Possibilistic Type-2 Fuzzy Regression with Application to TAIEX Forecasting

Narges Shafaei Bajestani¹, Ali Vahidian Kamya², Assef Zare³
¹-Department of Electrical Engineering, Tehran Science and Research Branch, Islamic Azad University, Tehran, Iran
Email: Narges.shafaei@gmail.com
²-Department of Electrical Engineering, Tehran Science and Research Branch, Islamic Azad University, Tehran, Iran
Email: Avkamyad@yahoo.com
³-Department of Electrical Engineering, Gonabad Branch, Islamic Azad University, Tehran, Iran
Email: Assef.zare@gmail.com

Received: September 2015 Revised: November 2015 Accepted: January 2016

ABSTRACT:
Interval type-2 fuzzy logic system permits us to model rule uncertainties and every membership value of an element is interval itself. This study aims to propose a framework for an interval type-2 possibilistic fuzzy regression model. In this model, vagueness is minimized, under the circumstances where the h-cut of observed value is included in predicted value. In this model both primary and secondary membership function of predicted value fit the observed value.

KEYWORDS: type-2 fuzzy regression, Possibilistic, Forecasting

1. INTRODUCTION
Statistical linear regression has long been used in almost every field of science. Developed by Tanaka et al. [1-3], the fuzzy regression model has been applied to modeling systems involving vague or imprecise phenomena. Fuzzy regression techniques can generally be classified into two distinct areas. The first area, proposed by Diamond [4], is an adaptation of the classical least squares method. The second area, called possibilistic regression [1-3], reduces the problem of finding fuzzy coefficients of a regression model to one of mathematical programming. The fuzzy regression analysis can be applied to many real-life problems [5-8] in which the strict assumptions of classical regression analysis cannot be satisfied. For example, people rarely use numerical expressions to judge things in daily life. The recognition, the judgment and evaluation that humans actually perform are commonly expressed in qualitative linguistic terms that can also be modeled with fuzzy sets. Many researchers have worked on the field of fuzzy linear regression (FLR) [9-12]. However, to this date, most models of fuzzy regression analysis have focused only on type1 fuzzy sets, while the type-2 fuzzy logic systems (FLS) can help deal with even higher levels of uncertainty in real world applications. Type-2 fuzzy rules are more complex than type-1 fuzzy rules, because they use of type-2 fuzzy sets in antecedent or consequent parts. Therefore, most of the type-2 FLS researches only concern with interval type-2 FLSs (IT2FLS).

Some researchers have used type-2 fuzzy regressions as the modeling structure. YichengWei and JunzoWatada have been build a Type-2 Fuzzy Qualitative Regression Model [13], which substantially is similar to our model when H=0. in their article, YichengWei and JunzoWatada imply that they have used a general type-2 fuzzy number while it seems that they have used interval type-2. O. Poleshchuk and E. Komarov present a regression model for interval type-2 fuzzy sets based on the least squares estimation technique [14]. Hosseinzadeh et al. present a weighted goal programming approach to fuzzy linear regression with crisp inputs and type-2 fuzzy outputs [15]. But this model only has tried to close the membership functions of observed and estimated responses by closing some of their parameters. It seems that none of the studies mentioned above could well model the Type2 fuzzy regression; instead they have reduced their model to only some points of type-2 fuzzy numbers.

This study aims to propose a framework for an interval Type 2 possibilistic fuzzy regression model. Interval type-2 fuzzy logic system permits us to model rule uncertainties and every membership value of an element is interval itself. The Taiwan stock index (TAIEX) is used to forecast with these model.

This paper is organized as follows. Section 2 reviews the relevant studies covered in this study, including type-1fuzzy regression models, type-2 fuzzy sets.
Section 3 describes proposed Type-2 fuzzy regression model. Section 4 applies the model to forecast the stock index. Section 5 is conclusion.

2. REWIVE

First of all, we assume that a fuzzy phenomenon can be presented as a fuzzy system of equations which, in turn, can be described by the fuzzy function

\[ \tilde{Y} = A_1X_1 + A_2X_2 + \ldots + A_nX_n \]

in which \( A_j \) is the fuzzy parameters with triangular membership functions \((a_j, b_j, c_j)\).

Given fuzzy parameters \( A_j \), the membership function of \( \tilde{Y} \) by the fuzzy linear function is obtained as the triangular membership function

\[ \left( \frac{1}{2} \right) \]

in which:

\[ C_1 = \sum_{i=1}^{n} a_i x_i, \quad C_2 = \sum_{i=1}^{n} b_i x_i, \quad C_3 = \sum_{i=1}^{n} c_i x_i \]  

(1)

The membership function of the output data \( Y \) also should have the same special form \((p_i, q_i, r_i)\).

If \( \tilde{A} \) and \( Y \) assumed to be symmetric triangular fuzzy number in which \( q \) and \( w \) represent the center value and spread width of \( Y \), respectively. The degree of the fitting of the estimated fuzzy linear regression model \( \tilde{Y} = \tilde{A}_1X_1 + \tilde{A}_2X_2 + \ldots + \tilde{A}_nX_n \) to the given data set \( Y = (q,w) \) is measured by the following index \( h \) which maximizes \( h \) subject to \( \tilde{h} \) for each element of this data in set \( Y = (q,w) \). The vagueness of the fuzzy regression model is defined by minimizing \( \tilde{V} \) for all data in \( Y = (q,w) \). The following linear programming problem expresses the above situation [2,9]:

\[ \min_{\tilde{V}} \sum_{x_j} \sum_{l_j} \sum_{s_j} \sum_{t_j} \]

\[ \text{subject to} \]

\[ \sum_{j=1}^{n} x_j (1-H) \sum_{j=1}^{n} l_j x_j \geq q + (1-H) w, \quad q \leq C_3 \]

\[ -\sum_{j=1}^{n} x_j (1-H) \sum_{j=1}^{n} l_j x_j \geq -q + (1-H) w, \quad q \geq C_3 \]

\[ l_j \geq 0, \quad w \geq 0 \]

If the input data \( x_j \) is a fuzzy number, \( x_j = (d_j, e_j, f_j) \). The membership function for the generalized fuzzy linear function \( \tilde{Y} = \tilde{A}_1X_1 + \tilde{A}_2X_2 + \ldots + \tilde{A}_nX_n \) is obtained from the following [9] (see fig.1):

\[ U_{Y_j}(y) = \begin{cases} \frac{-B_1}{2A_1} + \left[ \frac{(B_1^2/2A_1)^2 - (C_1 - y)^2}{A_1} \right]^{1/2}, & C_1 \leq y \leq C_3 \\ \frac{B_2}{2A_2} - \left[ \frac{(B_2^2/2A_2)^2 - (C_2 - y)^2}{A_2} \right]^{1/2}, & C_2 \leq y \leq C_3 \\ 0, & \text{o.w.} \end{cases} \]

(3)

In which

\[ A_1 = \sum_{j=1}^{n} (b_j - a_j)(e_j - d_j), \quad A_2 = \sum_{j=1}^{n} (c_j - b_j)(f_j - e_j) \]

\[ B_1 = \sum_{j=1}^{n} (a_j(e_j - d_j) + d_j(b_j - a_j)) \]

\[ B_2 = \sum_{j=1}^{n} (c_j(f_j - e_j) + e_j(c_j - b_j)) \]

\[ C_1 = \sum_{j=1}^{n} a_j d_j, \quad C_2 = \sum_{j=1}^{n} c_j f_j, \quad C_3 = \sum_{j=1}^{n} b_j e_j \]

For simplicity, in practice the approximation formula \( \tilde{Y} = (C_1, C_3, C_2) \) that represents triangular fuzzy number can be used. Regression model can be simplified as follows (see Fig. 2):

\[ \min_{\tilde{V}} \sum_{x_j} \sum_{l_j} \sum_{s_j} \sum_{t_j} \]

\[ \text{subject to} \]

\[ h = \begin{cases} \frac{1}{C_3 - C_1} \left( \frac{C_1(q - p) - p(C_3 - C_1)}{q - p} + (C_3 - C_1) - C_1 \right) \geq H & \text{if } q < C_3 \\ \frac{1}{C_3 - C_2} \left( \frac{C_2(q - r) - r(C_3 - C_2)}{q - r} - (C_3 - C_2) \right) \geq H & \text{if } q > C_3 \end{cases} \]

(4)

3. TYPE FUZZY REGRESSION

Type 1 FLSs cannot fully handle the high levels of
uncertainties available in the vast majority of real world applications. This is because type 1 FLSs employ crisp and precise type-1 fuzzy sets. A type-2 FLS can handle higher uncertainty levels to produce improved performance. Type-2 FLSs have a variety of real world applications, in the Business and Finance Domains, in the Electrical Energy Domain, Real-World Automatic Control, Healthcare and Medical Domains [10,11].

The membership functions of type-2 fuzzy sets are three dimensional and include a footprint of uncertainty, it is the third dimension of type-2 fuzzy sets and the footprint of uncertainty that provide additional degrees of freedom that make it possible to model and handle uncertainties [12].

Interval Type-2 fuzzy sets, has been widely investigated and applied in various contexts such as decision making, time series forecasting, control of mobile robots [13], etc. Interval Type-2 fuzzy sets are the most widely used Type-2 fuzzy sets because they are simple to use and because it is very difficult to justify the use of any kind of Type-2 fuzzy sets. In this case, the MF $\mu A(a)$ is an Interval Type-2 fuzzy set so that it can be represented only by its lower and upper bounds (i.e.by two Type-1 MFs).

Concerning Arianna Mencattini et al.[16], We will denote this kind of Type-2 interval As $[\left[ a_1, a_2 \right], \left[ b, a_4 \right]]$  Such that $a_1 \leq a_2 \leq b \leq a_3 \leq a_4$ , (see fig. 3): In this model, the input data $x_j$ is a scalar, $\tilde{A}_j$ is the interval type-2 fuzzy parameters with FOU triangular membership functions $\tilde{A}_j = \left[ \left[ a_{uj}, a_{lj} \right], b_j, \left[ c_{uj}, c_{lj} \right] \right]$. The membership function of the fuzzy linear $\tilde{Y} = \tilde{A}_1 x_1 + \tilde{A}_2 x_2 + \ldots + \tilde{A}_n x_n$ is obtained as the following:

$$C_{1LU} = \sum_{j=1}^{a} a_{uj} x_j \quad C_{1LL} = \sum_{j=1}^{a} a_{lj} x_j \quad C_{1U} = \sum_{j=1}^{a} b_j x_j$$

$$C_{1LU} = \sum_{j=1}^{a} c_{uj} x_j \quad C_{1LL} = \sum_{j=1}^{a} c_{lj} x_j$$

(5)

The output data $Y_i$ is the interval type-2 fuzzy parameters with FOU triangular membership functions $\tilde{Y}_i = \left( \left[ p_{li}, p_{ui} \right], q_i, \left[ r_{li}, r_{ui} \right] \right)$.

Solution of this problem is similar to type 1 fuzzy regression; there is some difference that should be noted.

i. Here, there are two type 1 fuzzy number( LM, UM), so we have two h (h1 , h2).

ii. There are additional constraints that should be noted

In all secondary membership function of x, The given output should be included in the estimated output , thus , interval $(y_1, y_2)$ should be included in interval $(y_1', y_2')$ [14]. $y_2'-y_1'$ should be minimized to minimize secondary uncertainty, too. To satisfy these assumption, UM of given output should be included UM of estimated output. Thus , $p_{mi}, q_{mi}, r_{mi}$ should approach to $C_{1LU}, C_{1LL}, C_{1U}, C_{1LU}$, in addition, here LMF and UMF are two type 1 fuzzy number that we use type 1 fuzzy regression for both of them (see fig.4 , fig 5).

If $A_j, Y_j$ be assumed Symmetric,

$$A_j = \left[ \left[ a_j - c_{mu}, a_j - c_{ml} \right], \left[ a_j - c_{mu}, a_j - c_{ml} \right] \right]$$

$$Y_j = \left[ \left[ y_j - e_{mu}, y_j - e_{ml} \right], \left[ y_j - e_{mu}, y_j - e_{ml} \right] \right]$$

Fig 4. Interval type 2 fuzzy regression

Fig 5. Second membership functions
And Formulation of interval type 2 fuzzy regressions can be expressed as the following QP problem (fig.6):

\[
\begin{align*}
\min & \sum c_i + \sum (x_i^T a - y_i)^2 + \\
& \sum (y_i - e_u) - (x_i^T a - \sum c_i |x_i|)^2 + \\
& \sum (y_i + e_u) - (x_i^T a - \sum c_i |x_i|)^2 + \\
& \sum (y_i - e_d) - (x_i^T a - \sum c_i |x_i|)^2 + \\
& + \sum (y_i + e_d) - (x_i^T a - \sum c_i |x_i|)^2
\end{align*}
\]

\[st \quad (x_i^T a - \sum c_i |x_i|) < y_i - e_u < (x_i^T a - \sum c_i |x_i|) \]
\[ (x_i^T a - \sum c_i |x_i|) < y_i - e_d < (x_i^T a - \sum c_i |x_i|) \]
\[ (x_i^T a + \sum c_i |x_i|) < y_i + e_d < (x_i^T a + \sum c_i |x_i|) \]
\[ (x_i^T a + \sum c_i |x_i|) < y_i + e_u < (x_i^T a + \sum c_i |x_i|) \]
\[ x_i^T a + (1 - H_1) \sum c_i |x_i| \geq y_i + (1 - H_1)e_d \]
\[ -x_i^T a + (1 - H_1) \sum c_i |x_i| \geq -y_i + (1 - H_1)e_d \]
\[ x_i^T a + (1 - H_2) \sum c_i |x_i| \geq y_i + (1 - H_2)e_u \]
\[ -x_i^T a + (1 - H_2) \sum c_i |x_i| \geq -y_i + (1 - H_2)e_u \]

4. TAIEX FORECASTING USING TYPE-2 FUZZY REGRESSION

If in this research, the TAIEX data in the year 2000 have been used to show the ability of these two models \[11, 16\].

Data from 11/2 to 12/15 are training data, 12/16 to 12/29 are testing data. The model is \( \hat{P} = \hat{A}_0 + \hat{A}_1 X_1 \). H1 and H2 are adopted as value 0.6, .5 respectively. The results are shown on fig. 9 and table 1. And fig 9 has shown forecasted time series after defuzzifying.

<table>
<thead>
<tr>
<th>Table 1. Model parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuzzy parameters</td>
</tr>
<tr>
<td>( a )</td>
</tr>
<tr>
<td>( c_l )</td>
</tr>
<tr>
<td>( c_u )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. RMSE comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 2 fuzzy model</td>
</tr>
<tr>
<td>RMSE on testing data</td>
</tr>
</tbody>
</table>

Fig. 6. Interval Type 2 fuzzy regression problem
5. CONCLUSION
This study presents a type-2 fuzzy regression model according to the possibilistic model. In this model, vagueness is minimized, under the circumstances where the h-cut of observed value is included in predicted value. In this case, both observed values and predicted values are interval type 2 fuzzy numbers. In this model both primary and secondary membership function of predicted value fit the observed value. Developing model to piecewise model makes it helpful in dealing with the fluctuating data. The numerical examples in this study are accurate and acceptable. Only one type of type-2 fuzzy sets is discussed here, but many types of type-2 fuzzy sets can be treated through developing same concept as in this paper. Yet, most of high vague phenomenon might be well identified by this model.

REFERENCES