On Cluster-Wise Fuzzy Regression Analysis

Miin-Shen Yang and Cheng-Hsiu Ko

Abstract—Since Tanaka et al. [19] proposed a study in linear regression analysis with fuzzy model, the fuzzy regression analysis has been widely studied and applied in a variety of substantive areas. We know that the regression analysis in the case of heterogeneity of observations is commonly presented in practice. In this paper, the main goal is to apply fuzzy clustering techniques to fuzzy regression analysis. The fuzzy clustering is used to overcome the heterogeneous problem in the fuzzy regression model. We present the cluster-wise fuzzy regression analysis in two approaches: the two-stage weighted fuzzy regression and the one-stage generalized fuzzy regression. The two-stage procedure extends the results of Jajuga [12] and Diamond [8]. The one-stage is created by embedding fuzzy clusterings into the fuzzy regression model fitting at each step of procedure. This kind of embedding in the one-stage procedure shall be more effective since the structure of regression-line shape encountered in the data set is taken into account at each iteration of algorithm. Numerical results give evidence that the one-stage procedure can be highly recommended to be of use in the cluster-wise fuzzy regression analysis.

I. INTRODUCTION

Regression analysis is used in evaluating the functional relationship between the dependent and independent variables and also in determining the best-fit model for describing the relationship. In the usual conventional model, deviations between the observed values and the estimated values are supposed to be due to measurement errors or random variations. Therefore, the statistical techniques are applied for estimation and inference in regression analysis. But sometimes the deviations are due to the imprecise observed data or the indefiniteness of the system structure. In this case, the uncertainty is not due to randomness but fuzziness. Regression analysis on fuzzy data in dealing with fuzziness is usually called fuzzy regression analysis. Tanaka et al. [19] first proposed this study in linear regression analysis with the fuzzy model. They considered the linear fuzzy regression model

\[ Y = a + bX, \quad a, b \in \mathbb{R}, \quad X \in \mathcal{F}_T(\mathbb{R}) \]

and

\[ Y = E + bX, \quad b \in \mathbb{R}, \quad E, X \in \mathcal{F}_T(\mathbb{R}). \]

The corresponding least-square optimization problems are

\[
\begin{align*}
\text{minimize } r(a, b) &= \sum_{i=1}^{n} d^2(Y_i, a + bX_i); \\
\text{minimize } r(E, b) &= \sum_{i=1}^{n} d^2(Y_i, E + bX_i); \\
\text{minimize } r(A, b) &= \sum_{i=1}^{n} d^2(Y_i, A + bX_i).
\end{align*}
\]

A generalization of the Tanaka approach for the general form of regression equations about \(LR\)-type fuzzy numbers is developed by Bárdozy [2].

We note that the Tanaka approach is quite complicated in solving the optimization problem. It is unclear what the relation is to a least-square concept. The measure of best-fit by residuals is not presented in the Tanaka approach. Therefore, Diamond [8] proposed the so-called fuzzy least squares. Based on a metric \(d_f\) on the space \(\mathcal{F}(\mathbb{R})\) of all normalized fuzzy numbers which was defined by Puri and Ralescu [15], Diamond gave a metric \(d\) on the space \(\mathcal{F}_T(\mathbb{R})\) of all triangular fuzzy numbers by

\[
d^2(X, Y) = (m_x - m_y)^2 + ((m_x - \alpha_x) - (m_y - \alpha_y))^2 + ((m_x + \beta_x) - (m_y - \beta_y))^2
\]

where \(X = (m_x, \alpha_x, \beta_x)_T\) and \(Y = (m_y, \alpha_y, \beta_y)_T\) are any two triangular fuzzy numbers in \(\mathcal{F}_T(\mathbb{R})\). There are three simple fuzzy regression models considered in Diamond [8].

1. Fuzzy input and fuzzy output

\[ Y = a + bX, \quad a, b \in \mathbb{R}, \quad X \in \mathcal{F}_T(\mathbb{R}) \]

and

\[ Y = E + bX, \quad b \in \mathbb{R}, \quad E, X \in \mathcal{F}_T(\mathbb{R}). \]

2. Numerical input and fuzzy output

\[ Y = A + Bx, \quad x \in \mathbb{R}, \quad A, B \in \mathcal{F}_T(\mathbb{R}). \]

The corresponding least-square optimization problems are

\[
\begin{align*}
\text{minimize } r(a, b) &= \sum_{i=1}^{n} d^2(Y_i, a + bX_i); \\
\text{minimize } r(E, b) &= \sum_{i=1}^{n} d^2(Y_i, E + bX_i); \\
\text{minimize } r(A, B) &= \sum_{i=1}^{n} d^2(Y_i, A + Bx_i).
\end{align*}
\]

Fuzzy clustering procedures have been widely applied in a diverse areas (see Yang, [23]). In the literature of fuzzy clustering, the fuzzy \(c\)-means (FCM) are the well known and powerful clustering algorithms. See for example, Dunn [10], Bezdek [3], Davenport et al. [6], Chan et al. [4], and Yang...
[21], [22]. But most of fuzzy clustering techniques are used for the crisp data. Recently, Yang and Ko [24] proposed a new type of fuzzy clustering procedures called the fuzzy $c$-numbers (FCN). These FCN clustering procedures are used in clustering the fuzzy data. Yang and Ko define a metric $d_{LR}$ on the space $F_{LR}(\mathcal{R})$ of all LR-type fuzzy numbers by

$$d_{LR}^2(X,Y) = (m_x - m_y)^2 + ((m_x - l_\alpha x) - (m_y - l_\alpha y))^2$$

$$+ ((m_x + r_\beta x) - (m_y + r_\beta y))^2$$

where $l = \int_0^1 L^{-1}(u)du$, $r = \int_0^1 R^{-1}(u)du$, and $X = (m_x, m_\alpha x, m_\beta x)_{LR}$ and $Y = (m_y, m_\alpha y, m_\beta y)_{LR}$ are LR-type fuzzy numbers in $F_{LR}(\mathcal{R})$.

We know that the regression analysis in the case of heterogeneity of observations is commonly presented in practice. In this paper, we embed the techniques of fuzzy clustering into fuzzy regression analysis. The techniques of fuzzy clustering are used to overcome the heterogeneous problem in the fuzzy regression model. We present this kind of embedding through two approaches. The first is a two-stage procedure. We first choose a fuzzy clustering procedure to get the class memberships of observations. Then we use these values of memberships as weights and consider a weighted fuzzy least-square optimization problem of the fuzzy regression model based on the metric $d_{LR}$. These results are presented in Section II. Some numerical comparisons with Diamond [8] are also presented. We note that the above two-stage procedure is easy to handle, but depends heavily on the chosen fuzzy clustering procedure. If we combine fuzzy clustering and fuzzy regression analysis into a one-stage procedure, we actually embed the fuzzy clustering into the fuzzy regression model fitting at each iteration of the procedure. This kind of embedding in the one-stage procedure should be more useful and effective in the model fitting than the two-stage procedure since it effectively takes into account the structure of clusters encountered in the data set at each iteration of algorithm. Section III describes this kind of one-stage procedure. Finally, some numerical comparisons and conclusions are presented in Section IV.

II. TWO-STAGE WEIGHTED FUZZY REGRESSION ANALYSIS

Regression analysis is used in the model-fitting of observations. The heterogeneous problem in the regression model is usually difficult to be handled. But the heterogeneity of observations is commonly presented in practice. We may think of the heterogeneity of observations because of different clusters of observations. If we first cluster the observations and then use their class memberships as the weights in the weighted least-squares estimation, it enables us to overcome the heterogeneous problem in the regression model fitting. Based on this kind of idea, we present the two-stage weighted fuzzy regression analysis in this section. We first choose a suitable fuzzy clustering procedure to get the class memberships of observations as weights. Then we consider the optimization problem of weighted fuzzy least-squares. We call it a two-stage procedure. We note that the conventional weighted least squares become the special case. In order to consider the weighted fuzzy linear regression for LR-type fuzzy numbers, it is necessary to have the following definitions:

**Definition 1**: (LR-type fuzzy numbers, see Zimmermann [25, pp. 62–63]). Let $L$ and $R$ be decreasing, shape functions from $\mathcal{R}$ to $[0,1]$ with $L(0) = 1; L(x) < 1$ for all $x > 0$; $L(x) > 0$ for all $x < 0$; $L(1) = 0$ or $L(x) > 0$ for all $x$ and $L(+\infty) = 0$. Then a fuzzy number $X$ is called of LR-type if for $m, \alpha, \beta > 0, \beta > 0$ in $\mathcal{R}$

$$X(x) = \left\{ \begin{array}{ll} L(\frac{m-x}{\alpha}) & \text{for } x \leq m \\ R(\frac{m-x}{\beta}) & \text{for } x \geq m \end{array} \right.$$
Thus, \( \mathcal{B} \) is a metric. Next, we claim that \( \mathcal{B} \) is complete as follows: Let \( \{x_n\}_{n \geq 1} \) be a Cauchy sequence in \( \mathcal{B} \) where \( \{x_n\}_{n \geq 1} \) is a Cauchy sequence in \( \mathcal{B} \). Then \( \lim_{n \to \infty} x_n = x \) as \( n \to \infty \). Similarly, \( \lim_{n \to \infty} \beta_n = \beta \) as \( n \to \infty \). That is, \( \{x_1, \ldots, x_n\}_{n \geq 1} \) is a Cauchy sequence in \( \mathcal{B} \).

Let us first choose a suitable fuzzy clustering procedure to get a good fuzzy \( \alpha \)-partition \( \{x_i\}_{i=1}^c \) of \( \mathcal{B} \) where the \( \alpha \)-component \( \beta_{x_i} \) represents the membership of \( x_i \) belonging to \( i \)th class. If we fit the data set in the sense of best fit with respect to the metric \( \mathcal{B} \), then the corresponding weighted fuzzy least-square optimization problem to the model (M1) is

\[
\minimize_i \left[ \sum_{j=1}^c u_{ij}^2 \right] \tag{F1}
\]

where \( u_{ij} \) is the index of fuzziness and \( a_{0i} \) and \( a_{1ij} \) are unknown coefficients and

\[
\sum_{j=1}^c u_{ij} = 1, \quad i = 1, \ldots, n,
\]

where \( a_0 \) and \( a_{1ij} \in \mathbb{R} \) are unknown coefficients.

A. Fuzzy Input and Fuzzy Output

Let \( \mathcal{G} \) be the set of observations \( (X_j, Y_j), j = 1, \ldots, n \) where \( X_j = (m_{x_j}, \alpha_{x_j}, \beta_{x_j})_R \) and \( Y_j = (m_{y_j}, \alpha_{y_j}, \beta_{y_j})_R \) are \( LR \)-type fuzzy numbers. Suppose these observations are heterogeneous and come from \( c \) clusters. Of course, if \( c = 1 \) then the observations are homogeneous. This kind of discussion is also presented in Cutsem and Gath [5]. Now we want to fit a data set to the cluster-wise fuzzy linear regression model

\[
Y_j = a_0 + a_{1i} X_j, \quad i = 1, \ldots, c, \quad j = 1, \ldots, n,
\]

where \( a_0 \) and \( a_{1ij} \in \mathbb{R} \) are unknown coefficients.

A partition of \( \mathcal{G} \) into \( c \) parts can be represented by mutually disjoint sets \( G_1, \ldots, G_c \) such that \( G_1 \cup \cdots \cup G_c = \mathcal{G} \) or equivalently by the indicator functions \( u_1, \ldots, u_c \) such that \( u_i(Z) = 1 \) if \( Z \in G_i \) and \( u_i(Z) = 0 \) if \( Z \notin G_i \) for all \( Z \in \mathcal{G} \) and for all \( i = 1, \ldots, c \). Here we call \( u = (u_1, \ldots, u_c) \) a hard \( c \)-partition of \( \mathcal{G} \). Ruspiní [17] introduced a fuzzy \( c \)-partition \( u = (u_1, \ldots, u_c) \) of \( \mathcal{G} \) by an extension to allow \( u_i(Z) \) to represent the membership of \( Z \) in \( i \)th class assuming values in the interval \([0, 1]\) such that \( \sum_{i=1}^c u_i(Z) = 1 \) for all \( Z \in \mathcal{G} \).

Let us first choose a suitable fuzzy clustering procedure to get a good fuzzy \( c \)-partition \( u = (u_1, \ldots, u_c) \) of \( \mathcal{G} \) where the \( j \)th component \( u_{1j} \) of \( u_i \) represents the membership of \( j \)th data point \( (X_j, Y_j) \) belonging to \( i \)th class. If we fit the data set in the sense of best fit with respect to the metric \( \mathcal{B} \), then the corresponding weighted fuzzy least-square optimization problem to the model (M1) is

\[
\minimize_i \left[ \sum_{j=1}^n u_{ij}^2 \right] \tag{F1}
\]

where \( m \geq 1 \) is the index of fuzziness and \( a_{0i} \) and \( a_{1ij} \) are unknown coefficients and

\[
\sum_{j=1}^n u_{ij} = 1, \quad i = 1, \ldots, c, \quad j = 1, \ldots, n,
\]

where \( a_0 \) and \( a_{1ij} \in \mathbb{R} \) are unknown coefficients.
We mention that the result here in the case of no clustering (i.e., $c = 1$) is similar to those of Diamond [8].

Now we see another special case when data are crisp. That is, $X$ and $Y$ have no spreads (i.e., $\alpha_x = \beta_x = \alpha_y = \beta_y = 0$). Then we have the following estimators $(\hat{\alpha}_0, \hat{\alpha}_1)$ of $(\alpha_0, \alpha_1)$ (see (9) and (10) at the bottom of page). The result here in the case of crisp data is similar to the formula which was obtained by Jajuga [12]. Next, we present some basic properties of cluster-wise fuzzy linear regression which are analogous to those of classical regression.

Define 3: (Mean value for LR-type fuzzy numbers).
The mean value of variables $Y = (m_y, \alpha_y, \beta_y)_L$ for the $i$th fuzzy cluster is defined as

$$E_i(Y) = \frac{\sum_{j=1}^{n} w_{ij} m_{yj} + (m_{yj} - \alpha_{yj}) + (m_{yj} + \beta_{yj})}{3\sum_{j=1}^{n} w_{ij}}$$

If $\alpha_{yj} = \beta_{yj} = 0$, then $E_i(Y) = \frac{\sum_{j=1}^{n} w_{ij} m_{yj}}{\sum_{j=1}^{n} w_{ij}}$.

Definition 4: (Variance for LR-type fuzzy numbers). The variance of variables $Y = (m_y, \alpha_y, \beta_y)_L$ for the $i$th fuzzy cluster is defined as

$$V_i(Y) = (E_i(Y) - E_i(Y))^2 = E_i(Y)^2 - E_i^2(Y)$$

$$= \frac{\sum_{j=1}^{n} w_{ij}^2 (m_{yj}^2 + m_{yj}^2 + m_{yj}^2)}{3\sum_{j=1}^{n} w_{ij}^2} - \left(\frac{\sum_{j=1}^{n} w_{ij} m_{yj}}{3\sum_{j=1}^{n} w_{ij}}\right)^2$$

If $\alpha_{yj} = \beta_{yj} = 0$, then $V_i(Y) = \frac{\sum_{j=1}^{n} w_{ij}^2 m_{yj}^2}{\sum_{j=1}^{n} w_{ij}} - \left(\frac{\sum_{j=1}^{n} w_{ij} m_{yj}}{\sum_{j=1}^{n} w_{ij}}\right)^2$. 

\begin{align*}
\hat{\alpha}_0 = & \quad \frac{\sum_{j=1}^{n} w_{ij} m_{yj} \sum_{j=1}^{n} u_{ij}^2 m_{xj} - \sum_{j=1}^{n} w_{ij} u_{ij}^2 m_{xj} \sum_{j=1}^{n} m_{yj} u_{ij} r_{ij}}{3\sum_{j=1}^{n} w_{ij}^2 m_{xj}^2 - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}, \quad i = 1, \ldots, c \\
\hat{\alpha}_1 = & \quad \frac{3\sum_{j=1}^{n} w_{ij}^2 m_{yj} \sum_{j=1}^{n} u_{ij}^2 m_{xj} - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}{3\sum_{j=1}^{n} w_{ij}^2 m_{xj}^2 - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}, \quad i = 1, \ldots, c
\end{align*}

\begin{align*}
\hat{\alpha}_0 = & \quad \frac{\sum_{j=1}^{n} w_{ij} m_{yj} \sum_{j=1}^{n} u_{ij}^2 m_{xj} - \sum_{j=1}^{n} w_{ij} u_{ij}^2 m_{xj} \sum_{j=1}^{n} u_{ij}^2 m_{yj} u_{ij} r_{ij}}{3\sum_{j=1}^{n} w_{ij}^2 m_{xj}^2 - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}, \quad i = 1, \ldots, c \\
\hat{\alpha}_1 = & \quad \frac{3\sum_{j=1}^{n} w_{ij}^2 m_{yj} \sum_{j=1}^{n} u_{ij}^2 m_{xj} - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}{3\sum_{j=1}^{n} w_{ij}^2 m_{xj}^2 - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}, \quad i = 1, \ldots, c
\end{align*}

\begin{align*}
\hat{\alpha}_0 = & \quad \frac{\sum_{j=1}^{n} w_{ij} m_{yj} \sum_{j=1}^{n} u_{ij}^2 m_{xj} - \sum_{j=1}^{n} w_{ij} u_{ij}^2 m_{xj} \sum_{j=1}^{n} u_{ij}^2 m_{yj} u_{ij} r_{ij}}{3\sum_{j=1}^{n} w_{ij}^2 m_{xj}^2 - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}, \quad i = 1, \ldots, c \\
\hat{\alpha}_1 = & \quad \frac{3\sum_{j=1}^{n} w_{ij}^2 m_{yj} \sum_{j=1}^{n} u_{ij}^2 m_{xj} - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}{3\sum_{j=1}^{n} w_{ij}^2 m_{xj}^2 - \left(\sum_{j=1}^{n} w_{ij} m_{xj}\right)^2}, \quad i = 1, \ldots, c
\end{align*}
Definition 5: (Covariance for LR-type fuzzy numbers). The covariance of variables $Y = (m_y, \alpha_y, \beta_y)_LR$ and $X = (m_x, \alpha_x, \beta_x)_LR$ for the $i$th fuzzy cluster is defined as

$$C_i(Y, X) = E(Y - E(Y))(X - E(X)) = E(Y - Y_i)(X - E(X)) = \frac{\sum_{j=1}^{n} u_{ij}^T (m_{xj} m_{yj} + m_{yj} m_{xj} + m_{rxj} m_{ryj})}{3 \sum_{j=1}^{n} u_{ij}^T} - \frac{\sum_{j=1}^{n} u_{ij}^T m_{yj}}{3 \sum_{j=1}^{n} u_{ij}^T} \cdot \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj}}{3 \sum_{j=1}^{n} u_{ij}^T}$$

$$= \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj} m_{yj} - \sum_{j=1}^{n} u_{ij}^T m_{yj} \sum_{j=1}^{n} u_{ij}^T m_{xj}}{3 \sum_{j=1}^{n} u_{ij}^T} \cdot \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj}}{3 \sum_{j=1}^{n} u_{ij}^T} = \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj} m_{yj} - \sum_{j=1}^{n} u_{ij}^T m_{xj} \sum_{j=1}^{n} u_{ij}^T m_{yj}}{3 \sum_{j=1}^{n} u_{ij}^T} \cdot \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj}}{3 \sum_{j=1}^{n} u_{ij}^T}.$$

If $\alpha_{xj} = \beta_{xj} = 0$ (i.e., $X \in \mathbb{R}$), then

$$C_i(Y, X) = \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj} m_{yj}}{3 \sum_{j=1}^{n} u_{ij}^T} - \frac{\sum_{j=1}^{n} u_{ij}^T m_{yj}}{3 \sum_{j=1}^{n} u_{ij}^T} \cdot \frac{\sum_{j=1}^{n} u_{ij}^T m_{xj}}{3 \sum_{j=1}^{n} u_{ij}^T}.$$

Definition 6: (Determination coefficient for LR-type fuzzy numbers). The determination coefficient of variables $Y = (m_y, \alpha_y, \beta_y)_LR$ and $X = (m_x, \alpha_x, \beta_x)_LR$ for the $i$th fuzzy cluster is defined as

$$R_i^2 = \left[ \frac{C_i(Y, X)}{V_i(Y)} \right]^2.$$

Proposition 2: For the model (M1) $Y_j = a_{0i} + a_{1i} X_j$, $j = 1, \ldots, c_i$, $i = 1, \ldots, \bar{c}$, we have that $a_{1i} = \frac{C(Y_i; X)}{V_i(Y)}$ and $a_{0i} = E_i(Y) - a_{1i} E_i(X)$. See proof at the bottom of the page.

We note that Definitions 5–6 and Proposition 2 are the extension of results in Jajuga [12] to the LR-type fuzzy numbers. That is, if we have $X_j$ and $Y_j$ crisp (i.e., $\alpha_x = \beta_x = 0$) then these definitions and proposition are the same as Jajuga [12].

In cluster analysis the decision of number $c$ of clusters is quite important but difficult. This kind of problem is usually called cluster validity. Many different kinds of criteria for cluster validity had been proposed in the literature (see Yang [23]). Here we adopt the suggestion from Jajuga [12]. We determine the number $c$ of clusters by choosing that value for which $R^2(c)$ achieves its maximum where $R^2(c)$ is the weighted mean of the determination coefficients $R_i^2$, $i = 1, \ldots, c$. It is of the form $R^2(c) = \sum_{i=1}^{c} w_i R_i^2$ where $w_i = \frac{\sum_{j=1}^{n} u_{ij}^T}{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^T}$. We note that $R^2(c)$ shall be a good validity criterion in the cluster-wise regression analysis since the determination coefficients $R_i^2$ are commonly used as the measure of goodness in the regression model fitting.

Next, let us consider another model

$$(M2) \quad Y_j = E_i \oplus a_{1i} X_j, \quad j = 1, \ldots, n,$$
where \( a_i \in \mathbb{R} \) and \( E_i = (m_{ei}, \alpha_{ei}, \beta_{ei}) \) \( \in F_{LR}(\mathbb{R}) \) are unknown coefficients.

Clearly, (M2) generalizes (M1). We also consider to fit to the data set in the sense of best fit with respect to the metric \( d_{LR} \). Then the corresponding weighted fuzzy least-squares optimization problem is

\[
\text{(F2) minimize} \quad r(E_i, a_i) = \sum_{j=1}^{n} \sum_{c=1}^{c} u_{ij}^m d_{LR}(E_i \oplus a_i X_j, Y_j)
\]

\[
= \sum_{j=1}^{n} \sum_{c=1}^{c} u_{ij}^m [(m_{ei} + a_i m_{xz_j} - m_{yj})^2 + s_i ((m_{ei} + a_i m_{xz_j} - \alpha_{ei} - m_{yj})^2 + (m_{ei} + a_i m_{xz_j} + r/\beta_{ei} - m_{yj})^2) + (1 - s_i) ((m_{ei} + a_i m_{xz_j} - \alpha_{ei} - m_{yj})^2 + (m_{ei} + a_i m_{xz_j} + r/\beta_{ei} - m_{yj})^2)]
\]

where \( m \geq 1 \) and \( u = (u_1, \ldots, u_c) \) is a fuzzy \( c \)-partition obtained by a chosen fuzzy clustering procedure and \( a_i \) and \( E_i \) are unknown coefficients and \( s_i = 1 \) if \( a_i \geq 0 \) and \( s_i = 0 \) if \( a_i < 0 \).

If \( a_i \geq 0 \) then the necessary conditions for a minimizers \((\hat{E}_i, \hat{a}_i)\) of \( r(E_i, a_i) \) are:

\[
\hat{m}_{ei} = \frac{\sum_{j=1}^{n} u_{ij}^m (m_{yj} - \hat{\alpha}_i m_{xz_j} + \hat{\beta}_i m_{yj})}{\sum_{j=1}^{n} u_{ij}^m}, \quad i = 1, \ldots, c \quad (11)
\]

\[
\hat{\alpha}_{ei} = \frac{\sum_{j=1}^{n} u_{ij}^m (\hat{m}_{ei} + \hat{\alpha}_i m_{xz_j} - m_{yj})}{\sum_{j=1}^{n} u_{ij}^m}, \quad i = 1, \ldots, c \quad (12)
\]

\[
\hat{\beta}_{ei} = \frac{\sum_{j=1}^{n} u_{ij}^m (m_{yj} - \hat{m}_{ei} - \hat{\alpha}_i m_{xz_j})}{\sum_{j=1}^{n} u_{ij}^m}, \quad i = 1, \ldots, c \quad (13)
\]

\[
\hat{\alpha}_i = \frac{\sum_{j=1}^{n} u_{ij}^m (m_{yj} - \hat{m}_{ei} - \hat{\alpha}_i m_{xz_j}) + \hat{\alpha}_i m_{xz_j} - r/\hat{\beta}_i m_{xz_j})}{\sum_{j=1}^{n} u_{ij}^m m_{xz_j}^2}, \quad i = 1, \ldots, c \quad (14)
\]

On the other hand, if \( a_i < 0 \), we have the following necessary conditions

\[
\hat{m}_{ei} = \frac{\sum_{j=1}^{n} u_{ij}^m (m_{yj} - \hat{\alpha}_i m_{xz_j} - \hat{\beta}_i m_{yj})}{\sum_{j=1}^{n} u_{ij}^m}, \quad i = 1, \ldots, c \quad (15)
\]

\[
\hat{\alpha}_{ei} = \frac{\sum_{j=1}^{n} u_{ij}^m (\hat{m}_{ei} - \hat{\alpha}_i m_{xz_j} - m_{yj})}{\sum_{j=1}^{n} u_{ij}^m}, \quad i = 1, \ldots, c \quad (16)
\]

\[
\hat{\beta}_{ei} = \frac{\sum_{j=1}^{n} u_{ij}^m (m_{yj} - \hat{m}_{ei} - \hat{\alpha}_i m_{xz_j})}{\sum_{j=1}^{n} u_{ij}^m m_{xz_j}^2}, \quad i = 1, \ldots, c \quad (17)
\]

Note that if \( \hat{\alpha}_{ei} < 0 \) and \( \hat{\beta}_{ei} < 0 \), then we have that \( \hat{E}_i = (\hat{m}_{ei}, -\hat{\alpha}_{ei}, -\hat{\beta}_{ei}) \) \( \in F_{LR}(\mathbb{R}) \). If \( \hat{\alpha}_{ei} > 0 \) and \( \hat{\beta}_{ei} < 0 \), then \( \hat{E}_i = (\hat{m}_{ei}, 0, -\hat{\beta}_{ei}) \) \( \in F_{LR}(\mathbb{R}) \) when \( |\hat{\beta}_{ei}| > |\hat{\alpha}_{ei}| \); and \( \hat{E}_i = (\hat{m}_{ei}, 0, -\hat{\beta}_{ei}) \) \( \in F_{LR}(\mathbb{R}) \) when \( |\hat{\beta}_{ei}| < |\hat{\alpha}_{ei}| \). If \( \hat{\alpha}_{ei} > 0 \) and \( \hat{\beta}_{ei} > 0 \), then \( \hat{E}_i = (\hat{m}_{ei}, \hat{\alpha}_{ei}, 0) \) \( \in F_{LR}(\mathbb{R}) \) when \( |\hat{\beta}_{ei}| > |\hat{\alpha}_{ei}| \); and \( \hat{E}_i = (\hat{m}_{ei}, -\hat{\beta}_{ei}, 0) \) \( \in F_{LR}(\mathbb{R}) \) when \( |\hat{\beta}_{ei}| < |\hat{\alpha}_{ei}| \).

Based on these necessary conditions, we can construct an iterative algorithm of the two-stage weighted fuzzy regression analysis for the model (M2) as follows.

**Algorithm 1:** (of the two-stage weighted fuzzy regression for the model (M2))


**TABLE II**

<table>
<thead>
<tr>
<th>Two-stage (FCN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>prediction ( Y^1 = 11.471 - 0.159X )</td>
<td>( Y^1 = -2.793 + 0.897X )</td>
</tr>
<tr>
<td>residual ( R^2 = 2.540 + 10.107 )</td>
<td>( R^2 = 53.242 + 74.560 )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>( R^2 = 0.025201 )</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>Two-stage (FCN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>prediction ( Y^1 = 16.802 - 0.223X )</td>
<td>( Y^1 = -0.875 + 0.778X )</td>
</tr>
<tr>
<td>( Y^2 = 4.829 + 0.090X )</td>
<td>( Y^2 = 22.887 - 1.112X )</td>
</tr>
<tr>
<td>residual ( R^2 = 227.26968 )</td>
<td>( R^2 = 62.89032 )</td>
</tr>
<tr>
<td>residual ( R^2 = 268.66066 )</td>
<td>( R^2 = 105.76119 )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>( R^2 = 0.012809 )</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Two-stage (FCN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>prediction ( Y^1 = 17.534 - 0.181X )</td>
<td>( Y^1 = -0.989 + 0.736X )</td>
</tr>
<tr>
<td>( Y^2 = 4.319 + 0.062X )</td>
<td>( Y^2 = 22.918 - 1.101X )</td>
</tr>
<tr>
<td>( Y^3 = 9.750 + 0.046X )</td>
<td>( Y^3 = 1.926 + 0.286X )</td>
</tr>
<tr>
<td>residual ( R^2 = 62.409358 )</td>
<td>( R^2 = 65.99791 )</td>
</tr>
<tr>
<td>residual ( R^2 = 105.43128 )</td>
<td>( R^2 = 13.35580 )</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>( R^2 = 0.001136 )</td>
</tr>
</tbody>
</table>

(c)

(S1) Fix \( m \geq 1 \); give \( c \in \{2, \ldots, n-1\} \); and fix any \( \varepsilon > 0 \).

Choose initials \((\hat{m}_i^0, \hat{c}_i^0, \hat{d}_i^0)\), \( i = 1, \ldots, c \).

(S2) Calculate \((\hat{m}_i^1, \hat{c}_i^1, \hat{d}_i^1)\) by using (11)–(14) if \( \hat{d}_i^0 \geq 0 \); using (15)–(18) if \( \hat{d}_i^0 < 0 \), \( i = 1, \ldots, c \).

(S3) If \( \hat{\beta}_e(\hat{\gamma}_e) < 0 \) and \( \hat{\beta}(\hat{\gamma}) > 0 \), then \( \hat{\beta}_e(\hat{\gamma}_e) = -\hat{\gamma}_e(\hat{\gamma}) \). If \( \hat{\beta}_e(\hat{\gamma}_e) < 0 \), and if \( \hat{\beta}_e(\hat{\gamma}_e) < 0 \), then \( \hat{\beta}_e(\hat{\gamma}_e) = 0 \) and \( \hat{\gamma}_e(\hat{\gamma}) = \max \{\hat{\gamma}_e(\hat{\gamma}) \} \), else \( \hat{\beta}_e(\hat{\gamma}_e) = 0 \) and \( \hat{\gamma}_e(\hat{\gamma}) = \max \{\hat{\gamma}_e(\hat{\gamma}) \} \).

(S4) Compare \((\hat{m}_i^1, \hat{c}_i^1, \hat{d}_i^1)\) to \((\hat{m}_i^0, \hat{c}_i^0, \hat{d}_i^0)\) by using any convenient criterion. For example, \( \max \{\|\hat{m}_1^0 - \hat{m}_1^1\|, |\hat{c}_1^0 - \hat{c}_1^1|, |\hat{d}_1^0 - \hat{d}_1^1|\} \leq \varepsilon \). If \( \hat{m}_i^1 = \hat{m}_i^0, \hat{c}_i^1 = \hat{c}_i^0, \hat{d}_i^1 = \hat{d}_i^0 \), \( i = 1, \ldots, c \) and go to (S2).

In the model (M2), if we consider the special case which is just one cluster, i.e., \( c = 1 \), then we have the results which is similar to the solutions solved by Diamond [8].

**B. Numerical Input and Fuzzy Output**

Consider the cluster-wise fuzzy linear regression model (M3) \( Y_j = A_j \oplus B_j x_j, j = 1, \ldots, c \), \( j = 1, \ldots, n \), where \( x_j \in R \) and \( A_j, B_j, Y_j \) are \( LR \)-type fuzzy numbers for which \( (x_j, Y_j) \) are data pairs and \( A_j = (m_{e_j}, c_{e_j}, \beta_{e_j})_{LR}, B_j = (m_{\bar{e}_j}, c_{\bar{e}_j}, \beta_{\bar{e}_j})_{LR} \) are unknown coefficients. The corresponding two-stage weighted fuzzy least-squares optimization problem to the model (M3) is

(F3) minimize

\[
\tau(A_j, B_j) = \sum_{j=1}^{n} \left( \sum_{i=1}^{c} u_{ji}^0 \|e_{i,j}^0\|_{LR}(A_j \oplus B_j x_j, Y_j) \right).
\]
where $u = (u_1, \ldots, u_c)$ is a fuzzy $c$-partition obtained by a chosen fuzzy clustering procedure and $A_i$ and $B_i$ are unknown coefficients. Then the necessary conditions for a minimizer $(\hat{A}_i, \hat{B}_i)$ are (19)–(24) (see bottom of next page). Note that if $\delta_{ai} < 0$ and $\beta_{ai} < 0$, then we have that $\hat{E}_i = (\hat{m}_{ai}, -\hat{\beta}_{ai}, -\hat{\delta}_{ai})_{RL}$. If $\delta_{ai} < 0$ and $\beta_{ai} > 0$, then $\hat{E}_i = (\hat{m}_{ai}, 0, \hat{\beta}_{ai})_{LR}$ when $|\beta_{ai}| > |\delta_{ai}|$; and $\hat{E}_i = (\hat{m}_{ai}, 0, -\hat{\beta}_{ai})_{RL}$ when $|\beta_{ai}| < |\delta_{ai}|$. If $\delta_{ai} > 0$ and $\beta_{ai} < 0$, then $\hat{E}_i = (\hat{m}_{ai}, -\delta_{ai}, 0)_{LR}$ when $|\delta_{ai}| > |\beta_{ai}|$; and $\hat{E}_i = (\hat{m}_{ai}, -\beta_{ai}, 0)_{RL}$ when $|\delta_{ai}| < |\beta_{ai}|$, and similarly for $B_i$.

Based on these necessary conditions, we have the algorithm of the two-stage weighted fuzzy linear regression for the model (M3) which is similar to the Algorithm 1.
In the model (M3), consider the special case which is \( c = 1 \). Then we have that
\[
\hat{\alpha}_a = \frac{\sum_{j=1}^{n} (m_{t,y_j} - x_j (3\hat{m}_b + r\hat{\beta}_b))}{3n}
\]
(25)
\[
\hat{\alpha}_b = \frac{\sum_{j=1}^{n} (m_{t,y_j} - x_j (3\hat{m}_a + r\hat{\beta}_a))}{\sum_{j=1}^{n} x_j^2}
\]
(26)
\[
\hat{\beta}_a = \frac{\sum_{j=1}^{n} x_j (m_{t,y_j} - \hat{m}_a - x_j (3\hat{m}_b + r\hat{\beta}_b))}{3\sum_{j=1}^{n} x_j^2}
\]
(27)
\[
\hat{\beta}_b = \frac{\sum_{j=1}^{n} x_j (m_{t,y_j} - \hat{m}_a - x_j (3\hat{m}_b + r\hat{\beta}_b))}{r\sum_{j=1}^{n} x_j^2}
\]
(28)

Note that the special results (25)–(30) here is similar to the solutions which was solved by Diamond [8].

III. ONE-STAGE GENERALIZED FUZZY REGRESSION ANALYSIS

In Section II, we dealt with the weighted fuzzy regression analysis. That is a two-stage procedure which may be easier to be handled, but it shall heavily depend on the chosen fuzzy clustering procedure. Now let us embed the fuzzy clustering to the fuzzy regression model-fitting and combine both into a one-stage generalized procedure. That is, we consider the class memberships to be unknown parameters in the weighted fuzzy least-square objective function. This kind of one-stage generalized fuzzy regression method should be more useful and effective than two-stage weighted procedure in cluster-wise fuzzy regression model-fitting since it effectively takes into account the structure of regression-line shape encountered in the data set at each iteration of algorithm. Next we shall set up these algorithms of one-stage generalized fuzzy regression for the models (M1), (M2), and (M3).

The one-stage generalized fuzzy linear regression with the corresponding least-square optimization problem for the model (M1) is
\[
\text{(F4) minimize } \rho_m(a_{0i}, a_{1i}, u) = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij} P_{ij}^2 (a_{0i} + a_{1i} X_j, Y_j)
\]
where \( m \geq 1 \) and \( a_{0i} \) and \( a_{1i} \) are unknown coefficients and \( u = (u_{11}, u_{21}, \ldots, u_{c1}) \) is an unknown fuzzy \( c \)-partition. Consider the Lagrangian \( L_m(a_{0i}, a_{1i}, u, \lambda) \) with
\[
L_m(a_{0i}, a_{1i}, u, \lambda) = \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij} P_{ij}^2 (a_{0i} + a_{1i} X_j, Y_j) - \lambda \left( \sum_{i=1}^{c} u_{ij} - 1 \right)
\]
If we have the first derivatives of \( L_m(a_{0i}, a_{1i}, u, \lambda) \) with respect to all parameters equal zero, then we can get the necessary conditions for a minimizers \((\hat{a}_{0i}, \hat{a}_{1i}, \hat{u})\) of \( \rho_m(a_{0i}, a_{1i}, u) \) as follows
\[
\hat{u}_{ij} = \left( \sum_{i=1}^{c} \frac{P_{ij}^2 (a_{0i} + a_{1i} X_j, Y_j)}{P_{ij}^2 (a_{0i} + a_{1i} X_j, Y_j)} \right)^{-1}, \quad i = 1, \ldots, c; \quad j = 1, \ldots, n
\]
(31)

where if \( a_{1i} \geq 0, \hat{a}_{0i}, \hat{a}_{1i} \) are the same as (1) and (2), respectively; if \( a_{1i} < 0, \hat{a}_{0i}, \hat{a}_{1i} \) are the same as (3) and (4), respectively. Based on these necessary conditions, we can construct the algorithm of one-stage generalized fuzzy linear regression for the model (M1) as follows.

\[
\hat{m}_{ai} = \frac{\sum_{j=1}^{n} u_{ij}^m (m_{t,y_j} - x_j (3\hat{m}_b + r\hat{\beta}_b))}{3\sum_{j=1}^{n} x_j^2 u_{ij}^m}, \quad i = 1, \ldots, c
\]
(19)
\[
\hat{\alpha}_{ai} = \frac{\sum_{j=1}^{n} u_{ij}^m (\hat{m}_{ai} + x_j (\hat{m}_b - \lambda \hat{\alpha}_b) - m_{t,y_j})}{\sum_{j=1}^{n} x_j^2 u_{ij}^m}, \quad i = 1, \ldots, c
\]
(20)
\[
\hat{\beta}_{ai} = \frac{\sum_{j=1}^{n} x_j (m_{t,y_j} - \hat{m}_a - x_j (3\hat{m}_b + r\hat{\beta}_b))}{3\sum_{j=1}^{n} x_j^2 u_{ij}^m}, \quad i = 1, \ldots, c
\]
(21)
\[
\hat{m}_{bi} = \frac{\sum_{j=1}^{n} u_{ij}^m x_j (m_{t,y_j} - x_j (3\hat{m}_a + r\hat{\beta}_a)) + \lambda x_j \hat{\alpha}_a}{r\sum_{j=1}^{n} x_j^2 u_{ij}^m}, \quad i = 1, \ldots, c
\]
(22)
\[
\hat{\alpha}_{bi} = \frac{\sum_{j=1}^{n} u_{ij}^m x_j (\hat{m}_{bi} + x_j \hat{m}_a - \lambda \hat{\alpha}_a - m_{t,y_j})}{\sum_{j=1}^{n} x_j^2 u_{ij}^m}, \quad i = 1, \ldots, c
\]
(23)
\[
\hat{\beta}_{bi} = \frac{\sum_{j=1}^{n} u_{ij}^m x_j (m_{t,y_j} - x_j \hat{m}_a - r\hat{\beta}_a)}{r\sum_{j=1}^{n} x_j^2 u_{ij}^m}, \quad i = 1, \ldots, c
\]
(24)
TABLE III

<table>
<thead>
<tr>
<th>Model (M2)</th>
<th>Two-stage (F.CN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>$Y^1 = (11.420, 0.961, 0.961)_{F} - 0.155X$</td>
<td>$Y^1 = (-0.261, 0.000, 3.434)_{F} + 0.709X$</td>
</tr>
<tr>
<td>Residual</td>
<td>$\text{Residual}^{1} = 253.365967$</td>
<td>$\text{Residual}^{1} = 643.012793$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2 = 0.025201$</td>
<td>$R^2 = 0.022633$</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>Model (M2)</th>
<th>Two-stage (F.CN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>$Y^2 = (16.766, 0.917, 0.965)_{F} - 0.155X$</td>
<td>$Y^2 = (-0.261, 0.000, 3.434)_{F} + 0.709X$</td>
</tr>
<tr>
<td>Residual</td>
<td>$\text{Residual}^{2} = 125.655497$</td>
<td>$\text{Residual}^{2} = 56.454891$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2 = 0.012809$</td>
<td>$R^2 = 0.449474$</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Model (M2)</th>
<th>Two-stage (F.CN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>$Y^2 = (4.924, 0.844, 0.791)_{F} + 0.090X$</td>
<td>$Y^2 = (22.891, 1.620, 1.703)_{F} - 1.115X$</td>
</tr>
<tr>
<td>Residual</td>
<td>$\text{Residual}^{3} = 61.256417$</td>
<td>$\text{Residual}^{3} = 100.867512$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2 = 0.001316$</td>
<td>$R^2 = 0.104021$</td>
</tr>
</tbody>
</table>

(c)

<table>
<thead>
<tr>
<th>Model (M2)</th>
<th>Two-stage (F.CN)</th>
<th>One-stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>$Y^2 = (10.028, 0.875, 0.835)_{F} + 0.024X$</td>
<td>$Y^2 = (19.840, 1.569, 1.647)_{F} - 0.978X$</td>
</tr>
<tr>
<td>Residual</td>
<td>$\text{Residual}^{4} = 63.351406$</td>
<td>$\text{Residual}^{4} = 36.945052$</td>
</tr>
<tr>
<td>$R^2$</td>
<td>$R^2 = 0.009314$</td>
<td>$R^2 = 0.104021$</td>
</tr>
</tbody>
</table>

(d)

Algorithm 2: (of one-stage generalized fuzzy regression for the model (M1))

(S1) Fix $m \geq 1$: fix $c \in \{2, \ldots, n-1\}$; and fix any $\epsilon > 0$. Choose an initial fuzzy $c$-partition $\tilde{U}^{(0)}$ and initial $(\tilde{\alpha}^{(0)}, \tilde{\beta}^{(0)})$, $i = 1, \ldots, c$.

(S2) Calculate $(\tilde{\alpha}^{(1)}_{\tilde{b}}, \tilde{\beta}^{(1)}_{\tilde{b}})$ by using $\tilde{U}^{(0)}$ and (1) and (2) if $\tilde{\alpha}^{(0)}_{\tilde{b}} \geq 0$; using (3) and (4) if $\tilde{\alpha}^{(0)}_{\tilde{b}} < 0$, $i = 1, \ldots, c$.

(S3) Calculate $\tilde{U}^{(1)}$ by using (31).

(S4) Compare $\tilde{U}^{(0)}$ to $\tilde{U}^{(1)}$, using any convenient criterion. For example, $\max_{i} \left[ \frac{1}{c} \sum \limits_{i=1}^{c} \left| \tilde{U}^{(0)}_{\tilde{b}} - \tilde{U}^{(1)}_{\tilde{b}} \right| \right] \leq \epsilon$. If $\tilde{U}^{(1)}$ is sufficiently close to $\tilde{U}^{(0)}$, stop. Otherwise, set $\tilde{U}^{(1)} = \tilde{U}^{(0)}, \tilde{\alpha}^{(1)}_{\tilde{b}} = \tilde{\alpha}^{(0)}_{\tilde{b}}, \tilde{\beta}^{(1)}_{\tilde{b}} = \tilde{\beta}^{(0)}_{\tilde{b}}, i = 1, \ldots, c$ and go to (S2).

The one-stage generalized fuzzy linear regression with the corresponding least-squares optimization problem for the model (M2) is

\[ \text{(F5)} \quad \rho_{m}(E_{i}, \alpha_{i}, u) = \sum_{j=1}^{n} \sum_{i=1}^{c} \tilde{U}_{\tilde{b}}^{m} \rho_{2}(E_{i} \oplus \alpha_{i} X_{j}, Y_{j}) \]

where $m \geq 1$ and $E_{i}$ and $\alpha_{i}$ are unknown coefficients and $u = \left( u_{1}, u_{2}, \ldots, u_{c} \right)$ is an unknown fuzzy $c$-partition. Similarly, we can get the necessary conditions for a minimizers $(\hat{E}_{i}, \hat{\alpha}_{i}, \hat{u})$ of $\rho_{m}(E_{i}, \alpha_{i}, u)$ as follows

\[ \hat{c}_{ij} = \left( \sum_{i=1}^{c} \frac{\left( d_{j}^{i R}(\hat{E}_{i} \oplus \hat{\alpha}_{i} X_{j}, Y_{j}) \right)^{2}}{\left( d_{j}^{i R}(\hat{E}_{i} \oplus \hat{\alpha}_{i} X_{j}, Y_{j}) \right)^{2}} \right)^{-1} \]

\[ \hat{\alpha}_{i} = \frac{\sum_{i=1}^{c} \tilde{u}_{i} \tilde{U}_{\tilde{b}}^{m} \rho_{2}(E_{i} \oplus \alpha_{i} X_{j}, Y_{j})}{\sum_{i=1}^{c} \tilde{u}_{i} \tilde{U}_{\tilde{b}}^{m} \rho_{2}(E_{i} \oplus \alpha_{i} X_{j}, Y_{j})} \]

where $\alpha_{i} \geq 0, \tilde{u}_{i}, \tilde{\alpha}^{(0)}_{\tilde{b}}, \tilde{\alpha}^{(1)}_{\tilde{b}}, \tilde{\beta}^{(0)}_{\tilde{b}}, \tilde{\beta}^{(1)}_{\tilde{b}}$ are the same as the (11)–(14), respectively; if $\alpha_{i} < 0$, $\tilde{u}_{i}, \tilde{\alpha}^{(0)}_{\tilde{b}}, \tilde{\alpha}^{(1)}_{\tilde{b}}, \tilde{\beta}^{(0)}_{\tilde{b}}, \tilde{\beta}^{(1)}_{\tilde{b}}$ are the same as (15)–(18), respectively. Based on these necessary conditions, we have the algorithm of one-stage generalized fuzzy linear regression for the model (M2) which is similar to the Algorithm 2.

The one-stage generalized fuzzy linear regression with the corresponding least-squares optimization problem for the model (M3) is
TABLE IV
THE PREDICTIONS, RESIDUALS, AND $R^2$ UNDER THE MODEL (M3) IN THE CASE OF (a) $c = 1$, (b) $c = 2$, (c) $c = 3$, AND (d) $c = 4$

(a)

<table>
<thead>
<tr>
<th>Two - stage (FC; N)</th>
<th>One - stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y} = (11.421, 0.828, 0.828)_T + (-0.155, 0.001, 0.000)_T m_x$</td>
<td>$\hat{Y} = (0.880, 2.135, 7.853)_T + (0.625, 0.644, 0.217)_T m_x$</td>
</tr>
<tr>
<td>residual = 2929.70996</td>
<td>residual = 3808.90823</td>
</tr>
<tr>
<td>$R^2 = 0.023814$</td>
<td>$R^2 = 0.023814$</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Two - stage (FC; N)</th>
<th>One - stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y}^1 = (17.765, 0.745, 0.779)_T + (-0.220, 0.000, 0.000)_T m_x$</td>
<td>$\hat{Y}^1 = (-0.849, 0.860, 0.870)_T + (0.776, 0.002, 0.001)_T m_x$</td>
</tr>
<tr>
<td>residual$^1 = 224.679997$</td>
<td>residual$^1 = 52.004986$</td>
</tr>
<tr>
<td>residual$^2 = 102.224430$</td>
<td>residual$^2 = 106.107475$</td>
</tr>
<tr>
<td>$R^2 = 0.010644$</td>
<td>$R^2 = 0.447476$</td>
</tr>
</tbody>
</table>

(c)

<table>
<thead>
<tr>
<th>Two - stage (FC; N)</th>
<th>One - stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Y}^1 = (17.496, 0.685, 0.783)_T + (-0.179, 0.001, 0.000)_T m_x$</td>
<td>$\hat{Y}^1 = (-0.974, 0.797, 0.873)_T + (0.780, 0.003, 0.000)_T m_x$</td>
</tr>
<tr>
<td>residual$^1 = 90.959285$</td>
<td>residual$^1 = 34.867188$</td>
</tr>
<tr>
<td>residual$^2 = 100.892919$</td>
<td>residual$^2 = 15.261712$</td>
</tr>
<tr>
<td>residual$^3 = 64.011940$</td>
<td>residual$^3 = 24.564290$</td>
</tr>
<tr>
<td>$R^2 = 0.009069$</td>
<td>$R^2 = 0.249599$</td>
</tr>
</tbody>
</table>

(d)

Example 1: In this example we use three data sets of Fig. 5 (a)–(c) from Diamond [8]. We run Algorithm 1 proposed in Section II on these data sets according to different models (M1), (M2), and (M3) in the special case of $c = 1$. We list the results correspondingly in Table I(a)–(c). We find that residuals from Algorithm 1 are always smaller than these from Diamond [8]. These results tell us that our metric defined on the space $\mathbb{F}^T(R)$ of all triangular fuzzy numbers is better than the metric proposed by Diamond [8].

Example 2: In this example we artificially give a data set of 40 triangular fuzzy data pairs shown in Fig. 6. Fig. 1 gives its geometric picture. Fig. 2 gives the location of observations in $\mathbb{R}^2$-plant. In two-stage procedure, we first choose FCN clustering algorithm proposed by Yang and Ko [24] to get the fuzzy c-partition $u = (u_1, u_2, \ldots, u_c)$ for the data set of Fig. 6. Then we use Algorithm 1 to get c predictions, residuals and $R^2$. We also run the same data set on Algorithm 2 of one-stage procedure which are described in Section III. The results are shown in Table II(a)–(d) according to different numbers of cluster $c = 1, 2, 3$ and 4. In two-stage procedure, we have that $R^2 = 0.025201$ when $c = 1$; $R^2 = 0.01289$ when $c = 2$; $R^2 = 0.00136$ when $c = 3$; $R^2 = 0.00441$ when $c = 4$. Based on the criterion $R^2$ for choosing the number of cluster which is proposed in Section II, we may choose $c = 1$ in two-stage procedure. In one-stage procedure, we have that $R^2 = 0.023033$ when $c = 1$; $R^2 = 0.449747$ when $c = 2$; $R^2 = 0.104021$ when $c = 3$; $R^2 = 0.167925$ when

(F6) minimize

$$
\rho_m(A_i, B_i, u) = \frac{1}{m} \sum_{j=1}^{c} \sum_{i=1}^{n} w_{ij}^m d_{LR}(A_i \oplus B_i x_j, Y_j)
$$

where $m \geq 1$ and $A_i$ and $B_i$ are unknown coefficients and $u = (u_1, u_2, \ldots, u_c)$ is an unknown fuzzy c-partition. Similarly, we can get the necessary conditions for a minimizer $(\hat{A}_i, \hat{B}_i, \hat{u})$ of $\rho_m(A_i, B_i, u)$ as follows

$$
\hat{u}_{ij} = \left( \sum_{k=1}^{c} \frac{(d_{LR}(A_k, B_k x_j, Y_j))^m}{(d_{LR}(A_i, B_i x_j, Y_j))^m} \right)^{-1},
$$

where $\hat{u}_{ij}, \hat{A}_{ij}, \hat{B}_{ij}, \hat{m}_{ij}, \beta_{ij}$, and $\hat{\beta}_{ij}$ are the same as (19)–(24), respectively. Based on these necessary conditions, we have the algorithm of one-stage generalized fuzzy linear regression for the model (M3) which is similar to the Algorithm 2. Note that the weighted mean $R^2(c)$ of determination coefficients $R^2_i, i = 1, \ldots, c$, defined in Section II are still used as a criterion on the decision of an optimal c here.

IV. NUMERICAL EXAMPLES AND CONCLUSION

In this section we present some numerical comparisons and then make conclusions. We choose the index of fuzziness $m = 2$ and the stopping criterion $\varepsilon = 0.0001$ in all the following numerical examples.
Fig. 6. Data pairs of 40 triangular fuzzy numbers.

We shall choose \( c = 2 \) in one-stage procedure. Let us see memberships of two clusters of observations in Fig. 6 based on two-stage and one-stage procedures for \( c = 2 \) which are presented in Table II(e). Figs. 3 and 4 show clustering results in \( \mathbb{R}^2 \)-plant. We get good clusters of regression-line shape in Fig. 4, but not good shape in Fig. 3. This is because one-stage procedure has taken into account the structure of regression-line shape in the data set at each iteration of algorithm. If we compare residuals, one-stage procedure is highly recommended. Similarly, we have results shown in Table III(a)–(d) for the model (M2) and Table IV(a)–(d) for the model (M3).

**Example 3:** In this example we consider a special case for the data set in Fig. 6, that is \( X = Y = (m_x, \sigma_x, \beta_x) \). In two-stage procedure, \( u = (u_1, \ldots, u_c) \) is a fuzzy \( c \)-partition which comes from the chosen fuzzy \( c \)-mean (FCM). We use Algorithm 1 to get \( c \) predicted regression lines. We mention that this kind of two-stage procedure was presented in Jajuga [12]. We also run the same data set on Algorithm 2 of one-stage procedure presented in Section III. The results are shown in Table V(a)–(d) correspondingly to different numbers of cluster \( c = 1, 2, 3 \) and 4. It is obvious that \( c = 2 \) is the best choice based on the \( R^2 \) criterion; and also one-stage procedure is better than two-stage procedure on comparison of residuals. We may say that the procedure proposed in Jajuga [12] is not a so good procedure for the cluster-wise fuzzy regression analysis.

In this paper we have presented the cluster-wise fuzzy regression analysis on two approaches: one is the two-stage weighted fuzzy regression and another is the one-stage generalized fuzzy regression. The two-stage procedure presented in Section II generalizes methods of Diamond [8] and Jajuga [12]. The cluster memberships out of a chosen clustering algorithm are used as weights in the two-stage weighted fuzzy regression. Therefore, results heavily depend on the chosen clustering algorithm. The one-stage procedure is a new created procedure. It embeds fuzzy clustering techniques into the fuzzy regression model fitting at each step of procedure. Therefore, the structure of regression-line shape is taken into account at each iteration of algorithm. Based on numerical experiments we conclude that one-stage procedure is highly recommended to be used in the cluster-wise regression analysis.
Note that the models (M1), (M2), and (M3) are all simple linear fuzzy regression models. We can use the summation operation to extend these to multiple linear fuzzy regression models. Except the models (M1), (M2), and (M3), we can also suggest the model of another type (M4) \( Y = A \oplus B \odot X \) where \( A, B, X, Y \in \mathcal{F}_{LR}(\mathbb{R}) \). Say for example, for \( B > 0, X > 0 \), by choosing different \( r_1, r_2, l_1 \) and \( l_2 \), we consider to minimize the objective function

\[
\rho(A_i, B_i) = \sum_{j=1}^{n} \sum_{i=1}^{c} \nu_{ij}^m b_{ij}^2 (A_i \oplus B_i \odot X_j, Y_j) = \sum_{j=1}^{n} \sum_{i=1}^{c} \nu_{ij}^m \left[ (m_{ai} + m_{bi}x_{ij} - m_{yij})^2 + (m_{ai} + m_{bi}x_{ij} - m_{yij})^2 - l_1(\alpha_{ai} + m_{bi}\alpha_{xj} + m_{xj}\alpha_{bi}) - (m_{yij} - l_2\alpha_{yij})^2 \right.
\]

Then we can construct algorithms of the cluster-wise fuzzy regression for the model (M4) which are similar to Algorithms 1 and 2.

Mixtures of distributions are widely applied to analysis of data (McLachlan and Basford [14]). Switching regressions are mixtures of regressions models. These have been applied in diverse areas such as economics and music perception, etc. (Quandt [16], Kiefer [13], and De Veaux [7]). The cluster-wise fuzzy regressions can be thought of fuzzy variations of mixtures of regressions models. These have been applied in diverse areas such as economics and music perception, etc. (Quandt [16], Kiefer [13], and De Veaux [7]). The cluster-wise fuzzy regressions can be thought of fuzzy variations of switching regressions. These fuzzy-type switching regressions can be applied in diverse areas under fuzzy environment.

ACKNOWLEDGMENT

The authors are grateful to the referees for their valuable suggestions and comments.

REFERENCES


Miin-Shen Yang was born in Taiwan in 1955. He received the B.S. degree in mathematics from the Chung-Yuan Christian University, Chung-Li, Taiwan, R.O.C., in 1977, the M.S. degree in applied mathematics from the National Chiao-Tung University, Hsinchu, Taiwan, in 1980, and Ph.D. degree in statistics from the University of South Carolina, Columbia, in 1989.

He is now a Professor, Department of Mathematics, Chung-Yuan Christian University. His current research interests include applications of statistics, cluster analysis, and fuzzy data analysis.

Cheng-Hsiu Ko was born in Fengshan, Taiwan, R.O.C., in 1969. She received the B.S. degree in mathematics in 1992 and the M.S. degree in applied mathematics in 1994 from the Chung-Yuan Christian University, Chung-Li, Taiwan.

She is a mathematics teacher, Yu-Dah Business High School, Chung-Li, Taiwan. She is also a part-time Instructor, Department of Mathematics, Chung-Yuan Christian University.