International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:8, No:9, 2014

A Fuzzy Linear Regression Model Based on Dissemblance Index

Shih-Pin Chen, Shih-Syuan You

Abstract**—**Fuzzy regression models are useful for investigating the relationship between explanatory variables and responses in fuzzy environments. To overcome the deficiencies of previous models and increase the explanatory power of fuzzy data, the graded mean integration (GMI) representation is applied to determine representative crisp regression coefficients. A fuzzy regression model is constructed based on the modified dissemblance index (MDI), which can precisely measure the actual total error. Compared with previous studies based on the proposed MDI and distance criterion, the results from commonly used test examples show that the proposed fuzzy linear regression model has higher explanatory power and forecasting accuracy.

Keywords**—**Dissemblance index, fuzzy linear regression, graded mean integration, mathematical programming.

I. Introduction

REGRESSION analysis has proven to be a powerful methodology for analyzing the relationship between dependent variables (also called response variables) and independent variables (also called explanatory variables) [1], [2]. However, in many real-world situations, the data involve subjective human judgment under incomplete data conditions [3]-[5]. Data of this kind are usually described by approximate linguistic terms such as "approximately 50 meters" instead of exact values due to a lack of historical information or inexact knowledge. Such incomplete information belongs to epistemic uncertainty [6], [7] and can be modelled by fuzzy sets [8], [9]. The articles proposed by Tanaka et al. [10], [11] are well-known pioneer papers on fuzzy regression analysis. Since then, a significant amount of research has been devoted to the use of fuzzy regression analysis to evaluate functional relationships involving data of this kind. Fuzzy regression analysis has been successfully used in practical applications such as business cycle analysis [12], productivity and consumer satisfaction [13], product life cycle prediction [14], R&D project evaluation [15], and reservoir operations [16].

Many fuzzy regression models have been developed over the past three decades. These models employed one of three different resolution techniques [17]: the least-squares (LS) method (for example, [17]-[21]), the mathematical programming (MP) method (for example, [22]-[28]), and the support vector machine (SVM) (for example, [29], [30]). Several articles have noted the drawbacks of existing fuzzy regression models [17], [21], [28], [31]. As Redden and Woodall [32] and Kao and Chyu [26] noted, the main drawback of the model developed by Tanaka et al. [11] and subsequent variations [22]-[24] is that increasing the number of observations to establish the regression model results in a wider spread of the estimated fuzzy response and then a fuzzier parameter estimation. This phenomenon contradicts the general observation that larger numbers of observations provide better estimations. To avoid this drawback, Diamond [18] and Kao and Chyu [21], [26] used crisp (numeric) regression coefficients to describe fuzzy relationships between dependent and independent variables. However, as Kao and Lin [31] noted, these models cannot address situations with decreasing spread or variable spread. Chen and Dang [28] proposed a variable spread method with a higher explanatory power to resolve this problem.

Moreover, many existing methods have been developed based on the criterion of Kim and Bishu [33], in which the estimation error is defined as the difference between the observed and estimated responses. However, Chen and Hsueh [17], [27] noted that the criterion fails when the observed and estimated fuzzy responses do not intersect. In this case, the measure of the estimation error defined by Kim and Bishu [33] cannot reflect the actual difference. To address this problem, Chen and Hsueh [17] proposed an approach based on the criterion of minimizing the sum of the average squared distance to measure the estimation error according to the differences between the observed and estimated fuzzy dependent variables at several α-levels. Theoretically, using larger numbers of α-cuts will yield estimates that are more precise. Chen and Hsueh noted that "previous studies suggest that only a few α-cuts are sufficient for fuzzy regression analysis if the fuzzy observations are characterized by fuzzy numbers with linear membership functions, such as triangular fuzzy numbers." However, this characterization may decrease the generality of their model and thus limits the applications of this method.

In this paper, a new approach is proposed to address this problem. For each pair of fuzzy observation and explanatory variables, the proposed modified dissemblance index (MDI) is adopted to measure the error in estimation. Accordingly, a two-stage approach with higher explanatory power is also developed. In the first stage, the graded mean integration (GMI) representation method is adopted to establish representative crisp values of fuzzy responses and independent variables based on the concept that the membership grade represents the fuzziness of the fuzzy set. The crisp LS method is then employed to estimate the regression coefficients. In the second stage, an MP model is formulated to determine the variable fuzzy error term for each pair of explanation variables and the corresponding response. The objective of each formulated

Shih-Pin Chen, Shih-Syuan You are with the Department of Business Administration, National Chung Cheng University, Ming-Hsiung, Chia-Yi 621, Taiwan (e-mail: chensp@ccu.edu.tw).

mathematical program is to minimize the total estimation errors measured by the proposed MDI subject to the established constraints, including that the spread of each estimated response is equal to that of the associated observed response. The concept of variable spread has been demonstrated to be useful in overcoming the problem of decreasing spread or variable spread [28].

II. Problem Statement

In classical regression analysis, the most frequently used type for evaluating the relationship between independent and dependent variables can be written in the linear form as follows:

$$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i, \quad i = 1,...,n, \tag{1}$$

where Y_i and X_{ij} are the dependent variable and the *j*th independent variable of the *i*th observation, respectively; β_j is the regression parameter associated with the *j*th independent variable; and ε_i is the error term associated with the ith observation.

The parameter β_j is usually estimated from sample data by statistical methods with sound theoretical properties, such as the least-squares method (LSM). If some of the observations Y_i and X_{ij} are collected under incomplete data conditions and described by fuzzy sets, then Model (1) becomes a fuzzy linear regression model. As a simple example, consider the case of a simple linear regression (SLR) model with one independent variable. Multiple regression cases can be straightforwardly generalized from this case. The fuzzy SLR model can be described as follows:

$$\tilde{Y}_i = \beta_0 + \beta_1 \tilde{X}_i + \tilde{\varepsilon}_i, \quad i = 1,...,n, \tag{2}$$

where $(\tilde{X}_i, \tilde{Y}_i)$ are n pairs of fuzzy observations and $\tilde{\varepsilon}_i$ represents the fuzzy errors. Without loss of generality, suppose that $\tilde{X}_i$ and $\tilde{Y}_i$, $i = 1,...,n$, are fuzzy numbers and are characterized by membership functions $\mu_{\tilde{X}_i}$ and $\mu_{\tilde{Y}_i}$, respectively. In fuzzy set theory, a membership function represents the fuzziness of the corresponding fuzzy set [32].

The problem is to determine the estimates for β_0, β_1, and $\tilde{\varepsilon}_i$ such that Model (2) has the highest explanatory power and forecasting accuracy under fuzzy environments. New ideas embedded in the proposed two-stage method are presented below.

III. The Idea

This paper proposes a new two-stage method that improves upon the models of Kao and Chyu [26] and Chen and Dang [28]. The main idea of the improved method is to use a more representative defuzzification method to establish crisp estimated regression parameters and to minimize the total estimation error using a more effective method to find the variable error terms.

A. *Crisp Regression Coefficients*

Ideal regression coefficients in a fuzzy linear regression model should be crisp numbers to avoid the main drawback of an increasing (or decreasing or irregular) spread of the estimated response [24]. Moreover, the regression coefficients should be representative in fuzzy environments because they are used to describe the contributions of the input variables in explaining the corresponding response. The regression coefficients derived in this paper follow these two criteria. Assume that all of the observations are triangular fuzzy numbers, which have been commonly employed by previous studies. The triangular fuzzy number $\tilde{X}_i$ can be described by a triple element set $\tilde{X}_i = (X_{iL}, X_{iM}, X_{iU})$ with the membership function $\mu_{\tilde{X}_i}(x)$ as follows:

$$\mu_{\tilde{X}_i}(x) = \begin{cases} L(x) = (x - X_{iL})/(X_{iM} - X_{iL}), & X_{iL} \le x \le X_{iM}, \\ R(x) = (X_{iU} - x)/(X_{iU} - X_{iM}), & X_{iM} \le x \le X_{iU}, \end{cases} \tag{3}$$

where X_{iL}, X_{iM}, and X_{iU} are the smallest, most likely, and largest values of $\tilde{X}_i$, respectively. The fuzzy number $\tilde{Y}_i$ can be described by $\tilde{Y}_i = (Y_{iL}, Y_{iM}, Y_{iU})$, where Y_{iL}, Y_{iM}, and Y_{iU} are the smallest, most likely, and largest values of $\tilde{Y}_i$, respectively, and the membership function $\mu_{\tilde{Y}_i}(y)$ is defined in a similar manner.

One idea of estimating β_0 and β_1 is to defuzzify the fuzzy observations $\tilde{X}_i$ and $\tilde{Y}_i$ to crisp values and then apply the conventional LSM (for example [24]). Many defuzzification methods have been proposed and discussed by previous studies, in which the center of gravity (COG) method, also called the centroid method [32], is commonly used. In fact, the defuzzified value X_i^C is the simple average of X_{iL}, X_{iM}, and X_{iU}; that is, $X_i^C = (X_{iL} + X_{iM} + X_{iU})/3$. Similarly, $Y_i^C = (Y_{iL} + Y_{iM} + Y_{iU})/3$.

The present paper adopts the GMI representation introduced by Chen and Hsieh [34]-[36]. They demonstrated the GMI method outperforms several methods proposed by previous studies [35], [36]. Let $L^{-1}(\alpha)$ and $R^{-1}(\alpha)$ be the inverse functions of $L(x)$ and $R(x)$, respectively. The GMI representation of $\tilde{X}_i$ can be calculated as follows:

$$\begin{aligned} X_i^{GMI} &= \frac{\int_0^1 0.5\alpha\left[L^{-1}(\alpha) + R^{-1}(\alpha)\right]d\alpha}{\int_0^1 \alpha d\alpha} \\ &= \int_0^1 \alpha\left[X_{iL} + \alpha(X_{iM} - X_{iL}) + X_{iU} - \alpha(X_{iU} - X_{iM})\right]d\alpha \\ &= \frac{1}{6}\left(X_{iL} + 4X_{iM} + X_{iU}\right). \end{aligned} \tag{4}$$

It is also easy to calculate $Y_i^{GMI} = \left(Y_{iL} + 4Y_{iM} + Y_{iU}\right)/6$ in the same manner. Based on the set of crisp values (X_i^{GMI}, Y_i^{GMI}), the classical LSM is applied to estimate representative values of b_0 and b_1 for the regression parameters of β_0 and β_1, respectively. These LS estimates are as follows:

$$\left(b_1\right)_{GMI} = \frac{\sum_{i=1}^{n} X_i^{GMI} Y_i^{GMI} - n\bar{X}_{GMI}\bar{Y}_{GMI}}{\sum_{i=1}^{n} (X_i^{GMI})^2 - n(\bar{X}_{GMI})^2}, \tag{5a}$$

$$\left(b_0\right)_{GMI} = \bar{Y}_{GMI} - b_1 \bar{X}_{GMI}, \tag{5b}$$

where $\bar{X}_{GMI} = \sum_{i=1}^{n} X_i^{GMI}/n$ and $\bar{Y}_{GMI} = \sum_{i=1}^{n} Y_i^{GMI}/n$.

It is clear that X_i^{GMI} and Y_i^{GMI} are more reasonable and representative than the defuzzified values X_i^C and Y_i^C, which are derived by the COG method, because the former are weighted averages with the most likely values of X_{iM} and Y_{iM} assigned quadruple weights. Therefore, the values of b_0 and b_1 obtained based on the values of X_i^{GMI} and Y_i^{GMI} are more representative.

B. Error Terms

Conceptually, the error term in (2) can be expressed as:

$$\tilde{\varepsilon}_i = \tilde{Y}_i - \left(\beta_0 + \beta_1 \tilde{X}_i\right), \quad i = 1, \ldots, n. \tag{6}$$

By applying the procedure proposed in the preceding subsection, β_0 and β_1 can be estimated from b_0 and b_1, respectively, denoted as $\left(b_0\right)_{GMI}$ and $\left(b_1\right)_{GMI}$. Substituting them into (6), we obtain the following:

$$\tilde{\varepsilon}_i = \tilde{Y}_i - \left[\left(b_0\right)_{GMI} + \left(b_1\right)_{GMI} \tilde{X}_i\right], \quad i = 1, \ldots, n. \tag{7}$$

when $\tilde{X}_i$ and $\tilde{Y}_i$ are triangular fuzzy numbers, the error terms $\tilde{\varepsilon}_i$ are triangular as well. Chen and Dang [28] noted that using a unique $\tilde{E}$ to estimate $\tilde{\varepsilon}_i$, $i = 1, \ldots, n$ cannot address the case of a decreasing or variable spread. The present paper follows the concept of variable spread proposed by Chen and Dang [28], where each $\tilde{\varepsilon}_i$ is estimated by a triangular fuzzy number $\tilde{E}_i = (-LS_i, 0, RS_i)$ and LS_i and RS_i are the estimated left spread and right spread of the ith fuzzy observation, respectively. Consequently, the estimated response is:

$$\hat{\tilde{Y}}_i = \left(b_0\right)_{GMI} + \left(b_1\right)_{GMI} \tilde{X}_i + \tilde{E}_i, \quad i = 1, \ldots, n. \tag{8}$$

To determine the best values for LS_i and RS_i, Chen and Dang [28] employed a mathematical program with the objective of minimizing the total estimation error. As in several related studies, the total estimation error is defined by the criterion of Kim and Bishu [33]; however, as Chen and Hsueh [17], [27] noted, this error cannot reflect the actual difference if the observed and estimated fuzzy responses do not intersect. Consequently, the explanatory power will be undermined if the fuzzy regression model is constructed based on Kim and Bishu's criterion. Therefore, an improved measurement of the estimation error is necessary.

To precisely evaluate the difference between two fuzzy numbers, the present paper proposes the MDI measurement based on Kaufmann and Gupta [37]. In (5), when $\tilde{X}_i$ and $\tilde{E}_i$ are triangular fuzzy numbers, $\hat{\tilde{Y}}_i$ is also a triangular fuzzy number. The MDI between the estimated response $\hat{\tilde{Y}}_i = (\hat{Y}_{iL}, \hat{Y}_{iM}, \hat{Y}_{iU})$ and the observed response $\tilde{Y}_i = (Y_{iL}, Y_{iM}, Y_{iU})$ is:

$$D_i\left(\hat{\tilde{Y}}_i, \tilde{Y}_i\right) = \int_0^1 \left(\left| (\hat{\tilde{Y}}_i)_\alpha^L - (\tilde{Y}_i)_\alpha^L \right| + \left| (\hat{\tilde{Y}}_i)_\alpha^U - (\tilde{Y}_i)_\alpha^U \right| \right) d\alpha, \; i = 1, \ldots, n, \tag{9}$$

where $(\hat{\tilde{Y}}_i)_\alpha^L$ and $(\hat{\tilde{Y}}_i)_\alpha^U$ are the lower and upper bounds of the α-cut of $\hat{\tilde{Y}}_i$, respectively, and $(\tilde{Y}_i)_\alpha^L$ and $(\tilde{Y}_i)_\alpha^U$ are those of the α-cut of $\tilde{Y}_i$, respectively. Because $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ is developed based on the concept of distance, as in the method proposed by Chen and Hsueh [17], it can reflect the actual difference even in the case where the observed and estimated fuzzy responses do not intersect; thus, this approach avoids the drawback of existing methods that are based on Kim and Bishu's criterion. If the observed and estimated fuzzy responses do not overlap, the value of $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ calculated from (9) becomes large because it includes the area between them. More importantly, MDI can also overcome the difficulty of determining the appropriate number of α-cuts used in the method proposed by Chen and Hsueh [15]. Therefore, the present paper adopts $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ as the estimation error to increase the generality of the proposed model and thus has wider applications.

A smaller value of $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ indicates that the estimated response of $\hat{\tilde{Y}}_i = (\hat{Y}_{iL}, \hat{Y}_{iM}, \hat{Y}_{iU})$ better fits the corresponding observed response of $\tilde{Y}_i = (Y_{iL}, Y_{iM}, Y_{iU})$. Therefore, the objective of the proposed fuzzy regression model is to find LS_i and RS_i, $i = 1, 2, \ldots, n$ such that the total estimation error of $\sum_{i=1}^{n} D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ is minimized. Because the MDI is a more representative measure of the estimation error, the resulting fuzzy regression model will have a higher explanatory power.

To handle the case of a decreasing or variable spread of the observed responses, the present paper follows Chen and Dang [28] and sets the width of the support of each estimated fuzzy response as equal to that of its associated observed fuzzy response:

$$\hat{Y}_{iU} - \hat{Y}_{iL} = Y_{iU} - Y_{iL}, \; i = 1, 2, \ldots, n. \tag{10}$$

Moreover, to obtain meaningful left and right spreads, two sets of constraints regarding the lower bound of the individual spread of the estimated response should be considered [26], [28]:

$$\hat{Y}_{iM} - \hat{Y}_{iL} \geq L_{\min},\ i = 1,\ldots,n, \tag{11a}$$

$$\hat{Y}_{iU} - \hat{Y}_{iM} \geq R_{\min},\ i = 1,\ldots,n, \tag{11b}$$

where $L_{\min}$ and $R_{\min}$ are the minimum left and right spreads of the observed responses, respectively. Incorporating (8), (9), (10), (11a), and (11b), the proposed mathematical program to find the optimal values of LS_i^* and RS_i^*, $i = 1,2,\ldots,n,$ such that the total estimation error is minimized, can be formulated as follows:

$$\begin{aligned}
&\min_{LS_i, RS_i} \sum_{i=1}^{n} D_i\left(\hat{\tilde{Y}}_i, \tilde{Y}_i\right) \\
&\text{s.t. } D_i\left(\hat{\tilde{Y}}_i, \tilde{Y}_i\right) = \int_0^1 \left(\left| (\hat{\tilde{Y}}_i)_\alpha^L - (\tilde{Y}_i)_\alpha^L \right| + \left| (\hat{\tilde{Y}}_i)_\alpha^U - (\tilde{Y}_i)_\alpha^U \right| \right) d\alpha,\ i = 1,\ldots,n, \\
&\quad \hat{\tilde{Y}}_i = (b_0)_{GMI} + (b_1)_{GMI}\, \tilde{X}_i + (-LS_i, 0, RS_i), \quad i = 1,\ldots,n, \\
&\quad \hat{Y}_{iU} - \hat{Y}_{iL} = Y_{iU} - Y_{iL},\ i = 1,\ldots,n. \\
&\quad \hat{Y}_{iM} - \hat{Y}_{iL} \geq L_{\min},\ i = 1,\ldots,n, \\
&\quad \hat{Y}_{iU} - \hat{Y}_{iM} \geq R_{\min},\ i = 1,\ldots,n.
\end{aligned} \tag{12}$$

Because Chen and Dang [28] proved that their proposed mathematical program is always feasible, Model (12) is always consistent as well.

Note that the difference $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ in the first set of constraints of Model (12) can be easily calculated as the sum of the area between the left-shape functions of the observed and estimated responses and the area between the right-shape functions of the observed and estimated responses. If the observed and estimated responses do not intersect, $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ is equal to the sum of the areas of the observed and estimated responses and the area between them. Moreover, although there are several cases involving the relative locations of the estimated and the observed responses [28], the calculation of $D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i)$ can be simplified by eliminating impossible cases because the spread of each estimated response is defined as equal to that of its associated observed response. In addition, the number of decision variables can be easily reduced, which improves the efficiency of solving Model (12). According to (10), we have

$$\hat{Y}_{iU} - \hat{Y}_{iL} = (b_1)_{GMI}(X_{iU} - X_{iL}) - RS_i + LS_i = Y_{iU} - Y_{iL},\ i = 1,2,\ldots,n,$$

or

$$RS_i = (b_1)_{GMI}(X_{iU} - X_{iL}) + LS_i - (Y_{iU} - Y_{iL}),\ i = 1,2,\ldots,n.$$

Therefore, Model (12) can be simplified as the following mathematical program with only *n* decision f variables:

$$\begin{aligned}
&\min_{LS_i} \sum_{i=1}^{n} D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i) \\
&\text{s.t. } D_i(\hat{\tilde{Y}}_i, \tilde{Y}_i) = \int_0^1 \left(\left| (\hat{\tilde{Y}}_i)_\alpha^L - (\tilde{Y}_i)_\alpha^L \right| + \left| (\hat{\tilde{Y}}_i)_\alpha^U - (\tilde{Y}_i)_\alpha^U \right| \right) d\alpha,\ i = 1,\ldots,n, \\
&\quad \hat{\tilde{Y}}_i = (b_0)_{GMI} + (b_1)_{GMI}\, \tilde{X}_i + (-LS_i, 0, RS_i), \quad i = 1,\ldots,n, \\
&\quad RS_i = (b_1)_{GMI}(X_{iU} - X_{iL}) + LS_i - (Y_{iU} - Y_{iL}), \quad i = 1,\ldots,n, \\
&\quad \hat{Y}_{iM} - \hat{Y}_{iL} \geq L_{\min},\ i = 1,\ldots,n, \\
&\quad \hat{Y}_{iU} - \hat{Y}_{iM} \geq R_{\min},\ i = 1,\ldots,n.
\end{aligned} \tag{13}$$

Consequently, the error terms of the proposed fuzzy regression model can be effectively and efficiently derived by solving Model (13) using mathematical programming solvers [38].

IV. Example

To illustrate the suitability of the proposed fuzzy linear regression model for solving different types of fuzzy regression problems and that it has a greater explanatory power than models in previous studies, we examine an example that is commonly discussed in the literature. This example was designed by Tanaka et al. [24]. The commonly used performance measure, Chen and Hsueh's distance criterion [17], [27], is adopted to compare the performance of the proposed fuzzy regression model with models proposed by previous studies. Moreover, the MDI criterion proposed in the present paper is also adopted.

Tanaka et al. [24] designed an example to illustrate their fuzzy regression model for addressing the problem of one crisp explanatory variable and fuzzy responses that are triangular fuzzy numbers. This example includes five pairs of observations $[x_i, \tilde{y}_i]$, which are shown in the first two columns of Table I.

TABLE I
THE ESTIMATION ERRORS FROM VARIOUS MODELS BASED ON MDI CRITERION

*i*th Obs.	$[x_i, \tilde{y}_i]$	MDI criterion					
		DM	KC-TS	KC-LSE	CH-LSE	CH-MP	Proposed CY
1	[1, (6.2, 8.0, 9.8)]	2.680	3.287	2.663	2.760	3.200	2.679
2	[2, (4.2, 6.4, 8.6)]	3.915	3.334	3.876	3.940	3.300	3.539
3	[3, (6.9, 9.5, 12.1)]	1.110	0.613	1.150	1.160	0.400	0.481
4	[4, (10.9, 13.5, 16.1)]	3.495	4.028	3.401	3.121	4.301	2.620
5	[5, (10.6, 13.0, 15.4)]	0.900	0.431	1.010	1.000	0.000	0.434
	Total error	12.100	11.693	12.100	11.981	11.201	9.753

By defuzzifying these five fuzzy observations using the GMI representation of (4) and applying the LS estimates of (5), the derived regression coefficients are $(b_1)_{GMI} = 1.71$ and $(b_0)_{GMI} = 4.95$. The mathematical program of Model (13) is then solved using the mathematical programming solver LINGO [39] to derive the estimates of the error terms $\tilde{E}_i, i = 1,2,\ldots,5$ as (-1.8,0,1.8), (-2.6,0,1.8), (-3.4,0,1.8), (-1.8,0,3.4), and (-3,0,1.8), respectively. Therefore, the fuzzy linear regression model constructed from the proposed method (denoted as CY) is as follows:

$$\tilde{Y}_{\mathrm{CY}} = 4.95 + 1.71X + \tilde{E}_i, \quad i = 1,...,5,$$

where $\tilde{E}_1 = (-1.8, 0, 1.8)$, $\tilde{E}_2 = (-2.6, 0, 1.8)$, $\tilde{E}_3 = (-3.4, 0, 1.8)$, $\tilde{E}_4 = (-1.8, 0, 3.4)$, and $\tilde{E}_5 = (-3, 0, 1.8)$.

Diamond [18] also applied a LSM of minimizing a distance function-Hansdorff metric, which is defined for the minimum, center point, and maximum between the observed and estimated responses, to determine the regression coefficients. The obtained model (denoted as DM) is

$$\tilde{Y}_{\mathrm{DM}} = (3.11,\ 4.95,\ 6.79) + (1.55,\ 1.71,\ 1.87)X.$$

Kao and Chyu [21] adopted a combination of Zadeh's extension principle [40] and Chen and Klein's ranking method and proposed a mathematical program to minimize the error sum of the squares (SSE) between the observed and estimated responses to find the numeric regression coefficients. The fuzzy regression model (denoted as KC-LSE) obtained is

$$\tilde{Y}_{\mathrm{KC\text{-}LSE}} = 4.808 + 1.718X + (-2.202,\ 0.118,\ 2.438).$$

Kao and Chyu [28] proposed a two-stage method to determine the crisp regression coefficients and a fuzzy error term, where the former is derived based on the defuzzification of the COG method and the latter is based on the criterion of Kim and Bishu [33]. Their fuzzy regression model (denoted as KC-TS) is given as

$$\tilde{Y}_{\mathrm{KC\text{-}TS}} = 4.95 + 1.71X + (-3.01,\ 0,\ 1.8).$$

Chen and Hsueh [27] proposed an MP approach based on a criterion of minimizing the distance between the observed and estimated fuzzy responses at certain α -cuts. Later, they improved the approach using the LSM [17]. These two fuzzy regression models (denoted as CH-MP and CH-LSE, respectively) are

$$\tilde{Y}_{\mathrm{CH\text{-}MP}} = 4.75 + 1.65X + (-2.4,\ 0,\ 2.4), \text{ and}$$
$$\tilde{Y}_{\mathrm{CH\text{-}LSE}} = 1.71X + (2.63, 4.95, 7.27).$$

Two performance measures, including the proposed MDI and Chen and Hsueh's distance criterion, are adopted to calculate the estimation errors to compare the performance of these six fuzzy regression models. Table I lists the estimation errors of the fuzzy responses for the models based on the proposed MDI criterion. The total estimation error of each model is listed in the last row. The total error of the proposed method (CY) is 9.753, which clearly smaller than the errors calculated using the other nine models. Table II compares the estimation errors based on Chen and Hsueh's distance criterion. The total estimation error of the proposed model is 5.120 using Chen and Hsueh's distance criterion. As expected, the total estimation error of the proposed model is less than those of the other five models.

In summary, the results listed in Tables I and II indicate that the total errors calculated using the proposed model based on all three criteria are smaller than those calculated from the other models.

TABLE II
THE ESTIMATION ERRORS FROM VARIOUS MODELS BASED ON CHEN AND HSUEH'S DISTANCE CRITERION

ith Obs.	$[x_i, \tilde{y}_i]$	Chen and Hsueh's distance criterion					
		DM	KC-TS	KC-LSE	CH-LSE	CH-MP	Proposed CY
1	[1, (6.2, 8.0, 9.8)]	1.340	1.643	1.356	1.340	1.600	1.340
2	[2, (4.2, 6.4, 8.6)]	1.970	1.668	1.962	1.970	1.650	1.770
3	[3, (6.9, 9.5, 12.1)]	0.580	0.388	0.580	0.580	0.200	0.400
4	[4, (10.9, 13.5, 16.1)]	1.710	2.013	1.702	1.710	2.150	1.310
5	[5, (10.6, 13.0, 15.4)]	0.500	0.303	0.516	0.500	0.000	0.300
	Total error	6.100	6.015	6.116	6.100	5.600	5.120

V. CONCLUSION

This paper proposes an idea for constructing the fuzzy linear regression model with higher explanatory power, where the GMI and LSM is adopted to determine the representative numeric regression coefficients and used the proposed MDI to determine the variable spreads of the fuzzy error terms for each observation. In contrast to Kim and Bishu's criterion, which has been adopted by several previous studies, the concept of dissemblance in the present study can measure the total estimation error even if the estimated and observed responses do not overlap. That is, the proposed MDI measurement reflects the actual estimation error. Furthermore, the proposed MDI is easier to use than Chen and Hsueh's distance criterion, in which the number of α -cuts is difficult to predetermine. An example shows that the proposed fuzzy regression model has high explanatory power and forecasting accuracy. More examples are being tested in progress.

REFERENCES

[1] J. Neter, W. Wasserman, and M. Kutner, *Applied Linear Regression Models*. Homewood, IL: Richard D. Irwin Inc., 1983.
[2] N.R. Draper and H. Smith, *Applied Regression Analysis*, New York: John Wiley & Sons, 1998.
[3] L.A. Zadeh, "Fuzzy sets," *Inf. and Control*, vol. 8, pp. 338–356, 1965.
[4] H. Tanaka, T. Okuda, and K. Asai, "On fuzzy-mathematical programming," *J. Cybern.*, vol. 3, pp. 37–46, 1974.
[5] J. Kacprzyk and M. Fedrizzi, *Fuzzy Regression Analysis*, Warsaw: Omnitech Press, 1992.
[6] S. Ferson and L. R. Ginzburg, "Different methods are needed to propagate ignorance and variability," *Rel. Eng. Syst. Safety*, vol. 54, pp. 133-144, 1996.
[7] D. Dubois and H. Prade, "Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets," *Fuzzy Sets Syst.*, vol. 192, pp. 3-24, 2012.
[8] K.Y. Cai, "System failure engineering and fuzzy methodology: an introductory overview," *Fuzzy Sets Syst.*, vol. 83, pp. 113–133, 1996.
[9] S. Fruehwirth-Schnatter, "On statistical inference for fuzzy data with applications to descriptive statistics," *Fuzzy Sets Syst.*, vol. 50, pp. 143–165, 1992.
[10] H. Tanaka, S. Uejima, and K. Asai, "Fuzzy linear regression model," *IEEE Trans. Syst., Man, Cybern.*, vol. 10, pp. 2933-2938, 1980.
[11] H. Tanaka, S. Uegima, and K. Asai, "Linear regression analysis with fuzzy model," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-12, pp. 903-907, 1982.
[12] B.Wu and N.F. Tseng, "A new approach to fuzzy regression models with application to business cycle analysis," *Fuzzy Sets Syst.*, vol. 130, pp. 33–42, 2002.

ISSN: 2517-9934
Vol:8, No:9, 2014

[13] Y.Q. He, L.K. Chan, and M.L. Wu, "Balancing productivity and consumer satisfaction for profitability: Statistical and fuzzy regression analysis," *Eur. J. Oper. Res.*, vol. 176, pp. 252-263, 2007.
[14] C.Y. Huang, and G.H. Tzeng, "Multiple generation product life cycle predictions using a novel two-stage fuzzy piecewise regression analysis method," *Technol. Forecasting & Social Change*, vol. 75, 12-31, 2008.
[15] S. Imoto, Y. Yabuuchi, and J.Watada, "Fuzzy regression model of R&D project evaluation," *Applied Soft Comput.*, vol. 8, pp. 1266-1273, 2008.
[16] S.J. Mousavia, K. Ponnambalamb, and F. Karray, "Inferring operating rules for reservoir operations using fuzzy regression and ANFIS," *Fuzzy Sets Syst.*, vol. 158, pp. 1064-1082, 2007.
[17] L.H. Chen and C.C. Hsueh, "Fuzzy regression models using the least-squares method based on the concept of distance," *IEEE Trans. Fuzzy Syst.,* vol. 17, pp. 1259-1272, 2009.
[18] P. Diamond, "Fuzzy least squares," *Inf. Sci.*, vol. 46, pp. 141-157, 1988.
[19] P.T. Chang and E.S. Lee, "A generalized fuzzy weighted least-squares regression," *Fuzzy Sets Syst.*, vol. 82, pp. 289-298, 1996.
[20] M. Ma, M. Friedman, and A. Kandel, "General fuzzy least squares," *Fuzzy Sets Syst.*, vol. 88, 107-118, 1997.
[21] C. Kao and C.L. Chyu, "Least-squares estimates in fuzzy regression analysis, *Eur. J. Oper. Res.*, vol. 148, pp. 426-435, 2003.
[22] H. Tanaka, "Fuzzy data analysis by possibilistic linear models," *Fuzzy Sets Syst.*, vol. 24, 363-375, 1987.
[23] H. Tanaka and J. Watada, "Possibilistic linear systems and their application to the linear regression model," *Fuzzy Sets Syst.*, vol. 27, pp. 275-289, 1988.
[24] H. Tanaka, I. Hayashi, and J. Watada, "Possibilistic linear regression analysis for fuzzy data," *Eur. J. Oper. Res.*, vol. 40, 389-396, 1989.
[25] D. Savic andW. Pedrycz, "Evaluation of fuzzy linear regression models," *Fuzzy Sets Syst.*, vol. 39, 51–63, 1991.
[26] C. Kao and C.L. Chyu, "A fuzzy linear regression model with better explanatory power," *Fuzzy Sets Syst.*, vol. 126, 401-409, 2002.
[27] L.H. Chen and C.C. Hsueh, "A mathematical programming method for formulating a fuzzy regression model based on distance criterion," *IEEE Trans. Syst., Man, Cybern. B, Cybern.,* vol. 37, 705-711, 2007.
[28] S.P. Chen and J.F. Dang, "A variable spread fuzzy linear regression model with higher explanatory power and forecasting accuracy," *Inf. Sci.* vol. 178, 3973–3988, 2008.
[29] D.H. Hong and C. Hwang, "Interval regression analysis using quadratic loss support vector machine," *IEEE Trans. Fuzzy Syst.,* vol. 13, 229–237, 2005.
[30] P.Y. Hao and J.H. Chiang, "Fuzzy regression analysis by support vector learning approach," *IEEE Trans. Fuzzy Syst.,* vol. 16, 428–441, 2008.
[31] C. Kao and P.H. Lin, "Entropy for fuzzy regression analysis," *Int. J. Syst. Sci.*, vol. 36, 869-876, 2005.
[32] D.T. Redden and W.H. Woodall, "Properties of certain fuzzy linear regression methods," *Fuzzy Sets Syst.*, vol. 64, 361-375, 1994.
[33] B. Kim and R.R. Bishu, "Evaluation of fuzzy linear regression models by comparing membership functions," *Fuzzy Sets Syst.*, vol. 100, 343-352, 1998.
[34] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. International Edition, Taiwan: Prentice-Hall, 2002.
[35] S.H. Chen and C.H. Hsieh, "Graded mean integration representation of generalized fuzzy numbers," *J. Chinese Fuzzy Syst. Association*, vol. 5, 1–7, 1991.
[36] S.H. Chen and C.H. Hsieh, "Representation, ranking, distance, and similarity of L-R type fuzzy number and application," *Australian J Intell. Process. Syst.*, vol. 6, 217–229, 2000.
[37] A. Kaufmann and M.M. Gupta, *Introduction to Fuzzy Arithmetic Theory and Applications.* Boston: International Thomson Computer Press, 1991.
[38] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Second ed. Singapore: JohnWiley& Sons, 1993.
[39] *LINGO User's Guide*, LINDO Syst. Inc., Chicago, IL, 2003.
[40] L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets Syst.*, vol. 1, 3-28, 1978.