Optimizing \( h \) value for fuzzy linear regression with asymmetric triangular fuzzy coefficients

Fangning Chen, Yizeng Chen, Jian Zhou *, Yuanyuan Liu

School of Management, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Available online 18 March 2015

Keywords:
Fuzzy linear regression
\( h \) value
System credibility
Asymmetric triangular fuzzy number
Crisp input–output

ABSTRACT

The parameter \( h \) in a fuzzy linear regression model is vital since it influences the degree of the fitting of the estimated fuzzy linear relationship to the given data directly. However, it is usually subjectively pre-selected by a decision-maker as an input to the model in practice. In Liu and Chen (2013), a new concept of system credibility was introduced by combining the system fuzziness with the system membership degree, and a systematic approach was proposed to optimize the \( h \) value for fuzzy linear regression analysis using the minimum fuzziness criterion with symmetric triangular fuzzy coefficients. As an extension, in this paper, their approach is extended to asymmetric cases, and the procedure to find the optimal \( h \) value to maximize the system credibility of the fuzzy linear regression model with asymmetric triangular fuzzy coefficients is described. Some illustrative examples are given to show the detailed procedure of this approach, and comparative studies are also conducted via the testing data sets.

1. Introduction

Fuzzy regression analysis was proposed by Tanaka et al. (1982) for the linear case using the fuzzy functions defined by Zadeh’s extension principle (Zadeh, 1975), in which the observed values can differ from the estimated values to a certain degree of belief. This method is recommended for practical situations where decisions often have to be made on the basis of imprecise and/or partially available data due to human estimations. Fuzzy regression analysis is a powerful tool in many decision domains in estimating relationships among variables with fuzzy, incomplete information (Chen et al., 2004; Höglund, 2013).

Generally, according to criterion functions, the existing fuzzy linear regression (FLR) methods can be roughly classified into two categories: FLR methods using the fuzzy least-squares criterion and FLR methods using the minimum fuzziness criterion. The first category of FLR methods aims to minimize the sum of squared errors in the estimated value based on the notion of distance between the predicted fuzzy outputs and the observed fuzzy outputs, which is indeed a fuzzy extension of the ordinary least squares. The second category of FLR methods, namely the minimum fuzziness approach, aims to build fuzzy linear models by minimizing the system fuzziness subject to including the data points of each sample within a specified feasible data interval. In this field, most of results focus on improving the linear regression method with the initial setting of symmetric triangular coefficients and the assumption of crisp input–output data in Tanaka et al. (1982). For example, many FLR models with some other types of fuzzy coefficients were proposed in the literature including asymmetric triangular fuzzy coefficients, symmetric fuzzy coefficients, LR-type coefficients, trapezoidal fuzzy coefficients, and exponential fuzzy coefficients (see, e.g., Ge and Wang, 2007; Kheirfam and Verdegay, 2013; Tanaka, 1997; Tanaka et al., 1995; Yen et al., 1999).

In the FLR methods using the minimum fuzziness criterion, there is an important input parameter \( h \) which refers to the degree of the fitting of the estimated fuzzy linear model to the given data, and is subjectively pre-selected by a decision-maker in real applications. The selection of a suitable \( h \) value for the FLR model is very vital since it determines the range of the possibility distributions of the obtained fuzzy coefficients directly. Actually, a higher \( h \) value will produce a large but unnecessary spread values of fuzzy coefficients, which has no operational definition or interpretation, while a lower \( h \) value may lead to a very narrow predictive interval so that the reliability of the FLR model is doubtful (Liu and Chen, 2013). In 1988, Tanaka and Watada (1988) first discussed the selection of a proper \( h \) value for the FLR model. They advised that a greater \( h \) value should be introduced if the observed data pairs are relative small. Besides, Bárdoossy (1990) stated that the selection of the \( h \) value could be generally based upon the decision-maker’s belief in the FLR model, and then recommended an \( h \) value between 0.5 and 0.7. Since these criteria are ad hoc, vague, and rather difficult to be justified or applied in

* Corresponding author. Tel.: +86 21 66134414 805.
E-mail address: zhou_jian@shu.edu.cn (J. Zhou).

http://dx.doi.org/10.1016/j.engappai.2015.02.011
0952-1976/© 2015 Elsevier Ltd. All rights reserved.
real situations, Moskowitz and Kim (1993) studied and examined the relationship among the \( h \) value, the membership function shape, and the spreads of fuzzy parameters in FLR with symmetric fuzzy numbers. Subsequently they developed a systematic approach to assessing a proper \( h \) value for the FLR model, which should satisfy the decision-maker's beliefs regarding the shape and range of the possibility distributions of fuzzy coefficients. Considering the situations that the observed data set contains a considerable level of noise or uncertainty, Ge and Wang (2007) suggested that the \( h \) value for the FLR model should be inversely proportional to the input noise. Besides, Shakouri and Nadimi (2009) introduced a measuring index of inequality of fuzzy numbers and proposed a programming with a new objective function as well as constraints to obtain an optimal \( h \) based on an idea of reducing the distance between the output of the possibilistic model and the measured output. Recently, Liu and Chen (2013) formulated a novel approach to optimizing the \( h \) value for FLR using the minimum fuzziness criterion based on a new notion of credibility, which may evaluate the reliability of FLR models. They focused on improving the classical and widely accepted FLR model proposed by Tanaka et al. (1982), and proposed a simple and easy-understood calculation process for the optimal \( h \) value by taking into account both the system fuzziness and the system membership degree. In their study, the given data are crisp input-output, and the coefficients are assumed to be symmetric triangular fuzzy numbers (TFNs). Considering that there are a great deal of data sets that generate scatter plots in which the data do not fall symmetrically on both sides of the regression line. Thus, the FLR with symmetric TFNs was extended asymmetrically by Yen et al. (1999) as follows.

If \( \hat{A}_j \) in (1) has an asymmetric triangular membership function, it can be uniquely defined by a triplet \( \hat{A}_j = (a^\ell_j, a^m_j, a^R_j) \), where \( a^m_j \) is the centre value, and \( a^\ell_j \) and \( a^R_j \) are the left and right spreads of \( \hat{A}_j \), respectively. The goal in FLR is to determine \( f(x_k) \) by minimizing the system fuzziness subject to the following inclusion conditions (Tanaka et al., 1982):

\[
y_i = [f(x_k)]^\alpha, \quad i = 1, 2, \ldots, m.
\]

Here \( h (0 \leq h < 1) \) is a parameter predetermined subjectively by the design team according to their engineering knowledge, and \( f(x_k)^\alpha \) is the \( h \)-level set of the predicted fuzzy output \( \tilde{y}_i = f(x_k) \) from the FLR model in (1) corresponding to the input vector \( x_k \), which is an interval. Since the fuzzy coefficients \( \hat{A}_j, j = 0, 1, \ldots, n, \) in (1) are all asymmetric TFNs, according to fuzzy arithmetic on TFNs, the predicted fuzzy output \( f(x_k) \) from the FLR model in (1) is also calculated as an asymmetric TFN. In order to further provide the expression of \( [f(x_k)]^\alpha \), we define

\[
x^+_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}
\]

and

\[
x^-_i = \begin{cases} 0 & \text{if } x_i \geq 0 \\ -x_i & \text{otherwise} \end{cases}
\]

for \( i = 1, 2, \ldots, m \) and \( j = 0, 1, \ldots, n \), respectively. It is obvious that \( x^+_i \) and \( x^-_i \) are both nonnegative and satisfy that \( x_i = x^+_i - x^-_i \) and \( |x_i| = x^+_i + x^-_i \).

Thus, if we denote the predicted fuzzy output \( f(x_k) \) via its left spread \( f^L(x_k) \), peak point \( f(x_k) \) and right spread \( f^R(x_k) \) as \( f(x_k) = (f^L(x_k), f(x_k), f^R(x_k)) \), then by the sum and product operations of TFNs, we can obtain \( f^L(x_k), f(x_k) \) and \( f^R(x_k) \), respectively, as

\[
f^L(x_k) = \sum_{j=0}^{n} a^\ell_j x^+_j + \sum_{j=0}^{n} a^m_j x^0_j,
\]

\[
f(x_k) = \sum_{j=0}^{n} a^m_j x^0_j,
\]

\[
f^R(x_k) = \sum_{j=0}^{n} a^R_j x^+_j + \sum_{j=0}^{n} a^m_j x^0_j.
\]

Besides, the \( h \)-level \((0 \leq h < 1)\) set of \( f(x_k) \) as \( (f^L(x_k), f(x_k), f^R(x_k)) \) is calculated as the following interval:

\[
[f(x_k)]^\alpha = [(1-h)f^L(x_k), f(x_k), (1-h)f^R(x_k)].
\]

As a result, the inclusion relation in (2) can be rewritten as

\[
f^L(x_k) - (1-h)f^R(x_k) \leq y_i \leq f^R(x_k) - (1-h)f^L(x_k), \quad i = 1, 2, \ldots, m.
\]

Furthermore, the system fuzziness, denoted by \( \Delta \), is defined by Tanaka et al. (1982) as the total covering area of predicted fuzzy outputs, i.e.

\[
\Delta = \sum_{i=1}^{m} \Delta Y_i = \sum_{i=1}^{m} \left[ \frac{1}{2} f^L(x_k) + f^R(x_k) \right] = \sum_{i=1}^{m} \left[ \frac{1}{2} (a^\ell_j + a^R_j)(x^+_j + x^-_j) \right]
\]

in which \( \Delta Y_i \) is the fuzziness with respect to \( \hat{y}_i \) in the asymmetric case, and can be given as

\[
\Delta Y_i = \frac{1}{2} \sum_{j=0}^{n} (a^\ell_j + a^R_j)(x^+_j + x^-_j).
\]
As depicted in Fig. 1, the covering area of the predicted fuzzy output \( \hat{y}_i = (y^L_i, y^C_i, y^R_i) \) is exactly the fuzziness \( \Delta y_i \), where \( y^L_i, y^C_i, \) and \( y^R_i \) represent the left spread, peak point, and right spread of \( \hat{y}_i \), respectively.

Henceforth, in terms of asymmetric TFNs, the fuzzy coefficients \( \hat{A}_j, j = 0, 1, \ldots, n \), can be obtained by solving the following linear programming model:

\[
\begin{align*}
\min \Delta &= \frac{1}{2} \sum_{i=0}^{n} \sum_{j=1}^{n} (a^L_{ij} + a^R_{ij})(x^C_j + x^C_i) \\
\text{subject to} & \\
(1-h) \left( \sum_{j=0}^{n} a^L_{ij} x^C_j + \sum_{j=0}^{n} a^R_{ij} x^C_j \right) & - n \sum_{j=0}^{n} a^L_{ij} x^C_j \geq -y_i, \quad i = 1, 2, \ldots, m \\
(1-h) \left( \sum_{j=0}^{n} a^L_{ij} x^C_j + \sum_{j=0}^{n} a^R_{ij} x^C_j \right) & + n \sum_{j=0}^{n} a^L_{ij} x^C_j \geq y_i, \quad i = 1, 2, \ldots, m \\
a^L_{ij}, a^R_{ij} & \geq 0, \quad j = 0, 1, \ldots, n.
\end{align*}
\]

in which the constraints (13b) and (13c) are corresponding to the inclusion condition in (2), the constraint (13d) guarantees that the left and right spreads of \( \hat{A}_j, j = 0, 1, \ldots, n \), are nonnegative.

In order to determine the asymmetric triangular fuzzy coefficients of the programming model in (13), a hybrid method using fuzzy regression combined with least squares regression has been designed in Ishibuchi and Nii (2001). First, it is required to determine the peak point vector \( \hat{a}^* \) by using least squares regression as follows:

\[
\hat{a}^* = (\hat{x^T} \hat{x})^{-1} \hat{x^T} \hat{y}.
\]

Subsequently, we utilize fuzzy regression to determine the parameters \( \hat{a}^L \) and \( \hat{a}^R \). That is, transform the above programming model in (13) into the following hybrid linear programming model

\[
\begin{align*}
\min \Delta &= \frac{1}{2} \sum_{i=0}^{n} \sum_{j=1}^{n} (a^L_{ij} + a^R_{ij})(x^C_j + x^C_i) \\
\text{subject to} & \\
(1-h) \left( \sum_{j=0}^{n} a^L_{ij} x^C_j + \sum_{j=0}^{n} a^R_{ij} x^C_j \right) & - n \sum_{j=0}^{n} a^L_{ij} x^C_j \geq -y_i, \quad i = 1, 2, \ldots, m \\
(1-h) \left( \sum_{j=0}^{n} a^L_{ij} x^C_j + \sum_{j=0}^{n} a^R_{ij} x^C_j \right) & + n \sum_{j=0}^{n} a^L_{ij} x^C_j \geq y_i, \quad i = 1, 2, \ldots, m \\
a^L_{ij}, a^R_{ij} & \geq 0, \quad j = 0, 1, \ldots, n.
\end{align*}
\]

3. Credibility for fuzzy linear regression

As we can see in Section 2, the value of parameter \( h \) determines the range of the possibility distributions of the fuzzy parameters, so it is important to select a suitable value for \( h \) in fuzzy regression analysis. To deal with the problem, a new concept of system credibility was suggested by Liu and Chen (2013), based on which FLR models can be achieved by taking into account both the system fuzziness and the system membership degree. In this section, the notion of system credibility is introduced, and then the system credibility of the FLR model in (13) is discussed.

3.1. System credibility

Assume that \( h_1 \) and \( h_2 \) are any two values for \( h \) with \( 0 \leq h_1 < 1 \) and \( 0 \leq h_2 < 1 \). Suppose that \( A_{1j} \) and \( A_{2j} \), \( j = 0, 1, \ldots, n \), are two sets of triangular fuzzy coefficients with respect to \( h_1 \) and \( h_2 \), respectively, and \( y^1_i \) and \( y^2_i \) are the corresponding estimations of \( y_i \) from (1), which are calculated as

\[
y^1_i = \sum_{j=0}^{n} A_{1j} x_j \quad \text{and} \quad y^2_i = \sum_{j=0}^{n} A_{2j} x_j.
\]

Now we are interested in which one, out of \( y^1_i \) and \( y^2_i \), is better as the estimation of \( y_i \). That is to say, when the sample data pairs are given, which one, out of \( h_1 \) and \( h_2 \), is better to be used to build an FLR model in (13). To address this problem, two critical factors, i.e., the estimated fuzziness \( \Delta y_i \) and the membership degree \( \mu_i(y_i) \), are naturally taken into consideration. In general, the smaller \( \Delta y_i \) and \( \mu_i(y_i) \) is, the better the performance of \( \hat{y}_i \) in representing \( y_i \) will be. Moreover, Liu and Chen (2013) initially presented some new concepts to tackle this problem as follows.

Definition 1. The credibility of \( \hat{y}_i \) in representing \( y_i \), denoted by \( z_i \), is defined as

\[
z_i = \frac{\mu_i(y_i)}{\Delta y_i}.
\]

The higher the \( z_i \) is, the better the performance of \( \hat{y}_i \) in representing \( y_i \) will be.

As we can see from (18), \( z_i \) can be improved by increasing \( \mu_i(y_i) \), reducing \( \Delta y_i \), or both simultaneously. Furthermore, based on (18), a conception of system credibility is proposed to measure the reliability of the FLR model in (1) as follows.

Definition 2. The system credibility of the FLR model in (1), denoted by \( z \), is defined as the total credibility of \( \hat{y}_i \) in representing \( y_i \) for \( i = 1, 2, \ldots, m \), which can be calculated as

\[
z = \sum_{i=1}^{m} z_i = \sum_{i=1}^{m} \frac{\mu_i(y_i)}{\Delta y_i}.
\]

The higher the \( z \) is, the better the performance of the FLR model will be.

3.2. Further discussion

Now let us consider the FLR model with asymmetric TFNs. Assume that a sample of \( m \) observations \( (y_{i1}, x_{1i}), (y_{i2}, x_{2i}), \ldots, (y_{im}, x_{mi}) \) are given. As introduced in Section 2, for any predetermined \( h \), the peak point vector \( \hat{a}^* = (a^L_{i0}, a^C_{i0}, \ldots, a^R_{im}) \) of the optimal fuzzy coefficients \( \hat{A}_{ij} \), \( j = 0, 1, \ldots, n \), can be determined via (14), and

\[
\begin{align*}
\text{Fig. 1. Covering area of the predicted fuzzy output } \hat{y}_i = (y^L_i, y^C_i, y^R_i).
\end{align*}
\]
then the optimal left and the right spreads \((a_j^L)_{h^*}\) and \((a_j^R)_{h^*}\), \(j = 0, 1, \ldots, n\), will be figured out by solving the linear programming model in (15) according to the minimum fuzziness criterion. Subsequently, we obtain the corresponding fuzzy estimation \(y_i^h\) of the observed crisp output \(y_i\) as

\[
y_i^h = \sum_{j=0}^n A_j^h x_j = \sum_{j=0}^n \left( (a_j^L)_{h^*}, \tilde{a}_j^C, (a_j^R)_{h^*} \right) x_j
\]

It is clear that for different values of parameter \(h\), the fuzzy estimation \(y_i^h\) is an asymmetric triangular fuzzy number with the same centre value \(\sum_{j=0}^n \tilde{a}_j^C x_j\). Fig. 2 shows the relationships between the fuzzy outputs \(y_i^{h_1}\) and \(y_i^{h_2}\) with respect to different levels \(h_1\) and \(h_2\).

Furthermore, the systematic credibility \(z^h\) of the FLR model in (1) with respect to \(h\) can be calculated as

\[
z^h = \sum_{i=1}^m z_i^h = \sum_{i=1}^m \frac{\mu_{A_i}}{\Delta y_i^h}, \quad (21)
\]

A specific \(h\) value will result in a unique \(z^h\), which implies that \(z^h\) is a function of the parameter \(h\).

In practice, it is possible that the given value of \(h\) predetermined by the design team or domain experts corresponds a very low system credibility \(z^h\), which implies a bad performance of the FLR model in (1). Therefore, in FLR, it is natural to determine an appropriate \(h\) value with a high system credibility \(z^h\). In our paper, we select the optimal \(h\) value with the maximum system credibility, which will be figured out in the following section.

### 4. Formulating an approach to optimizing the \(h\) value

In this section, based on the concept of system credibility introduced in Section 3, a systematic approach to selecting the optimal \(h\) value for a given set of sample data pairs is proposed by combining the criterion of minimizing the system fuzziness with that of maximizing the system credibility of an FLR model with asymmetric TFNs. The procedure to find the optimal \(h\) value is described in detail as follows.

Suppose that \(h_1\) and \(h_2\) are any two values in the interval \((0, 1)\). Let us denote the optimal solution, the optimal fuzziness, and the optimal objective function value of (15) with regard to \(h_1\) as \(\hat{A}_j^{h_1} = ((a_j^L)_{h_1}, \tilde{a}_j^C, (a_j^R)_{h_1})\), \(j = 0, 1, \ldots, n\), \(\Delta y_i^{h_1}\), \(i = 1, 2, \ldots, m\), and \(\Delta^h_1\), respectively. According to Moskowitz and Kim (1993), the corresponding optimal solution, the optimal fuzziness, and the optimal objective function value of the model in (15) with regard to \(h_2\) can be obtained as

\[
\begin{align*}
A_j^{h_2} &= \left( \frac{1-h_1}{1-h_2}, (a_j^L)_{h_2}, \tilde{a}_j^C, (a_j^R)_{h_2} \right), \quad j = 0, 1, \ldots, n \\
\Delta y_i^{h_2} &= \frac{1-h_1}{1-h_2} \Delta y_i^{h_1}, \quad i = 1, 2, \ldots, m \\
\Delta^h_2 &= \frac{1-h_1}{1-h_2} \Delta^h_1.
\end{align*}
\]

It is shown in (22) that when the FR model with asymmetric TFNs is used, the values of the left spread \((a_j^L)_{h^*}\) and the right spread \((a_j^R)_{h^*}\) of the optimal solution \(A_j^{h^*}\), \(j = 0, 1, \ldots, n\), the optimal fuzziness \(\Delta y_i^{h^*}\), \(i = 1, 2, \ldots, m\), and the optimal objective function value \(\Delta^h\) becomes \((1-h_1)/(1-h_2)\) times when the \(h\) value changes from \(h_1\) to \(h_2\). Based on (22), we may deduce the relationship between the corresponding system credibilities \(z^{h_1}\) and \(z^{h_2}\) with respect to \(h_1\) and \(h_2\) defined in (21) as follows.

Denote the optimal fuzzy output \(y_i^{h^*}\) with respect to \(h_1\) as \(y_i^{h_1} = (\tilde{y}_i^{h_1}, \tilde{y}_i^{C}, (\tilde{y}_i^{h_1})^R)\). Hence the membership degree \((\mu_{y_i^{h_1}})_{h^*}\) is calculated as

\[
(\mu_{y_i^{h_1}})_{h^*} = \begin{cases} 
1 - \tilde{y}_i^{C} - y_i \left( \frac{\tilde{y}_i^{h_1}}{\tilde{y}_i^{C}} \right) & \text{if } y_i \leq \tilde{y}_i^{C} \\
1 - \tilde{y}_i^{C} \left( \frac{\tilde{y}_i^{h_1}}{\tilde{y}_i^{C}} \right) & \text{if } y_i > \tilde{y}_i^{C},
\end{cases} \quad i = 1, 2, \ldots, m.
\]

If we define \(Q_i^{h_1}\) by

\[
Q_i^{h_1} = \begin{cases} 
\tilde{y}_i^{C} - y_i \left( \frac{\tilde{y}_i^{h_1}}{\tilde{y}_i^{C}} \right) & \text{if } y_i \leq \tilde{y}_i^{C} \\
\tilde{y}_i^{C} \left( \frac{\tilde{y}_i^{h_1}}{\tilde{y}_i^{C}} \right) & \text{if } y_i > \tilde{y}_i^{C},
\end{cases} \quad i = 1, 2, \ldots, m.
\]

then (23) can be rewritten as

\[
(\mu_{y_i^{h_1}})_{h^*} = 1 - Q_i^{h_1}, \quad i = 1, 2, \ldots, m.
\]

It follows from (12), (18), and (25) that the estimated credibility of the predicted fuzzy output \(y_i^{h^*}\) in representing \(y_i\) with regard to \(h_1\) can be expressed as

\[
z^{h_1} = \frac{(\mu_{y_i^{h_1}})_{h^*}}{\Delta y_i^{h_1}} = \frac{2(1-Q_i^{h_1})}{\sum_{j=0}^n (a_j^L)^{h_1} + (a_j^R)^{h_1}(x_{i,j}^L + x_{i,j}^R)}, \quad i = 1, 2, \ldots, m.
\]

Therefore, the system credibility of the FR model in (1) with respect to \(h_1\) can be obtained as

\[
z^{h_1} = \sum_{i=1}^m z_i^{h_1} = \sum_{i=1}^m \frac{1-Q_i^{h_1}}{\sum_{j=0}^n (a_j^L)^{h_1} + (a_j^R)^{h_1}(x_{i,j}^L + x_{i,j}^R)}
\]

Combining (22) and (24), with regard to \(h_2\), we have

\[
Q_i^{h_1} = \begin{cases} 
\tilde{y}_i^{C} - y_i \left( \frac{\tilde{y}_i^{h_1}}{\tilde{y}_i^{C}} \right) & \text{if } y_i \leq \tilde{y}_i^{C} \\
\tilde{y}_i^{C} \left( \frac{\tilde{y}_i^{h_1}}{\tilde{y}_i^{C}} \right) & \text{if } y_i > \tilde{y}_i^{C},
\end{cases} \quad i = 1, 2, \ldots, m.
\]

which follows that

\[
(\mu_{y_i^{h_1}})_{h^*} = 1 - Q_i^{h_1}, \quad i = 1, 2, \ldots, m.
\]

Therefore, according to (25) and (29), the membership degree
\((\mu_y(y_i))^{p_i^2}\) with regard to \(h_2\) can be calculated as
\[(\mu_y(y_i))^{p_i^2} = 1 - Q_i^{0^2} = 1 - \frac{1 - h_2 Q_i^h}{1 - Q_i^h}, \quad i = 1, 2, ..., m. \tag{30}\]

Henceforth, according to (18), (22), and (30), the estimated credibility of the predicted fuzzy output \(\hat{y}_i^{p_i^2}\) in representing \(y_i\) with regard to \(h_2\) can be expressed as
\[\hat{z}_i^{p_i^2} = \left(\frac{\mu_y(y_i))^{p_i^2}}{\Delta y_i^{p_i^2}}\right) = \frac{1 - h_2 Q_i^h + h_2 Q_i^h (1 - h_2) (1 - Q_i^h)}{1 - h_2 Q_i^h}, \quad i = 1, 2, ..., m. \tag{31}\]

in which
\[\rho_i^h = \frac{Q_i^h}{\Delta y_i^{p_i^2}}, \quad i = 1, 2, ..., m. \tag{32}\]

Consequently, the estimated system credibility with regard to \(h_2\) for the FLR model in (1) can be obtained as
\[z^{p_i^2} = \sum_{i=1}^{m} z_i^{p_i^2} = \sum_{i=1}^{m} \left(1 - h_2 Q_i^h + h_2 Q_i^h (1 - h_2) (1 - Q_i^h)\right), \tag{33}\]

in which
\[\rho_i^h = \sum_{i=1}^{m} \rho_i^h. \tag{34}\]

Moreover, if we set \(h_1 = 0\) and \(h_2 = h\), then (22), (28), (30), (31), and (33) can be rewritten as
\[
\begin{align*}
\hat{y}_j^{n^2} &= \left(1 - h \right) \left( (\tilde{a}_j^{p_0})^{\sigma'} \tilde{a}_j^{p_0} \right), \quad j = 0, 1, ..., n \\
\Delta \hat{y}_i^{p_i^2} &= \frac{1 - h}{1 - h} \Delta \hat{y}_i^{p_i^2}, \quad i = 1, 2, ..., m \\
\Delta \hat{y}_i^{p_i^2} &= \frac{1 - h}{1 - h} \Delta \hat{y}_i^{p_i^2}.
\end{align*} \tag{35}\]

and
\[Q_i^h = (1 - h) Q_i^h, \quad i = 1, 2, ..., m, \tag{36}\]

\[\mu_y(y_i))^{p_i^2} = \mu_y(y_i))^{p_i^2} + h Q_i^h, \quad i = 1, 2, ..., m. \tag{37}\]

\[z_i^{p_i^2} = -P_i^{h^2} + (P_i^{h^0} - z_i^{p_i^2}) h + z_i^{p_i^2}, \quad i = 1, 2, ..., m. \tag{38}\]

\[z_i^{p_i^2} = -P_i^{h^2} + (P_i^{h^0} - z_i^{p_i^2}) h + z_i^{p_i^2}. \tag{39}\]

In order to derive a formula for the optimal \(h\) value based on minimizing the estimated system credibility of the FLR model in (1), the derivative of (39) with respect to \(h\) needs to be derived and set to zero, which can be given as follows:
\[
\frac{\partial z_i^{p_i^2}}{\partial h} = -2 P_i^{h^2} + P_i^{h^0} - z_i^{p_i^2} = 0. \tag{40}\]

By solving (40), the optimal value for \(h\) can be calculated as
\[h^* = \frac{1}{2} \left(1 - \frac{z_i^{p_i^2}}{P_i^{h^0}} \right) \tag{41}\]

providing \(0 \leq h^* < 1\). Since \(z_i^{p_i^2} \geq 0\) and \(P_i^{h^0} > 0\), we have
\[\frac{z_i^{p_i^2}}{P_i^{h^0}} \geq 0. \tag{42}\]

Therefore, we can conclude that
\[h^* \leq 0.5. \tag{43}\]

However, in general, (41) cannot guarantee that \(h^* \geq 0\), and sometimes the value for \(h^*\) may be negative, i.e., \(z_i^{p_i^2} > P_i^{h^0}\). In such a case, we should obtain the optimal value for \(h\) by solving the following programming model:
\[
\begin{align*}
\text{max } & z_i^{p_i^2} = -P_i^{h^2} h^2 + (P_i^{h^0} - z_i^{p_i^2}) h + z_i^{p_i^2} \\
\text{subject to } & 0 \leq h < 1.
\end{align*} \tag{44}\]

It follows from \(z_i^{p_i^2} \geq 0, P_i^{h^0} \geq 0\), and \(z_i^{p_i^2} > P_i^{h^0}\) that the optimal solution of (44) is
\[h^* = 0. \tag{45}\]

Comparing the proposed approach for the FLR model with asymmetric TNFS with that in Liu and Chen (2013) for the symmetric case, it can be seen that the same form for the system credibility of the models with symmetric and asymmetric TNFSs is obtained as described in (39), and the only difference is the calculation of parameters \(P_i^{h^0}\) and \(z_i^{p_i^2}\). For the asymmetric case, \(P_i^{h^0}\) and \(z_i^{p_i^2}\) are calculated by (32), (34), and (27). In addition, it should be noted that in both cases, the optimal value for \(h\) would not exceed 0.5 based on the maximum system credibility criterion.

5. Numerical simulations

In this section, some numerical simulations are used to illustrate the performances of the systematic approach proposed above, and some comparative studies with the approach in Liu and Chen (2013) for the FLR with symmetric TNFSs are also made.

5.1. Example 1

First, the numerical example in Chang and Ayyub (2001) is used to show how the optimal \(h\) value can be obtained using the proposed approach. The eight testing data pairs are listed in Table 1.

When \(h\) is set to 0, the fuzzy coefficients in terms of asymmetric triangular fuzzy numbers can be obtained by calculating the peak point vector \(\tilde{a}^{p_0}\) via (14) and then solving the programming model given in (15) as \(A_0^{p_0} = (1.9286, 12.9286, 0.0714)\) and \(A_0^{p_0} = (0.0357, 0.3537, 0.2143)\). Hence the fuzzy regression model with respect to \(h=0\) is given as
\[\tilde{y}_i^{p_0} = (1.9286, 12.9286, 0.0714) + (0.0357, 0.3537, 0.2143)x. \tag{46}\]

Accordingly, the fuzziness \(\Delta \tilde{y}_i^{p_0}\), parameters \(Q_i^{h_0}\) and \(P_i^{h_0}\), membership degrees \(\mu_y(y_i))^{p_i^2} \), and credibility \(\rho_i^h\), \(i = 1, 2, ..., 8\), are calculated, respectively, as shown in Table 2.

| Table 1|
| Testing data set I. |
|---|---|
| **y** | **x** |
| 14 | 2 |
| 16 | 4 |
| 14 | 6 |
| 18 | 8 |
| 18 | 10 |
| 22 | 12 |
| 18 | 14 |
| 22 | 16 |
From Table 2, \( z^0 \) is calculated as 1.7546, and \( P^0 \) as 2.3249. According to (39), the system credibility of the fuzzy regression model with respect to \( h \) can be represented as follows:

\[
z^0 = -2.3249h^2 + 0.5703h + 1.7546. \tag{47}
\]

Since \( z^0 < P^0 \), according to (41), the optimal value \( h^* \) is calculated as 0.1227, and the optimal coefficients in the form of asymmetric case are obtained as \( \tilde{A}_0 = (2.1983, 12.9286, 0.0814) \) and \( \tilde{A}_1 = (0.0407, 0.5357, 0.2443) \). Therefore, the optimal fuzzy linear relationship in the asymmetric case is given as

\[
y^* = (2.1983, 12.9286, 0.0814) + (0.0407, 0.5357, 0.2443)x. \tag{48}
\]

The system credibility \( z^* \) of the model in (48) is calculated as 1.7896, which is higher than that of the model in (46), i.e., 1.7546.

---

Table 2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta \psi_i )</td>
<td>1.2500</td>
<td>1.5000</td>
<td>1.7500</td>
<td>2.0000</td>
<td>2.2500</td>
<td>2.5000</td>
<td>2.7500</td>
<td>3.0000</td>
</tr>
<tr>
<td>( Q_i^0 )</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.4400</td>
<td>0.1250</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.1429</td>
</tr>
<tr>
<td>( P_i^0 )</td>
<td>0.0000</td>
<td>0.6667</td>
<td>0.5714</td>
<td>0.2200</td>
<td>0.0556</td>
<td>0.4000</td>
<td>0.3636</td>
<td>0.0476</td>
</tr>
<tr>
<td>( \mu_{\psi_i}(y_i)^2 )</td>
<td>1.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.5600</td>
<td>0.8750</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.8571</td>
</tr>
<tr>
<td>( z_i^0 )</td>
<td>0.8000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2800</td>
<td>0.3889</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2857</td>
</tr>
</tbody>
</table>

---

Fig. 3. The typical change processes with respect to \( h \). (a) Fuzziness of the eight sample data. (b) Membership degrees of the eight sample data. (c) Credibility of the eight sample data. (d) System credibility of the FLR model.
Besides, comparing with the symmetric case in Liu and Chen (2013) with the same data set \( (z^* = 1.4960 \) in this case), the system credibility of the FLR model with asymmetric TFNs will improve 19.63%.

According to (35), (37)–(39), the changes of the fuzziness \( \Delta y_i^h \), the membership degrees \( \mu_i(y_i) \), and the system credibility \( z^h \) with different \( h \) values are depicted in Fig. 3(a), (b), (c) and (d), respectively. Notably, the vertical axis is logarithmic scale in Fig. 3(a) to enable us to see small changes in fuzziness among eight sample data, and the membership degrees for sample data 2, 3, 6, and 7 are overlapped in Fig. 3(b).

From Fig. 3(a) and (b), we can see that the fuzziness of all sample data will increase, and the membership degrees of all sample data will increase or keep constant with the augment of the \( h \) value. When the \( h \) value is close to 1, the fuzziness of all sample data will be extremely large, and the membership degrees of all sample data will converge to 1. It is clearly shown in Fig. 3(c) that when the \( h \) value is close to 1, the credibility for all sample data will be close to zero due to the extremely large fuzziness. Fig. 3(d) shows that when the \( h \) value is 0.1227, the system credibility of the FLR model will achieve the maximum, based on which the best FLR model with asymmetric TFNs can be obtained.

5.2. Example 2

To further demonstrate the effectiveness of the approach proposed in this paper, another numerical example with two independent variables used in Liu and Chen (2013) is given as follows. The 20 observations in the testing data set II are listed in Table 3.

Similarly, the corresponding results with the setting of \( h = 0 \) for the FLR model can be obtained, and the values of \( P_0 \) and \( z_0 \) are calculated as 0.1458 and 0.2042. Since \( z_0 > P_0 \), it follows from (44) and (45) that the optimal value \( h^* \) is zero. When \( h \) is specified as 0, 0.1, 0.2, 0.3, 0.4, and 0.5, the fuzzy coefficients in terms of asymmetric TFNs and the corresponding system credibility can be obtained by solving the programming model in (15). The corresponding results involving the best fuzzy linear model and the system credibility \( z^h \) are summarized in Table 4.

From Table 4, we can see that when the \( h \) value is set to 0, the system credibility of the FLR model with asymmetric TFNs will achieve the maximum, that is, 0.2042, which indicates that the approach proposed in this paper is effective and reasonable. Similarly, comparing with the symmetric case in Liu and Chen (2013) with the same data set \( (z^* = 0.1371 \) with regard to the optimal \( h \) value, \( h^* = 0.3184) \), it is not surprise that the system credibility of the FLR model with asymmetric TFNs will improve 48.94%. We also depict the changes of the optimal system credibility \( z^h \) with respect to different \( h \) values in the interval \( [0, 1] \) as shown in Fig. 4. From that, we can see when the \( h \) value is 0, the system credibility of the FLR model will achieve the maximum, based on which the best FLR model with asymmetric TFNs can be obtained.

5.3. Comparative studies

In order to examine the performance of the FLR with symmetric and asymmetric TFNs, some comparative studies are made by using the first five, the first six, the first seven, and the first eight sample data in Table 1. The changes of the optimal \( h \) value, the system fuzziness, the system membership degree, and the system credibility in both symmetric and asymmetric cases with the augment of sample data size are given in Fig. 5(a), (b), (b) and (d), respectively.

From Fig. 5(a), it can be seen that the optimal \( h \) values in both symmetric and asymmetric cases will decrease with the augment of sample data size, and the optimal \( h \) value in the asymmetric case is lower than that in the symmetric case when the volume of the sample data pairs is given. As shown in Fig. 5(b), when the volume of sample data pairs is given, the system fuzziness with respect to the optimal \( h \) value in the asymmetric case, which is calculated as the sum of individual fuzziness of all sample data, is lower than that in the symmetric case. As demonstrated in Fig. 5(c), the system membership degree with respect to the optimal \( h \) value in the asymmetric case, which is calculated as the sum of individual membership degrees of all sample data, is not more than that in the symmetric case. It is clearly shown in Fig. 5(d) that the system credibility of the FLR model in both symmetric and asymmetric cases will increase with the augment of sample data size, and the system credibility with the optimal \( h \) value in the asymmetric case is higher than that in the symmetric case when the volume of sample data pairs is given.

In summary, the optimal \( h \) values in both symmetric and asymmetric cases will decrease with the augment of sample data size, while the system credibility of the FLR model will increase with the augment of the sample data size. When the volume of sample data pairs is determined, the credibility of the FLR model with respect to the optimal \( h \) value in the asymmetric
case is better than that in the symmetric case. Furthermore, based on the above numerical simulations, we can also see that the larger the volume of sample data pairs is, the better the performances of FLR models will be in both symmetric and asymmetric cases, and it is more appealing using the FLR model with asymmetric fuzzy coefficients than that using the symmetric fuzzy coefficients when the volume of sample data pairs is determined.

6. Conclusions

In this paper, a systematic approach was proposed to select the optimal $h$ value for FLR analysis with asymmetric TFNs. Firstly, a new concept of credibility defined in Liu and Chen (2013) was introduced by taking into consideration both the system fuzziness and the system membership degree, which can be used to assess the performance of FLR analysis with different $h$ values using a given set of sample data pairs. Secondly, a procedure to obtain the optimal $h$ value was formulated for the FLR model with asymmetric TFNs by maximizing the system credibility of FLR models. The approach proposed in this paper suggests that the $h$ value should not be more than 0.5 when FLR analysis is used with symmetric TFNs or asymmetric TFNs in order to obtain more reliable FLR models. Besides, the optimal $h$ value will decrease with the augment of the sample data size. Comparative studies show that the result of the FLR model with asymmetric TFNs is more reliable than that of the FLR model with symmetric TFNs when the same $h$ value is employed.

For the further work, firstly, the theoretical study in this direction involves extending the approach proposed in this paper for FLR analysis with other forms of fuzzy sets, e.g., symmetric or asymmetric trapezoidal fuzzy numbers, LR fuzzy numbers, etc. Besides, the crisp outputs of the fuzzy regression models in this paper could also be expanded into fuzzy outputs, i.e., developing a procedure to select the optimal $h$ value for the FLR with fuzzy outputs using symmetric TFNs or asymmetric TFNs. Secondly, the proposed fuzzy linear regression approach with an optimized $h$ value could be well utilized in many practical applications to settle the involved uncertain variables. For instance, for the product design, the relations between the customer satisfaction and the product characteristics could be expressed with fuzzy numbers and identified through the FLR models with optimized $h$ values. Besides, in the fields of customer identification, market segmentation, and risk control, etc., the approach proposed in this paper could also provide good solutions for the uncertain relations in the practical problems.

**Fig. 5.** Changes with the augment of sample data size. (a) Changes of the optimal $h$ value. (b) Changes of system fuzziness. (c) Changes of system membership degree. (d) Changes of system credibility.
Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China Grant (No. 71272177/G020902), the National Social Science Foundation of China (No. 13CGL057), and Innovation Program of Shanghai Municipal Education Commission (No. 12ZS101).

References