression model (17), according to Theorem 2, we can obtain the standard deviations as shown in Table 2.

From the standard deviations in Table 2, after deleting the 14-th data point, one can find that the standard deviation of the estimate of the fuzzy regression equation is larger than those of others. This can be observed evidently from Figure 2. Thus, the 14-th data point is likely the outlier point, which means the estimation error obtained in Table 1 is the least-accurate estimation error.
Suppose that Definitions used in this paper are as follows:

Appendix

Definitions used in this paper are as follows:

Definition 1 (see [14, 31]). Suppose that \( F \) is a nonempty fuzzy set and that the expression of type-2 fuzzy numbers is as follows:

\[
F = \{ ((x, u), \mu_F(x, u) \mid \forall x \in X, \forall u \in J_x, 0 \leq \mu_F(x, u) \leq 1 \},
\]

(A.1)
in which \( X \) is the universe of discourse, \( \mu_F \) denotes the type-2 membership function, and \( J_x \subseteq [0, 1] \). The fuzzy numbers can also be represented by

\[
F = \int_{x \in X} \int_{u \in J_x} \mu_F(x, u) \frac{\mu_F(x, u)}{\mu_F(x, u)} (x, u)
\]

(A.2)

Definition 2 (see [14, 31]). For a type-2 fuzzy set \( A \), if for all the type-2 membership functions \( \mu_F(x, u) = 1 \), then \( F \) is called the interval type-2 fuzzy set, which is expressed by

\[
F = \int_{x \in X} \int_{u \in J_x} 1 \frac{\mu_F(x, u)}{\mu_F(x, u)} (x, u),
\]

(A.3)

where \( X \) is the universe of discourse and \( J_x \subseteq [0, 1] \).

Definition 3. If the upper- and lower-bound membership functions of the interval type-2 fuzzy numbers are trapezoidal fuzzy numbers, then they are called trapezoidal interval type-2 fuzzy numbers, expressed by

\[
F = (F^L, F^U) = \left( (a_{1i}^L, a_{1i}^U, a_{2i}^L, a_{2i}^U, a_{3i}^L, a_{3i}^U, a_{4i}^L, a_{4i}^U; h_i^L(F^L), h_i^U(F^U)),
\right.
\]

\[
\left. (a_{1i}^L, a_{1i}^U, a_{2i}^L, a_{2i}^U, a_{3i}^L, a_{3i}^U, a_{4i}^L, a_{4i}^U; h_i^L(F^L), h_i^U(F^U)) \right)
\]

(A.4)

where \( h_i^L(F) \) and \( h_i^U(F) \) denote the membership grades of the second and third elements \( a_2^L \) and \( a_2^U (\kappa \in \{ U, L \}) \) above. If \( h_i^L = h_i^U = 1 \), then the generalized fuzzy number \( F \) is called a normal trapezoidal type-2 fuzzy number denoted \( F = ((a_{1i}^L, a_{1i}^U, a_{2i}^L, a_{2i}^U, a_{3i}^L, a_{3i}^U, a_{4i}^L, a_{4i}^U)) \).

Definition 4. Let \( F_1 \) and \( F_2 \) be expressed as two quadrilateral interval type-2 fuzzy numbers:

\[
F_1 = (F_1^L, F_1^U) = \left( (a_{11}^L, a_{11}^U, a_{12}^L, a_{12}^U, a_{13}^L, a_{13}^U, a_{14}^L, a_{14}^U; h_{11}(F_1^L), h_{11}(F_1^U)) \right),
\]

\[
\left. (a_{11}^L, a_{11}^U, a_{12}^L, a_{12}^U, a_{13}^L, a_{13}^U, a_{14}^L, a_{14}^U; h_{11}(F_1^L), h_{11}(F_1^U)) \right)
\]

(A.5)

Then, the Euclidean distance between \( F_1 \) and \( F_2 \) is defined as

\[
d(F_1, F_2) = a \cdot d^L + b \cdot d^U, \quad a + b = 1,
\]

(A.6)

with \( 0 \leq a \leq 1, 0 \leq b \leq 1 \), and

\[
d^L = \frac{1}{4} \left[ (a_{11}^U - a_{12}^U)^2 + (h_i^L(F_1^L) \cdot a_{12}^L - h_i^L(F_2^L) \cdot a_{22}^L)^2 + (h_i^L(F_1^L) \cdot a_{13}^L - h_i^L(F_2^L) \cdot a_{23}^L)^2 + (a_{14}^U - a_{24}^U)^2 \right]^{1/2},
\]

(A.7)

Definition 5. Consider the observation pair \((x_i, \tilde{y}_i)\) of the regression equation in (3), with crisp \( x_i \) and fuzzy \( \tilde{y}_i \), for \( i = 1, 2, \ldots, n \). Denote \( \tilde{y}_i \) the estimate of \( \tilde{y}_i \). Then, the function

\[
\sigma_j = \left( \frac{1}{n-2} \sum_{i=1}^{n} d_i^2(\tilde{y}_i, \tilde{y}_i) \right)^{1/2}
\]

(A.8)

is called the Euclidean distance-based standard deviation of the fuzzy regression equation, where \( \tilde{y}_i = \alpha + \beta x_i \).
Data Availability
No data were used to support this study.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

References
[1] R. J. Brook, Applied Regression Analysis and Experimental Design, Routledge, Abingdon, UK, 2018.
[2] G. W. Bohrnstedt and T. M. Carter, “Robustness in regression analysis,” Sociological Methodology, vol. 3, pp. 118–146, 1971.
[3] J. Fox, Applied Regression Analysis and Generalized Linear Models, Sage Publications, Thousand Oaks, CA, USA, 2015.
[4] K. Yao and B. Liu, “Uncertain regression analysis: an approach for imprecise observations,” in Soft Computing, Springer, Berlin, Germany, 2018.
[5] K. Asai, H. Tanaka, and S. Uegima, “Linear regression analysis with fuzzy model,” IEEE Transactions on Systems, Man and Cybernetics, vol. 12, no. 6, pp. 903–907, 1982.
[6] H. Tanaka, “Fuzzy data analysis by possibilistic linear models,” Fuzzy Sets and Systems, vol. 24, no. 3, pp. 363–375, 1987.
[7] H. Tanaka, I. Hayashi, and J. Watada, “Possibilistic linear regression analysis for fuzzy data,” European Journal of Operational Research, vol. 40, no. 3, pp. 389–396, 1989.
[8] L. A. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, no. 3, pp. 338–353, 1965.
[9] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility,” Fuzzy Sets and Systems, vol. 100, pp. 9–34, 1999.
[10] J. M. Mendel, Uncertain Rule-Based Fuzzy Systems, Springer, Berlin, Germany, 2017.
[11] J. Liu, Y. Gao, W. Luo, and L. Wu, “Takagi–Sugeno fuzzy-model-based control of three-phase AC/DC voltage source converters using adaptive sliding mode technique,” IET Control Theory & Applications, vol. 11, no. 8, pp. 1255–1265, 2016.
[12] Y. Gao, F. Xiao, J. Liu, and R. Wang, “Distributed soft fault detection for interval type-2 fuzzy-model-based stochastic systems with wireless sensor networks,” IEEE Transactions on Industrial Informatics, vol. 15, no. 1, pp. 334–347, 2018.
[13] K. T. Atanassov, “Interval valued intuitionistic fuzzy sets,” in Intuitionistic Fuzzy Sets, pp. 139–177, Springer, Berlin, Germany, 1999.
[14] J. M. Mendel, R. I. John, and F. Liu, “Interval type-2 fuzzy logic systems made simple,” IEEE Transactions on Fuzzy Systems, vol. 14, no. 6, pp. 808–821, 2006.
[15] A. Bissierer, R. Boukezzoulou, and S. Galichet, “A revisited approach to linear fuzzy regression using trapezoidal fuzzy intervals,” Information Sciences, vol. 180, no. 19, pp. 3653–3673, 2010.
[16] J. Nowakovz and M. Pokorný, “Fuzzy linear regression analysis,” IFAC Proceedings Volumes, vol. 46, no. 28, pp. 245–249, 2013.
[17] S. Roychowdhury and W. Pedrycz, “Modeling temporal functions with granular regression and fuzzy rules,” Fuzzy Sets and Systems, vol. 126, no. 3, pp. 377–387, 2002.
[18] J. Chachi and M. Roozbeh, “A fuzzy robust regression approach applied to bedload transport data,” Communications in Statistics-Simulation and Computation, vol. 46, no. 3, pp. 1703–1714, 2017.
[19] P. D’Urso, “Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data,” Computational Statistics & Data Analysis, vol. 42, no. 1-2, pp. 47–72, 2003.
[20] H.-C. Wu, “Fuzzy estimates of regression parameters in linear regression models for imprecise input and output data,” Computational Statistics & Data Analysis, vol. 42, no. 1-2, pp. 203–217, 2003.
[21] J. Chachi, “A weighted least-squares fuzzy regression for crisp input-fuzzy output data,” IEEE Transactions on Fuzzy Systems, vol. 27, no. 4, pp. 739–748, 2018.
[22] Y.-H. O. Chang and B. M. Ayyub, “Fuzzy regression methods-comparative assessment,” Fuzzy Sets and Systems, vol. 119, no. 2, pp. 187–203, 2001.
[23] A. Arabpour and M. Tata, “Estimating the parameters of a fuzzy linear regression model,” Iranian Journal of Fuzzy Systems, vol. 5, no. 2, pp. 1–19, 2008.
[24] 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, vol. 17, no. 6, pp. 1259–1272, 2009.
[25] M. Černý and M. Hladík, “Possibilistic linear regression with fuzzy data: tolerance approach with prior information,” Fuzzy Sets and Systems, vol. 340, pp. 127–144, 2018.
[26] M. Soltani, A. J. Telmoudi, L. Chaouech, M. Ali, and A. Chaari, “Design of a robust interval-valued type-2 fuzzy c-regression model for a nonlinear system with noise and outliers,” Soft Computing, vol. 23, no. 15, pp. 6125–6134, 2018.
[27] W.-L. Hung and M.-S. Yang, “An omission approach for detecting outliers in fuzzy regression models,” Fuzzy Sets and Systems, vol. 157, no. 23, pp. 3109–3122, 2006.
[28] K. Y. Chan, C. K. Kwong, and T. C. Fogarty, “Modeling manufacturing processes using a genetic programming-based fuzzy regression with detection of outliers,” Information Sciences, vol. 180, no. 4, pp. 506–518, 2010.
[29] L.-H. Lee and S.-M. Chen, “A new method for fuzzy multiple attributes group decision-making based on the arithmetic operations of interval type-2 fuzzy sets,” in Proceedings of the 2008 International Conference on Machine Learning and Cybernetics, vol. 6, pp. 3084–3089, IEEE, Kunming, China, July 2008.
[30] S.-P. Chen and J.-F. Dang, “A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy,” Information Sciences, vol. 178, no. 20, pp. 3973–3988, 2008.
[31] L. Abdulllah, C. W. R. Adawiyah, and C. W. Kamal, “A decision making method based on interval type-2 fuzzy sets: an approach for ambulance location preference,” Applied Computing and Informatics, vol. 14, no. 1, pp. 65–72, 2018.
