Necessity Analysis of Fuzzy Regression Equations Using a Fuzzy Goal Programming Model

Ruey-Chyn Tsaur and Hsiao-Fan Wang

Abstract

In this study, using necessity analysis, we located a fuzzy regression interval that is possibly included in collected interval data by reducing the effect of fluctuating points. First, we developed a fuzzy regression model with a regression interval that is close to the sum of the radius values of collected interval data. Next, we enlarged the feasible region so that the fuzzy regression interval may be possibly included in the collected data by fuzzifying the constraints of the fuzzy regression model within a given tolerance value. Finally, we derived a satisfactory fuzzy regression equation with maximum possibility. The application of the proposed model is illustrated by means of an example.

Keywords*: *Fuzzy regression model, fuzzy data, necessity analysis, fuzzy goal programming, tolerance value.

1. Introduction

The process of fuzzy regression through possibility analysis was first studied by Tanaka et al. [12] in 1982 to propose an alternative approach to evaluate the fuzzy relationship between independent and dependent variables; subsequently, necessity analysis was also clearly defined [7]. A considerable amount of research in various fields has focused on using possibility analysis of a fuzzy regression model, including model extensions [3] [6] [8-9] [11] [14] [18-23], business forecasting [15-16], and engineering [1][4]. However, necessity analysis in a fuzzy regression model usually leads to an infeasible solution owing to large variability in the collected data or large fluctuations in the given data. Yu et al. [22-23] proposed a piecewise model to cope with such problems, but it is still difficult to set the change points. In addition, setting the change points to derive a piecewise fuzzy regression model does not lead to a useful model, because we still do not know where to set the next change point. In other words, it is more important to obtain a trend from the collected data to determine where the change point will be.

In our present study, to determine trends from variable collected data using the maximum possible amount of information, we propose a linear fuzzy regression model with a fuzzy interval by using a fuzzy goal programming model to reduce the effect of fluctuating points. First, we determine an interval of the fuzzy linear regression model that is as close as possible to the sum of the radius values of the collected interval data. Next, in order to obtain the optimal solution for the fuzzy regression model, we ensure that the fuzzy regression interval can possibly be included in the collected data so as to fuzzify the constraints of the fuzzy regression model within a given tolerance value.

The remainder of this paper is organized as follows. In Section 2, we provide reviews of (1) the fuzzy regression model using necessity analysis and (2) non-preemptive fuzzy goal programming. In Section 3, we propose an extension to the fuzzy regression model using the fuzzy goal programming model. In Section 4, we provide an example to illustrate our approach. Finally, in Section 5, we draw our conclusions.

2. Review of Fuzzy Regression Model and Fuzzy Goal Programming Model

In this section, we introduce the fuzzy regression model and the fuzzy goal programming model in order to derive our proposed fuzzy regression model under necessity analysis.

2.1. Fuzzy Regression Using Necessity Analysis

The fuzzy regression model assumes that a linear interval regression equation is represented as given below:

Corresponding Author: Ruey-Chyn Tsaur is with the Department of Management Sciences and Decision Making, Tamkang University, 151 Ying-chuan Road, Tamsui, Taipei, Taiwan, 25137.
E-mail: rctsaur@yahoo.com.tw
Manuscript received 25 Nov. 2008; revised 8 May 2009; accepted 19 June 2009.

$$Y_i^* = A_0 + A_1 X_{i1} + ... + A_N X_{iN} \quad (1)$$

where Y_i^* is the predicted interval corresponding to theinput vector $X_i = [X_{i1}, \dots, X_{iN}]^T$ of independent variables for the i^{th} data point; $[A_0, A_1, \dots, A_N]$ is a vector of interval parameter A_j, which is given by

$$A_j = (a_{cj}, a_{wj}) = \{a_j : a_{cj} - a_{wj} \le a \le a_{cj} + a_{wj}\}, \quad (2)$$

where a_{cj} is the center and a_{wj} is the radius of A_j, j=0, 1,…, N. Then, addition and multiplication operations involving intervals A and B are carried out as follows:

$$A + B = (a_c, a_w) + (b_c, b_w) = (a_c + b_c, a_w + b_w) \quad (3)$$

$$AB = (a_c, a_w)(b_c, b_w) = [a_c - a_w, a_c + a_w][b_c - b_w, b_c + b_w] = (a_c b_c + a_w b_w, a_c b_w + a_w b_c) \quad (4)$$

$$rA = r(a_c, a_w) = (ra_c, ra_w) \quad (5)$$

where $a_c \ge a_w \ge 0$, $b_c \ge b_w \ge 0$, and r is a real number. By interval arithmetic, the basic linear interval regression (1) can be represented as follows:

$$Y_i^* = (a_{0c}, a_{0w}) + (a_{1c}, a_{1w})X_{i1} + ... + (a_{Nc}, a_{Nw})X_{iN} = (Y_{ic}^*, Y_{iw}^*) \quad (6)$$

where $Y_{ic}^* = a_{0c} + a_{1c}X_{i1} + ... + a_{Nc}X_{iN}$, and $Y_{iw}^* = a_{0w} + a_{1w}X_{i1} + ... + a_{Nw}X_{iN}$.

Then, necessity analysis of the interval regression equation is satisfied by ensuring that the predicted intervals are included in the collected interval values as follows:

$$Y_i^* \subseteq \breve{y}_i, \quad i = 1, 2, ..., M \quad (7)$$

where $\breve{y}_i$ is the collected interval value with radius value e_i; consequently, its lower bound $\breve{y}_{iL} = y_i - e_i$ and upper bound $\breve{y}_{iR} = y_i + e_i$. Let predicted interval Y_i^* belong to collected interval values $\breve{y}_i$ as fit as possible; necessity analysis is then used to maximize the radius of the predicted interval as follows:

$$\begin{aligned} &\text{Maximize } Y_{1w}^* + Y_{2w}^* + ... + Y_{Mw}^* \\ &\text{Subject to } Y_i^* \subseteq \breve{y}_i, i = 1, 2, ..., M \\ &\qquad a_{iw} \ge 0, i = 0,1,2,...., M; j = 1,2,...., N. \end{aligned} \quad (8)$$

This linear programming problem can be rewritten as follows:

$$\begin{aligned} &\text{Maximize} \sum_{i=1}^{M} \left(a_{0w} + \sum_{j=1}^{N} a_{jw} |X_{ij}| \right) \\ &s.t. \left(a_{0c} + \sum_{j=1}^{N} a_{jc} X_{ij} \right) - \left(a_{0w} + \sum_{j=1}^{N} a_{jw} |X_{ij}| \right) \ge \breve{y}_{iL}, \ i = 1, 2, ..., M \\ &\quad \left(a_{0c} + \sum_{j=1}^{N} a_{jc} X_{ij} \right) + \left(a_{0w} + \sum_{j=1}^{N} a_{jw} |X_{ij}| \right) \le \breve{y}_{iR}, \ i = 1, 2, ..., M \\ &\quad a_{jc} \in R, a_{jw} \ge 0, \ j = 0, 1, 2, ..., N \end{aligned} \quad (9)$$

By solving (9), parameters a_{jc}, a_{jw}, $\forall j = 0, 1, 2, ..., N$, can be obtained; then, (1) can be rewritten as follows:

$$Y_i^* = (a_{0c}, a_{0w}) + (a_{1c}, a_{1w})X_{i1} + ... + (a_{Nc}, a_{Nw})X_{iN} \quad (10)$$

Each predicted value Y_i^* of the dependent variable can be estimated as an interval number $Y_i^* = [Y_{iL}^*, Y_{iR}^*]$, $i = 1,2,..., M$, where the lower bound of Y_i^* is $Y_{iL}^* = (a_{0c} - a_{0w}) + (a_{1c} - a_{1w})X_{i1} + ... + (a_{Nc} - a_{Nw})X_{iN}$, and the upper bound of Y_i^* is $Y_{iR}^* = (a_{0c} + a_{0w}) + (a_{1c} + a_{1w})X_{i1} + ... + (a_{Nc} + a_{Nw})X_{iN}$.

2.2. Fuzzy Goal Programming Model

Goal programming (GP) is an appropriate alternative for modeling real-world decision problems, and it has been used extensively in previous studies that involve obtaining solutions to decision-making problems. However, a major limitation of GP is the imprecision of the goal. Therefore, in fuzzy goal programming (FGP), fuzzy goals are considered at imprecise levels [2][5]. The FGP approach was first introduced by Narasimhan [10] modeled FGP by solving the set of 2^k linear programming problems with equal and unequal fuzzy weights under the assumption of linear membership functions, each of which contains 3k constraints, where k denotes the number of goals in the original problem. Later, Hannan [17] simplified the procedure to formulate a single LP problem with 2k goal-related constraints that can be preemptive or non-preemptive. Therefore, when fuzzy goals are presented as "essentially equal to b," a FGP problem can be written in a general form as follows:

$$\begin{aligned} &Ax \cong b \\ s.t. \quad &Cx \le d, \\ &x \ge 0 \end{aligned} \quad (11)$$

where x is an $n \times 1$ alternative set, A is a $k \times n$ matrix of coefficients of the objective function, C is a $k\times n$ matrix of coefficients of the constraints, and d is right-hand side with a $k\times 1$ matrix. The membership function for fuzzy goals is postulated in the triangular form as follows, i=1, 2,…, M:

$$u_i(AX)=\begin{cases}1, & (AX)_i=b_i\\ \left[(b_i+p_i)-(AX)_i\right]\big/p_i, & b_i\le (AX)_i\le b_i+p_i\\ \left[(AX)_i-(b_i-p_i)\right]\big/p_i, & b_i-p_i\le (AX)_i\le b_i\\ 0, & ow.\end{cases}\tag{12}$$

When the decision is made by fulfilling all the fuzzy goals to the greatest degree, the max-min operator is used such that the optimal decision D is obtained by the following expression:

$$u_D(x)=\underset{x}{Max}\,\underset{i}{Min}\;u_i(AX)\tag{13}$$

For a FGP problem, Hannan [5] proved that if λ_i is the optimal solution to subproblem i with $u_i(AX)\ge\lambda_i$, then there exists $\lambda^*=\underset{i=1,2,\ldots,M}{Max}\lambda_i$ such that the optimal solution in (9) will be equal to λ^*. Therefore, model (11) can be written as follows:

$$\begin{aligned}&Max\ \lambda\\ &s.t.\frac{(AX)_i}{p_i}-d_i^++d_i^-=\frac{e_i}{p_i}\ \forall i=1,2,\ldots,M\\ &Cx\le b\\ &\lambda+d_i^++d_i^-\le 1,\quad \forall i=1,2,\ldots,M\\ &\lambda\in[0,1]\\ &0\le d_i^+,d_i^-\le 1,\ ;\forall i=1,2,\ldots,M,\end{aligned}\tag{14}$$

However, the membership function of Hannan's model is computationally complicated, and Narasimhan's model does not take into account the priorities of fuzzy goals. Therefore, Tiwari et al. [12] proposed an algorithm for solving a FGP problem using symmetrical triangular membership functions of fuzzy goals with priority structure. Still later, Chen [2] and Wang and Fu [17] proposed more efficient methods for determining the properties of the models.

3. Fuzzy Regression Model Using FGP Approach

As described in Section 2, in necessity analysis, the fuzzy regression model is derived by maximizing the total radiuses $\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)$ for the predicted fuzzy regression interval. If a decision maker (DM) desires to obtain a maximum fuzzy regression interval, then the level of radiuses $\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)$ should be as close as possible to the sum of radiuses $\sum_{i=1}^{M}e_i$ of the collected interval data $\breve{y}_i,\forall i=1,2,\ldots,M$; thus, we can formulate the imprecise goal as $\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)\cong\sum_{i=1}^{M}e_i$. Let p_1 denote the tolerance value for fuzzy goal $\sum_{i=1}^{M}e_i$; then, the fuzzy goal can be described by a triangular membership function as per (15). In order to obtain the maximum membership degree for the goal, (15) must be greater than membership level λ_1 with $u\left(\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)\right)\ge\lambda_1$. Then,

$$u\left(\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)\right)=\begin{cases}1, & \sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)=\sum_{i=1}^{M}e_i\\ \dfrac{(\sum_{i=1}^{M}e_i+p_1)-\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)}{p_1}, & \sum_{i=1}^{M}e_i\le\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)\le\sum_{i=1}^{M}e_i+p_1\\ \dfrac{\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)-(\sum_{i=1}^{M}e_i-p_1)}{p_1}, & \sum_{i=1}^{M}e_i-p_1\le\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)\le\sum_{i=1}^{M}e_i\\ 0, & ow\end{cases}\tag{15}$$

Therefore, we rewrite (9) as (16) below:

$$\begin{aligned}&\text{Maximize}\,\lambda_1\\ &s.t.\frac{\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)}{p_1}-(1-\lambda_1)\le\frac{\sum_{i=1}^{M}e_i}{p_1}\\ &\left(a_{0c}+\sum_{j=1}^{N}a_{jc}X_{ij}\right)-\left(a_{0w}+\sum_{j=1}^{N}a_{jw}|X_{ij}|\right)\ge\breve{y}_{iL},\ i=1,2,\ldots,M\\ &\left(a_{0c}+\sum_{j=1}^{N}a_{jc}X_{ij}\right)+\left(a_{0w}+\sum_{j=1}^{N}a_{jw}|X_{ij}|\right)\le\breve{y}_{iR},\ i=1,2,\ldots,M\\ &a_{jc}\in R, a_{jw}\ge 0, 0\le\lambda_1\le 1\ j=0,1,2,\ldots,N\end{aligned}\tag{16}$$

Although the constraints of (9) are relaxed to formulate the fuzzy goal by requiring the radiuses $\sum_{i=1}^{M}\left(\sum_{j=0}^{N}a_{jw}|X_{ij}|\right)$ of the predicted values to be possibly equal to $\sum_{i=1}^{M} e_i$, the feasible region is still not enlarged; therefore, (16) is usually infeasible. We cannot obtain a fuzzy regression equation that is fully included in the collected interval data because of variability or large fluctuations among the given data. In order to derive a feasible solution based on these fluctuating data, we must modify the constraints in (16) so that constraints containing $\geq$ and $\leq$ are relaxed to obtain fuzzy constraints containing $\underset{\sim}{\geq}$ and $\underset{\sim}{\leq}$, respectively; then, a larger feasible region is available [24]. Now, (16) can be rewritten as follows:

$$\begin{aligned}
&\text{Maximize}\ \lambda_1 \\
&s.t.\ \frac{\sum_{i=1}^{M}\left(\sum_{j=0}^{N} a_{jw}\left|X_{ij}\right|\right)}{p_1} - (1-\lambda_1) \leq \frac{\sum_{i=1}^{M} e_i}{p_1} \\
&\left(a_{0c} + \sum_{j=1}^{N} a_{jc}X_{ij}\right) - \left(a_{0w} + \sum_{j=1}^{N} a_{jw}\left|X_{ij}\right|\right) \underset{\sim}{\geq} \breve{y}_{iL},\ i = 1, 2, ..., M \\
&\left(a_{0c} + \sum_{j=1}^{N} a_{jc}X_{ij}\right) + \left(a_{0w} + \sum_{j=1}^{N} a_{jw}\left|X_{ij}\right|\right) \underset{\sim}{\leq} \breve{y}_{iR},\ i = 1, 2, ..., M \\
&a_{jc} \in R, a_{jw} \geq 0, 0 \leq \lambda_1 \leq 1\ j = 0, 1, 2, ..., N
\end{aligned} \tag{17}$$

where fuzzy constraints $\underset{\sim}{\geq}$ and $\underset{\sim}{\leq}$ are shown, respectively, in Figure 1 and Figure 2 with membership functions defined by (18) and (19).

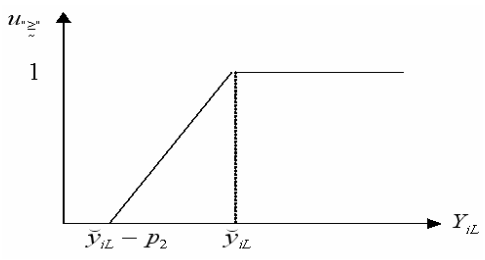

Figure 1. Membership function of $\underset{\sim}{\geq}$.

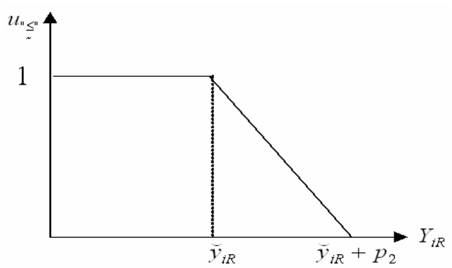

Figure 2. Membership function of $\underset{\sim}{\leq}$.

$$u_{"\underset{\sim}{\geq}"} = \begin{cases} 1, & \text{for } Y_{iL}^* \geq \breve{y}_{iL} \\ 1 - \dfrac{\breve{y}_{iL} - Y_{iL}^*}{p_2}, & \text{for } \breve{y}_{iL} - p_2 \leq Y_{iL}^* \leq \breve{y}_{iL} \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

$$u_{"\underset{\sim}{\leq}"} = \begin{cases} 1, & \text{for } Y_{iR}^* \leq \breve{y}_{iR} \\ 1 - \dfrac{Y_{iR}^* - \breve{y}_{iR}}{p_2}, & \text{for } \breve{y}_{iR} \leq Y_{iR}^* \leq \breve{y}_{iR} + p_2 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

In (18), the tolerance value p_2 suggests that the predicted lower bound Y_{iL}^* is possibly less than the lower bound of the collected data $\breve{y}_{iL}$; a membership degree of 1 means that the lower bound Y_{iL}^* is absolutely larger than $\breve{y}_{iL}$ and that the lower bound Y_{iL}^* could be less than $\breve{y}_{iL}$ with a membership degree of at least λ_2 under p_2. In addition, in (19), p_2 suggests that the predicted upper bound Y_{iR}^* is possibly larger than the upper bound of the collected data $\breve{y}_{iR}$. Therefore, in this case, a membership degree of 1 means that the upper bound Y_{iR}^* is absolutely less than $\breve{y}_{iR}$ and that Y_{iR}^* could be larger than $\breve{y}_{iR}$ with a membership degree of at least λ_2 under p_2. Thus, we can obtain a linear programming model as follows:

$$\begin{aligned}
&\text{Maximize}\ \min(\lambda_1, \lambda_2) \\
&s.t.\ \frac{\sum_{i=1}^{M}\left(\sum_{j=0}^{N} a_{jw}\left|X_{ij}\right|\right)}{p_1} - (1-\lambda_1) \leq \frac{\sum_{i=1}^{M} e_i}{p_1} \\
&\left(a_{0c} + \sum_{j=1}^{N} a_{jc}X_{ij}\right) - \left(a_{0w} + \sum_{j=1}^{N} a_{jw}\left|X_{ij}\right|\right) - \lambda_2 p_2 \geq \breve{y}_{iL} - p_2,\ i = 1, 2, ..., M \\
&\left(a_{0c} + \sum_{j=1}^{N} a_{jc}X_{ij}\right) + \left(a_{0w} + \sum_{j=1}^{N} a_{jw}\left|X_{ij}\right|\right) + \lambda_2 p_2 \leq \breve{y}_{iR} + p_2,\ i = 1, 2, ..., M \\
&a_{jc} \in R, a_{jw} \geq 0, 0 \leq \lambda_1 \leq 1, 0 \leq \lambda_2 \leq 1, j = 0, 1, 2, ..., N
\end{aligned} \tag{20}$$

Theorem 1. If (20) is infeasible, given a larger value $p_i, \forall i = 1,2,$ then an optimal solution of (20) absolutely exists.

Proof

(1) For a special case, we assume that $\sum_{i=1}^{M}\left(\sum_{j=0}^{N} a_{jw}X_{ij}\right) = \sum_{i=1}^{M} e_i$ with $\lambda_1 = 1$. If $\breve{y}_{iL} \leq Y_{iL}^*$, $Y_{iR}^* \leq \breve{y}_{iR}$, and $\forall i = 1, 2, ..., M$, then $\lambda_2 = 1$, which means that the derived fuzzy regression equation is absolutely in the collected interval data.

(2) (i) Without losing generality, given a suitable value p_2, we assume that $\sum_{i=1}^{M}\left(\sum_{j=0}^{N} a_{jw} X_{ij}\right) \le \sum_{i=1}^{M} e_i$ with $\lambda_1 \le 1$, $\breve{y}_{iL} - p_2 \le Y_{iL}^* \le \breve{y}_{iL}$, and $Y_{iR}^* \le \breve{y}_{iR}$ where $\forall i = 1, 2, ..., M$; then,

$$\lambda_2 = \min\left(1 - \frac{\breve{y}_{1L} - Y_{iL}^*}{p_2}, ..., 1 - \frac{\breve{y}_{iL} - Y_{iL}^*}{p_2}, ..., 1 - \frac{\breve{y}_{ML} - Y_{ML}^*}{p_2}\right) \le 1$$

, which means the derived fuzzy regression equation is possibly included in the collected interval data.

(ii) Otherwise, $\forall i = 1, 2, ..., k, ..., M \ni Y_{kL}^*$ s.t. $Y_{kL}^* \le \breve{y}_{kL} - p_2$. Then, λ_2 is negative, which leads to an infeasible solution. If we reset a larger value of p_2' s.t. $\breve{y}_{iL} - p_2' \le Y_{iL}^* \le \breve{y}_{iL}$, then we can obtain

$$\lambda_2 = \min\left(1 - \frac{\breve{y}_{1L} - Y_{iL}^*}{p_2'}, ..., 1 - \frac{\breve{y}_{iL} - Y_{iL}^*}{p_2'}, ..., 1 - \frac{\breve{y}_{ML} - Y_{ML}^*}{p_2'}\right) \le 1 ,$$

and an optimal solution is available.

In addition, by introducing a slack variable d_i^- as a negative deviation variable, assuming $(1 - \lambda_1) = d_i^+$ to be a positive deviation variable in the first constraint of (20), and substituting $\lambda = \min(\lambda_1, \lambda_2)$, we have $(1 - d_i^+) = \lambda_1 \ge \lambda$. Introducing a slack variable, d_i^-, implies that $\lambda + d_i^+ + d_i^- = 1$. Thus, (20) can be rewritten as (21):

$$\begin{aligned}
&\text{Maximize } \lambda \\
&s.t. \frac{\sum_{i=1}^{M}\left(\sum_{j=0}^{N} a_{jw}|X_{ij}|\right)}{p_1} - d_i^+ + d_i^- = \frac{\sum_{i=1}^{M} e_i}{p_1} \\
&\left(a_{0c} + \sum_{j=1}^{N} a_{jc} X_{ij}\right) - \left(a_{0w} + \sum_{j=1}^{N} a_{jw}|X_{ij}|\right) - \lambda p_2 \ge \breve{y}_{iL} - p_2,\ i = 1, 2, ..., M \\
&\left(a_{0c} + \sum_{j=1}^{N} a_{jc} X_{ij}\right) + \left(a_{0w} + \sum_{j=1}^{N} a_{jw}|X_{ij}|\right) + \lambda p_2 \le \breve{y}_{iR} + p_2,\ i = 1, 2, ..., M \\
&\lambda + d_i^+ + d_i^- = 1 \\
&a_{jc} \in R, a_{jw} \ge 0, 0 \le \lambda \le 1, j = 0, 1, 2, ..., N
\end{aligned} \tag{21}$$

The meaning of the objective value λ is that a satisfactory level for the given tolerance values of predictive value Y_i^* can *possibly* be included in the collected interval data $\breve{y}_i$ under the given tolerance values p_1 and p_2. A larger λ does not imply a better solution but merely allows for the possibility than an optimal solution can be obtained under the given tolerance values. Therefore, Peter [10] suggested that choosing larger tolerance values p_1 and p_2 can contribute to obtaining an optimal solution in (21). In contrast, if the selected tolerance values still make (21) infeasible, then larger tolerance values should be chosen in order to obtain the optimal solution for (21).

4. Illustrated Example

4.1. Example Using Proposed Model

Interval data are collected as follows: $\{x_i, \breve{y}_i\}$ = $\{x_i, [\breve{y}_{iL}; \breve{y}_{iR}]\}$ = (3; [12; 17]); (6; [10; 13]); (9; [13; 18]); (12; [14; 18]); (15; [19; 24]); (18; [16; 19]). In this illustrated example, Tanaka and Ishibuchi [7] used this data set of the linear necessity model but were unable to obtain a feasible solution. In modeling a fuzzy regression model, we first apply our proposed (21), and we obtain the solutions for the variables in Table 1 under different tolerance values. In Table 1, it is obvious that variable a_{0w} is affected by different fuzzy regression intervals, while the other variables remain constant, exhibiting near identical trends for different tolerance values. In addition, objective values for λ are listed in the final column of Table 1. Thus, by definition, λ provides a satisfactory level for given tolerance values. A larger λ does not imply a better solution but merely allows for the possibility that the optimal solution can be obtained with the given tolerance values. Therefore, choosing suitable tolerance values can contribute to obtaining a different optimal solution using (21).

Table 1. Solutions for variables.

Variables	a_{0c}	a_{0w}	a_{1c}	a_{1w}	λ
Solution with values $p_1 = 1, p_2 = 3$	10.75	1.9342	0.5	0	0.1053
Solution with values $p_1 = 1, p_2 = 5$	10.75	1.9919	0.5	0	0.4516
Solution with values $p_1 = 1, p_2 = 7$	10.75	2.0174	0.5	0	0.6047
Solution with values $p_1 = 1, p_2 = 9$	10.75	2.0318	0.5	0	0.6909
Solution with values $p_1 = 3, p_2 = 3$	10.75	1.6786	0.5	0	0.1905
Solution with values $p_1 = 9, p_2 = 3$	10.75	1.1389	0.5	0	0.3704
Solution with values $p_1 = 15, p_2 = 3$	10.75	0.7955	0.5	0	0.4848
Solution with values $p_1 = 20, p_2 = 3$	10.75	0.5921	0.5	0	0.5526

In order to determine the reliability and validity of the proposed model, we use suitable tolerance values $p_1 < p_2$ and $p_1 \ge p_2$ together with their necessity areas, as listed in Table 2 and Table 3. From Table 2, Figure 3, and Figure 4, it is evident that a larger p_2 leads to a larger fuzzy regression interval. Thus, by maintaining a constant value of p_1, a given tolerance value p_2 leads to a predicted lower bound Y_{iL}^* in the fuzzy regression interval that is possibly less than the lower bound of the collected data y_{iL} or to a predicted upper bound Y_{iR}^* in the fuzzy regression interval that is possibly larger than the upper bound of the collected data y_{iR}. Finally, from Table 3, Figure 5, and Figure 6, it is evident that a larger p_1 leads to a smaller regression interval. Therefore, if p_2 is maintained at a constant value, then the tolerance value p_1 is used to derive the tolerance difference between the radiuses $\sum_{i=1}^{M}\left(\sum_{j=0}^{N} a_{jw}\left|X_{ij}\right|\right)$ of the predicted values and the radiuses $\sum_{i=1}^{M} e_i$ of the collected data. In the illustrated example, Figures 3–6 show that when $p_2 > p_1$, the proposed model can be used to obtain a fuzzy regression equation in which the radiuses of the predicted values are as possibly fit as the radiuses of the collected data with a larger p_2. In that case, the predicted lower bound or upper bound in the fuzzy regression interval is possibly lesser than or greater than the corresponding bounds of the collected data with a smaller p_1. Therefore, in order to obtain a fuzzy regression interval that is as near as the radius values of the collected data, we must choose a smaller p_1 value (e.g., $p_1 = 1$) and a larger p_2 value (e.g., $p_2 = 9$) to obtain the optimal predictive fuzzy regression equation.

Table 2. Forecast intervals with tolerance values $p_1 < p_2$.

data i	Forecast for $p_1 = 1, p_2 = 3$	Forecast for $p_1 = 1, p_2 = 5$	Forecast for $p_1 = 1, p_2 = 7$	Forecast for $p_1 = 1, p_2 = 9$
1	[10.316, 14.184]	[10.258, 14.242]	[10.233, 14.267]	[10.218, 14.282]
2	[11.816, 15.684]	[11.758, 15.742]	[11.733, 15.767]	[11.718, 15.782]
3	[13.316, 17.184]	[13.258, 17.242]	[13.233, 17.267]	[13.218, 17.282]
4	[14.816, 18.684]	[14.758, 18.742]	[14.733, 18.767]	[14.718, 18.782]
5	[16.316, 20.184]	[16.258, 20.242]	[16.233, 20.267]	[16.218, 20.282]
6	[17.816, 21.684]	[17.758, 21.742]	[17.733, 21.767]	[17.718, 21.782]

Table 3. Forecast intervals with tolerance values $p_1 \geq p_2$.

data i	Forecast for $p_1 = 3, p_2 = 3$	Forecast for $p_1 = 9, p_2 = 3$	Forecast for $p_1 = 15, p_2 = 3$	Forecast for $p_1 = 20, p_2 = 3$
1	[10.571, 13.929]	[11.111, 13.389]	[11.455, 13.046]	[11.658, 12.842]
2	[12.071, 15.429]	[12.611, 14.889]	[12.955, 14.546]	[13.158, 14.342]
3	[13.571, 16.929]	[14.111, 16.389]	[14.455, 16.046]	[14.658, 15.842]
4	[15.071, 18.429]	[15.611, 17.889]	[15.955, 17.546]	[16.158, 17.342]
5	[16.571, 19.929]	[17.111, 19.389]	[17.455, 19.046]	[17.658, 18.842]
6	[18.071, 21.429]	[18.611, 20.889]	[18.955, 20.546]	[19.158, 20.342]

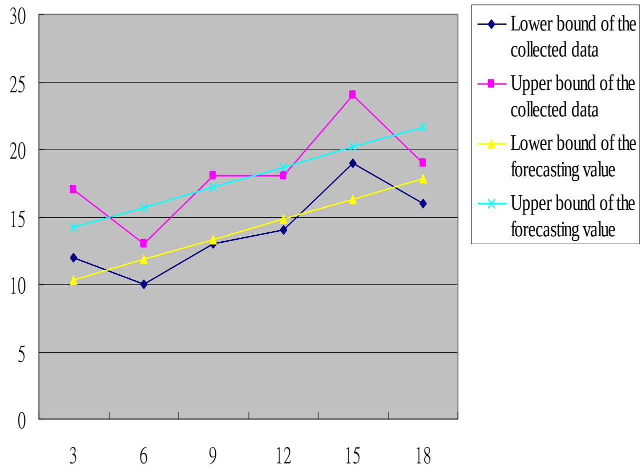

Figure 3. Forecasting results with $p_1 = 1$ and $p_2 = 3$.

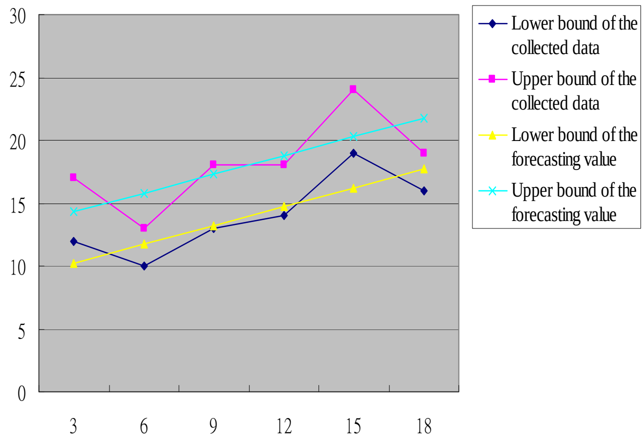

Figure 4. Forecasting results with $p_1 = 1$ and $p_2 = 9$.

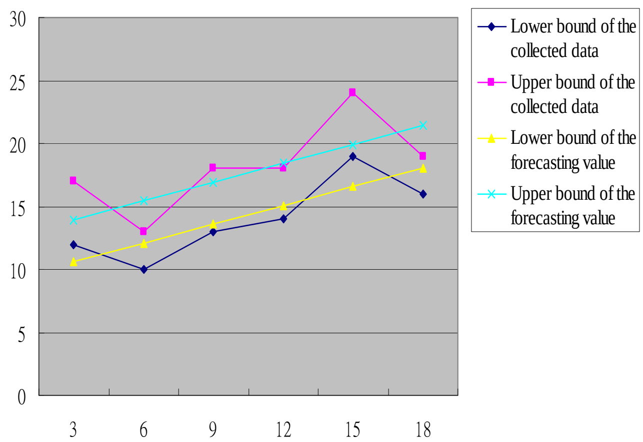

Figure 5. Forecasting results with $p_1 = 3$ and $p_2 = 3$.

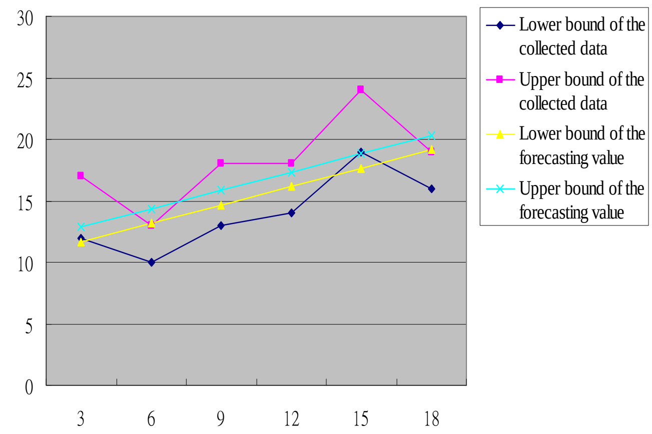

Figure 6. Forecasting results with $p_1 = 20$ and $p_2 = 3$.

4.2. Comparisons and Analysis

For comparison to our proposed model, we use the model proposed by Yu et al [22]. to derive a piecewise fuzzy regression equation by setting five and two change points for modeling, as shown in Figure 7 and Figure 8, respectively. It is clear that different numbers of change points lead to different bands of fuzzy regression intervals, and we still do not know how to determine the optimal change points. In addition, we see that the greater the number of change points, the better the derived piecewise fuzzy regression equation fits the collected data. However, we cannot be absolutely certain that an equation fitted to training data will have improved extrapolative ability. In fact, we would need to determine the location of the next change point for forecasting every time. On the other hand, our proposed model derives most of the information from the collected data, and it suggests suitable tolerance values and a well-fitting linear trend by which to determine future values. A comparison of three different methods is shown in Table 4.

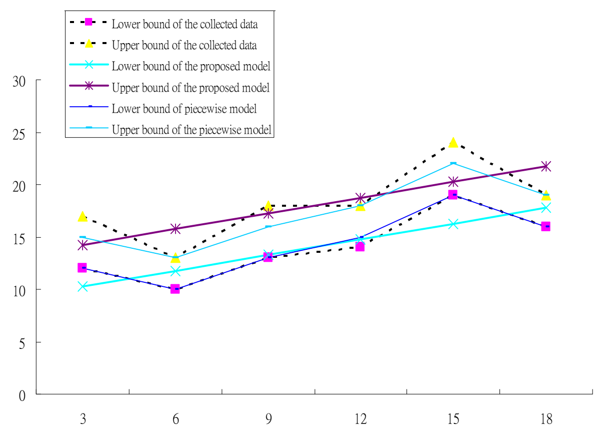

Figure 7. Comparison results to the piecewise model with 5 change points.

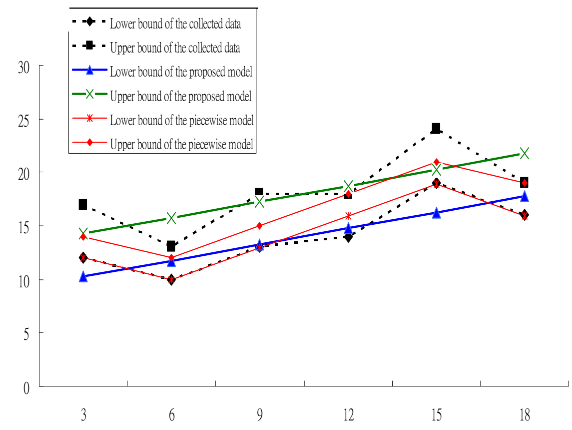

Figure 8. Comparison to the piecewise model with two change points.

Table 4. Comparison of three methods.

Tanaka & Ishibuchi 's Model	**Yu et al.'s Model**	**Our Proposed Model**
Necessity analysis	Necessity analysis using change points for modeling piecewise fuzzy regression equation.	Necessity analysis using fuzzy goal programming model with fuzzy interval.
Usually infeasible	Do not know how to determine change points	Tolerance values are required to be determined.
	Piecewise model cannot easily be used for extrapolation.	Linear model is feasible for forecasting with a possible forecasting interval.

5. Conclusion

In this study, based on the concept of necessity analysis, we proposed a fuzzy regression equation in which the fuzzy regression interval is included in the collected fuzzy data to a desired degree. Because the regression interval must be as near as possible to the sum of the radius values of the collected interval data, the developed fuzzy goal programming (FGP) model enlarges the feasible region by fuzzifying the constraints of the fuzzy regression model with a given tolerance value. Moreover, we derived a satisfactory fuzzy regression equation with maximum possibility. As a result, the fuzzy regression equation can be obtained by fitting the trend of the collected data absolutely from the general data, thereby reducing the effects of large variability in data or of large fluctuations in the given data.

Acknowledgment

The authors acknowledge financial support from the National Science Council of Taiwan, R.O.C., under project number NSC 92-2213-E-364-001.

References

[1] P. T. Chang, S. A. Konz, and E. S. Lee, "Applying fuzzy linear regression to VDT legibility," *Fuzzy Sets and Systems*," vol. 80, no. 22, pp. 197-204, 1996.

[2] H. K. Chen, "A note on a fuzzy goal programming algorithm by Tiwari, Dharmar and Rao," *Fuzzy Sets and Systems*, vol. 62, no. 33, pp. 287-290, 1994.

[3] C. C. Chuang, J. T. Jeng, and C. W. Tao, "Two-Stages Support Vector Regression for Fuzzy Neural Networks with Outliers," *International Journal of Fuzzy Systems*, vol. 11, no. 1, pp. 20-28, 2009.

[4] R. Y. K. Fung, Y. Chen, and J. Tang, "Estimating the functional relationships for quality function deployment under uncertainties," *Fuzzy Sets and Systems* , vol. 157, no. 1, pp. 98-120, 2006.

[5] E. L. Hannan, "On fuzzy goal programming," *Decision Sciences*, vol. 12, no. 3, pp.522-531, 1981.

[6] P. Y. Hao and J. H. Chiang, "A Fuzzy model of support vector regression machine," *International Journal of Fuzzy Systems,* vol. 9, no. 1, pp. 45-50, 2007.

[7] J. Kacprzyk and M. Fedrizzi, Fuzzy Regression Analysis, Omnitech Press, Warsaw, Poland, 1992.

[8] K. J. Kim, H. Moskowitz, and M. Koksalan, "Fuzzy versus statistical linear regression," *European Journal of Operational Research*, vol. 92, no. 2, pp. 417-434, 1996.

[9] K. S. Kula and A. Apaydin, "Fuzzy robust regression analysis based on the ranking of fuzzy sets," *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, vol. 16, no. 5, pp. 663-681, 2008.

[10] R. Narasimhan, "Goal programming in a fuzzy environment," *Decision Sciences*, vol. 11, no. 2, pp. 325-336, 1980.

[11] G. Peters, "Fuzzy linear regression with fuzzy intervals," *Fuzzy Sets and Systems*, vol. 63, no. 1, pp. 45-55, 1994.

[12] H. Tanaka, S. Uejima, and K. Asai, "Linear regression analysis with fuzzy model," *IEEE Transactions on System Man and Cybernetics*, vol. 12, no. 66, pp. 903-907, 1982.

[13] R. N. Tiwari, S. Dharmar, and J. R. Rao, "Priority structure in fuzzy goal programming," *Fuzzy Sets and Systems*, vol. 19, no. 33, pp. 251-259, 1986.

[14] R. C. Tsaur and H. F. Wang, "Outliers in fuzzy regression analysis," *International Journal of Fuzzy Systems*, vol. 1, no. 2, pp. 113-119, 1999.

[15] R. C. Tsaur and H. F. Wang, and J. C. O Yang, "Fuzzy regression for seasonal time series analysis," *International Journal of Information Technology and Decision Making*, vol. 1, no. 1, pp. 165-175, 2002.

[16] R. C. Tsaur, "Extrapolate internet users in Taiwan by risk assessment," *Computers and Mathematics with Applications*, vol. 46, no. 10-11, pp. 1725-1734, 2003.

[17] H. F. Wang and C. C. Fu, "A generalization of fuzzy goal programming with preemptive structure," *Computers and Operations Research*, vol. 24, no. 9, pp. 819-828, 1997.

[18] H. F. Wang and R. C. Tsaur, "Insight of a Fuzzy Regression Model," *Fuzzy Sets and Systems*, vol. 112, no. 3, pp. 355-369, 2000.

[19] H. F. Wang, R. C. Tsaur, "Bi-Criteria Variable Selection in Fuzzy Regression Equation," *Computers and Mathematics with Applications*, vol. 40, no. 6-7 , pp. 877-883, 2000.

[20] H. F. Wang and R. C. Tsaur, "Resolution of fuzzy regression model," *European Journal of Operation Research*, vol. 126, no. 3, pp. 637-650, 2000.

[21] H. C. Wu, "Fuzzy linear regression model based on fuzzy scalar product," *Soft Computing*, vol. 12, no. 5, pp. 469-477, 2008.

[22] J. R. Yu, G. H. Tzeng, and H. L. Li, "A general piecewise necessity regression analysis based on linear programming," *Fuzzy Sets and Systems*, vol. 105, no. 3, pp. 429-436, 1999.

[23] J. R. Yu, G. H. Tzeng, and H. L. Li, "General fuzzy piecewise regression analysis with automatic change-point detection," *Fuzzy Sets and Systems*, vol. 119, no. 2, pp. 247-257, 2001.

[24] H. J. Zimmermann, *Fuzzy Sets, Decision Making, and Expert Systems,* Kluwer, Boston, 1987.

Ruey-Chyn Tsaur received a B.B. degree in business mathematics from Soochow University, Taipei, Taiwan, R.O.C., in 1992, and the M.S. & Ph.D. degrees in Industrial Engineering and Engineering Management from National Tsing Hua University, in 1995 and 1999, respectively. He is currently a Professor with the Department of Management Sciences and Decision Making, Tamkang University, Taiwan. His research interests are on the Multicriteria Decision Making, Fuzzy Set Theory and Business Forecasting.

Hsiao-Fan Wang is the Distinguished Chair Professor of National Tsing Hua University, Taiwan, Republic of China. She has been teaching at the Department of Industrial Engineering and Engineering Management at the same university, NTHU after she graduated from Cambridge University, UK in 1981 and now is also the Vice Dean of the College of Engineering. She used to be the Head of the Department of IEEM, NTHU, President of Chinese Fuzzy Systems Association, Vice President of International Fuzzy Systems Association and Erskine Fellow of Canterbury University, NZ. Also, she has been awarded the Distinguished Research Award from National Science Council of Taiwan, ROC; Distinguished Contracted Research Fellow of NSC and Distinguished Teaching Award of Engineering College, NTHU. She used to be the editor-in-chief of the Journal of Chinese Industrial Engineering Association; also the Journal of Chinese Fuzzy Set and Theories and now is the area editor of several international journals. Her research interests are in Multicriteria Decision Making, Fuzzy Set Theory and Green Supply Chain Management.