Simulation of Continuous Qualitative Variables in Econometric Models Using Fuzzy Functions and Numbers

1Ahmad Jafari samimi, 2Bijan Bidabad, 3Roohollah Mohammadi

1Professor of Economics, University of Mazandaran, Babolsar, Iran.  
2PhD in Economics – Researcher  
3MSc Econom & Social Systems Engineering

Abstract: The present paper deals with explanatory dummy variables in econometrics models in two different forms namely, discrete and continuous. In other words we will concentrate not only on a usual binary zero and one dummy variables but also using a fuzzy system to include the case in which dummy variable may appear in a non-discrete form. After introducing a fuzzy system, the application of this system in continuous dummy variables in regression analysis has been investigated. The results indicate that using continuous dummy variables in a fuzzy system version that comparing to the ordinary classic usage of a binary dummy variable we will have a better performance on F and t statistics as well as D-W and R-squares. In other words, using continuous dummy variables in a fuzzy system version will increase the level of significance as well as the explanatory power of the model.

Key words:

INTRODUCTION

In general, explanatory variables used in econometric models are of two types: quantitative and qualitative. In the classical practice, explanatory variables (such as wartime, revolution, earthquake, severe economic momentum, etc.), regardless of being discrete or continuous, are used in regression analysis with values of either zero or one. This classification leads to sudden changes in models and if there are intermediate levels and stages between zero and one, these models cannot always include desirable results. Only linguistic variables are able to define intermediate levels and to give numerical identity to the respective variable using Fuzzy sets. Zero and one variables are called virtual variables and are dealt with like other explanatory variables. Virtual variables can be used in terms of y-intercept, gradient or both. According to the classical principles, everything has a fixed rule according to which the trueness or faultiness of problem is determined. Using these presuppositions one cannot turn qualitative variables to quantitative elements. In fact, phenomena cannot always be divided into correct and incorrect or one and zero, but something between zero and one. In other words, real world expressions are between completely true or completely false; between zero and one; not of two values, but of multiple or fuzzy values.

Regarding the changeable and ambiguous nature of qualitative variables, sometimes only man can determine the stature of variables and decides on continuity or discreteness of them with recourse to his overall perception of signs. Because what is compatible with logic doesn’t mandatory happen in reality, sometimes the use of explicit two-value quantities to recognize patterns, including continuous qualitative variables, won’t yield desirable results. Therefore, in order to include qualitative variables, one should use a method which enjoys a high potential to mix qualitative variables. It seems that fuzzy systems, given their special characteristics, can explain the role of continuous qualitative variables in econometric models.

Research Methodology:

Literature on Fuzzy Method:

Till early 1960s, classical theory was the most dominant theory in most scientific calculations and planning. At the beginning of the decade, Professor Lotfi Zadeh offered a new theory. He was of the opinion that the classical theory placed too much emphasis on precision and, in turn, could not address complex systems. Lotfi Zadeh proposed fuzzy algorithms in 1968, fuzzy decision making in 1970, and fuzzy order in
In 1971, Max Black published an article about the logic of analysis titled "ambiguity" in Science Journal and for the first time, defined fuzzy sets with something which is today called Membership Function. In fact, in 1970s, fuzzy theory developed and its scientific applications emerged.

In 1990, Wang and Raz proposed a method for structuring variables based on linguistic data and stated that under these conditions, we can calculate quantitative values of qualitative-linguistic terms using membership functions. In this method, numerical values for fuzzy criteria are obtained using any of the four common fuzzy methods (Fuzzy Mode, Mid Level Number, Fuzzy Median and Fuzzy Average).

In 1995, Wang and Chen designed Fuzzy planning mathematical model and proposed a heuristic solution for economic design of statistical control diagrams.

In 1996, Khoo and Ho presented a framework for Fuzzy Quality Function Development System. In this model, customer satisfaction is stated in terms of numerical and linguistic variables.

In 1998, Glushokofsky and Florescu explained how to use fuzzy set theory in promoting qualitative tools where linguistic data are existent. They also recognized and introduced steps for standardizing qualitative-linguistic features.

In 2005, Shapiro used fuzzy regression to discuss the term structure of interest rates as well as using different Fuzzy members.

In 2009, Feng & Giles concentrated on Bayesian Fuzzy regression in order to select the most suitable model. They also showed

**Comparison Between Classical and Fuzzy Sets:**

Classical and fuzzy sets are subsets of a reference set. Classical set A has precise, explicit and definite borders; thus, any member of reference set X either belong to set A or not. In other words, an element’s membership to set A can be stated in a true-false statement for which values 1 or 0 are assigned. On the basis of the above definition, it can be concluded that membership degree of different elements in classical set A is either 1 or 0. Unlike classical sets, Fuzzy Sets’ borders are not clear and precise; thus, membership degree of different elements of fuzzy set \( A \) cannot be 1 or 0, but a number between 0 and 1. In summary, the difference between the concepts of membership in classical and fuzzy sets can be stated as follows.

\[
\mu_A(x) : x \rightarrow \{0, 1\}
\]

\[
\mu_A(x) : x \rightarrow [0, 1]
\]

**Qualitative Explanatory Variables and Their Application in Econometric Models:**

To study the effects of qualitative variables (such as revolution, earthquake, oil shocks, etc.) we can use virtual variables ranging from 0 to 1. In other words:

\[
DR = \begin{cases} 
1 & \text{If qualitative variable occurs} \\
0 & \text{If qualitative variable doesn't occur} 
\end{cases}
\]

The pattern variable can be studied in three ways:

- y-intercept
- gradient
- Both of them (gradient of the pattern and y-intercept)

Let dependent variable \( Y \) be explained by variables \( X \) and \( Z \) through the following pattern:

\[
Y = c_1 + c_2 X + c_3 Z + u
\]

Post estimation pattern will be as follows:

\[
\tilde{Y} = \tilde{c}_1 + \tilde{c}_2 X + \tilde{c}_3 Z + \tilde{c}_4 DR
\]

In this case, the effect of qualitative variable on the pattern will be y-intercept change.
In the second mode (gradient), qualitative pattern is placed in the model as follows:

\[ SDR = DR \times X \]
\[ Y = c_1 + c_2 X + c_3 Z + c_4 SDR + u \]
\[ Y = c_1 + c_2 X + c_3 Z + c_4 (DR \times X) + u \]

Post estimation model will be as follows:

\[ \hat{Y} = \hat{c}_1 + \hat{c}_2 X + \hat{c}_3 Z + \hat{c}_4 (DR \times X) \]

If qualitative variable occurs, then:

\[ \hat{Y} = \hat{c}_1 + \hat{c}_2 X + \hat{c}_3 Z + \hat{c}_4 X \]
\[ \hat{Y} = \hat{c}_1 + (\hat{c}_2 + \hat{c}_4) X + \hat{c}_3 Z \]

The effect of qualitative variable will be gradient change.

The third mode is a combination of the above two modes. In this mode, the effect of qualitative value results in change in both y-intercept and gradient.
Virtual DR variable application in each of the above modes increases accuracy and explanation power of the model. But in practice, qualitative variable effect on gradient and y-intercept cannot always be in one of the three above modes. For example, in Fig. 1, qualitative variable changed y-intercept. In a short period, this change has moved the model dependent variable curve while in most cases qualitative variable gradually changes y-intercept. Also, in Fig. 2, qualitative variable changed gradient of the model and its effect has been supposed as reliable while qualitative variable can change gradient in sectional basis. In other words, the effect of qualitative variable on dependent variable can be neutralized after some time. It should be mentioned that the effect of qualitative variables is not related to linear or non-linear nature of the pattern. An example of such states is mentioned in Fig. 4. As the figures show, in all cases, the effect of quantitative variable on y-intercept and gradient is continuous and unstable. In such cases, utilizing classical virtual variable (with values of zero and one), although may lead to increased explanation power of the model, does not always contain optimal results. Therefore, this paper aims at compensating shortcomings of utilization of virtual variables in such cases with the use of fuzzy numbers and functions.

Fig 4: Effect of qualitative variable on pattern in different cases
Introduction to Fuzzy System, Set, Numbers and Functions:

Fuzzy System:
Generally speaking, fuzzy system is very simple and includes input, output and processing. Input unit gets the output of sensors or inputs of other values and passes them to the related membership functions and calculates the respective membership values. Processing unit calls rules and produces a result for each of them, combines results together and finally produces output.

Fuzzy Set:
Linguistic variable \( x \) in the form of a word or expression forms a fuzzy set if each \( x_i \) in \( T(x) \) set gets a numerical identity with \( \mu(x) \). In fact, \( x \) is quality level criteria which accepts values between 0 and 1, i.e. it changes in the range \([0,1]\).

If \( X \) is a reference set and any of its members is shown with \( x \), fuzzy set \( A \) in \( X \) is stated with the following ordered pairs:
\[
A = \{(x, \mu_A(x)) | x \in X\}
\]

Where \( \mu_A(x) \) shows membership degree of \( x \) in \( A \), \( \mu_A(x) \) is membership function which shows membership degree of \( x \) to fuzzy set \( A \). The range of the function includes non-negative real numbers in the range \([0,1]\) in normal state (if the range of this function is 0 or 1 it turns to a precise set).

Fuzzy Numbers:
Fuzzy numbers are in fact natural generalizations of ordinary numbers. An ordinary number like \( a \) can be shown with the following membership function:
\[
\mu_a(x) = \begin{cases} 
1 & ; \text{if } x = a \\
0 & ; \text{if } x \neq a 
\end{cases}
\]

Therefore, any real number can be stated as a fuzzy number. The simplest fuzzy numbers are triangular fuzzy numbers. Membership faction of a triangular fuzzy number includes increasing and decreasing linear functions in triangular form with the following definition:

Fuzzy number \( A = (a_1, a_2, a_3) \) \((a_1, a_2, a_3) \) is a triangular fuzzy number if its membership function is as follows:
\[
\mu_A(x) = \begin{cases} 
0 & ; x < a_1 \\
(x - a_1) / (a_2 - a_1) & ; a_1 \leq x < a_2 \\
(a_3 - x) / (a_3 - a_2) & ; a_2 \leq x < a_3 \\
0 & ; x \geq a_3 
\end{cases}
\]

In an especial state, triangular fuzzy numbers are defined in terms of semi limited fuzzy numbers \((a_1, a_2, \infty)\) as follows:

**Fig 5:** Membership Function of Triangular Fuzzy Number \((a_1, a_2, a_3)\)
Another simple form of fuzzy numbers is trapezoidal after its membership function form. Fuzzy number $A=(a_1, a_2, a_3, a_4)$ is a trapezoidal fuzzy number if its membership function is as follows:

$$
\mu_i(x) = \begin{cases} 
0 & ; x \leq a_i \\
(x-a_i)/(a_2-a_1) & ; a_i < x \leq a_2 \\
1 & ; a_2 < x \\
\end{cases}
$$

for $(\infty, a_2, a_3)$,

$$
\mu_i(x) = \begin{cases} 
1 & ; x \leq a_2 \\
(a_3-x)/(a_3-a_2) & ; a_2 < x \leq a_3 \\
0 & ; a_3 < x \\
\end{cases}
$$

Fig 6: Membership Function of Trapezoidal Fuzzy Number $(a_1, a_2, a_3, a_4)$

In fact, it can be said that triangular fuzzy number is an especial case of trapezoidal function. Like triangular fuzzy number, different states of qualitative variables can be simulated using trapezoidal fuzzy numbers. Triangular and trapezoidal fuzzy numbers are very simple examples of fuzzy numbers which can be used to improve explanation power of regression models in different states.

Fuzzy Functions:

In terms of the form of membership function and fuzziness degree, fuzzy sets are divided into some categories. There are different methods for determining membership function of a fuzzy variable including three-stage method, increasing methods, fuzzy multi-state statistical methods, etc. A summary of increasing method is mentioned below.

If $A$ is a set including fuzzy concept and $A \in [a, b]$ then $\mu(u)$ can simply be defined as follows:

$$
\mu(u) = \mu_A(u) \quad u \in [a, b]
$$

Suppose $\Delta \mu$ is an accidental amount of $\mu$. Similarly, increasing amount of $\Delta \mu$ is determined by $\mu$. If $\Delta \mu$ is proportionate with $\Delta \mu$ and supposing that $\mu$ increases up to 1; then:
\[ \Delta \mu = k.\Delta \mu .u(1-\mu) \]

K is a fix value.
Dividing the above relation by \(\Delta \mu\) we will have:

\[ \Delta \mu /\Delta u = k.u(1-\mu) \]

When \(\Delta u \to 0\) then:
\[ \frac{d\mu}{du} = k.u.(1-\mu) = 1 - ce^{-k\frac{x^2}{2}} \]

C is a fixed integral number. Assigning numbers to c and k we can calculate. \(\Delta \mu\)

In terms of the related membership function it should be mentioned that there is no theory on the type and form of the functions. Infinite functions can be considered but, regarding applications, functions which match reality are selected.

The most important forms of fuzzy functions in terms of use are \(s\), \(z\), \(u\) and \(\pi\) categories. These functions are as follows:

**S Function:**
\[
S(x, a, b, c) = \begin{cases} 
0 ; x \leq a \\
2((x-b)/(c-a))^2 ; a \leq x \leq b \\
1-2((x-c)/(c-a))^2 ; b \leq x \leq c \\
1 ; x \leq c 
\end{cases}
\]

**Z Function:**
\[
Z(x, a, b, c) = \begin{cases} 
0 ; x < c \\
2((x-b)/(c-a))^2 ; b \leq x \leq c \\
1-2((x-c)/(c-a))^2 ; a \leq x \leq b \\
1 ; x \leq a 
\end{cases}
\]

*Fig. 7: S Function Curve*

a, b and c are lower limit, middle limit and upper limit of the function respectively.
Fig 8: Z Function Curve

\( \pi \) Function:
This function is in fact a combination of s and z functions. Other terms used for this function include convex or bell-shaped function:

\[
\pi(x, a, b, c) = \begin{cases} 
S_{(x,b-a,b-a/2,b)} & ; x \leq b \\
S_{(x,b,(b+ca)/2,c)} & ; x > b
\end{cases}
\]

Fig 9: \( \pi \) function curve

\( u \) Function:
This function is also a combination of s and z functions. This function is also called concave.

\[
v(x, a, b, c) = \begin{cases} 
Z_{(x,b-a,b-a/2,b)} & ; x \leq b \\
S_{(x,a,(b+ca)/2,c)} & ; x > b
\end{cases}
\]

Fig 10: \( u \) Function Curve
Depending on the effect of continuous qualitative variable on dependent variable one of the above membership functions can be used to study the effect of qualitative variable on the pattern. For this purpose, thresholds a, b and c in the related charts should be determined and instead of virtual variable with 0 and 1 range, depending on the effect of qualitative variable, one of the above membership functions should be used.

Identification of thresholds a, b and c of the above diagrams is of crucial importance in determining the precision of the patterns. Where the above thresholds may not be determined easily or the effect of qualitative variable results in convolution in the dependent variable pattern in a specific time period, in order to disambiguate the distances between levels, a, b and c are defined in a fuzzy way, in compliance with their natures. In other words, for any grade a membership function is determined and patterns in different modes are assessed and with regard to the assess parameters the best pattern is used to explain model realities.

Fig. 11: disorder in the pattern by Qualitative Variable

Regarding the above explanations, generally, the effect of continuous and instable qualitative variable on regression patterns can be considered as a fuzzy system. In this system, input is made with regard to the effect of qualitative variable on the pattern dependent variable.

Continuous qualitative variables (fuzzy) are entered in the model like discrete qualitative variables (classic). In the simple state:

\[ \hat{y}_i = \alpha + \beta_i X_i \]

Coefficients of this equation, i.e. \( \alpha \) \( \beta \), are fuzzy numbers and visible input variables, i.e. \( X_i \) are normal numbers. In this state, the aim is to assess the model based on visible data set in a way to have the best fitting equation.

Thus, the above equation can be rewritten as follows:

\[ \hat{y}_i = (a, c_0) + (\beta_i, c_i) X_i \]

To solve the above linear equation different algorithms have been suggested, one of which is transforming fuzzy linear regression problem to a linear programming problem. In an especial state, when skilled man defines membership function for continuous qualitative variables with regard to the manner of affecting pattern dependent variable process by continuous qualitative variable, the model can be assessed using OLS. Of course, the answer in this state where coefficients appear as fixed values is a special form of fuzzy linear regression.

**Estimation of Simple Econometric Model including Continuous Qualitative Variable:**

In this section, application of continuous and unstable qualitative variables in regression analysis is studied with an example:

The example is related to cement industry. In this example, the true price of cement during 1971 to 2004 has been considered as a function of production amount. Since cement pricing till 1991 was done by the government and later by manufacturers; thus, the role of manufacturers in cement pricing has been included in the model with virtual variable DR

\[ P = f (Q, DR) \]

In this relation:

P: true price of cement
R: annual cement production
DR: qualitative variable related to the role of manufacturers in cement pricing which is zero for the years 1971 to 1991 and one for the following years. The model was assessed in different modes and finally the best pattern for estimation is as follows:

\[ P = \alpha + \beta_1 Q + \beta_2 DR + \beta_3 SDR \]

The results out of estimation of the above model through normal Least Squares Method are as follows:

\[
\begin{align*}
\hat{P} &= 1444.77 - 0.08Q - 894.07DR + 0.093SDR \\
t &\begin{pmatrix} 18.43 \\ -10.12 \\ -5.08 \\ 8.82 \end{pmatrix} \\
R^2 &= 0.788 \quad F-Statistics = 37.27 \\
D.W &= 2.21
\end{align*}
\]

In fuzzy mode, fuzzy linear regression equation will be as follows:

\[ \tilde{p} = \tilde{\alpha} + \tilde{\beta}_1 Q + \tilde{\beta}_2 DR + \tilde{\beta}_3 SDR \]

Triangular numbers \( \tilde{\alpha} = (\alpha, c_0, c_1) \), \( \tilde{\beta}_1 = (\beta_1, c_1, c_2) \), and \( \tilde{\beta}_3 = (\beta_3, c_2, c_3) \) are the coefficient of the equation in which \( \alpha, \beta_1, \beta_2 \) and \( \beta_3 \) are centers and \( c_0, c_1, c_2 \) and \( c_3 \) are widths of the said coefficients. Different algorithms have been suggested for solving fuzzy linear regression problem one of which is transforming fuzzy linear regression problem to linear programming problem. In this method, the aim of regression model is determination of optimal amounts of \( \tilde{\alpha} \), \( \tilde{\beta}_1 \), \( \tilde{\beta}_2 \) and \( \tilde{\beta}_3 \), in a way to have membership degree of fuzzy output variable for all data be larger than a definite amount such as \( h \) which is determined by the user. Transforming fuzzy linear regression to linear programming problem, object function and its constraints will be as follows:

\[
\begin{align*}
\min \quad Z &= \left[ 34C_0 + C_1 \sum_{i=1}^{34} Q + C_2 \sum_{i=1}^{34} DR + C_3 \sum_{i=1}^{34} SDR \right] \\
\alpha + \beta_1 \sum_{i=1}^{34} Q_i + \beta_2 \sum_{i=1}^{34} DR_i + \beta_3 \sum_{i=1}^{34} SDR_i - 0.5 \left[ C_0 + C_1 \sum_{i=1}^{34} Q_i + C_2 \sum_{i=1}^{34} DR_i + C_3 \sum_{i=1}^{34} SDR_i \right] &\leq P_i \\
\alpha + \beta_1 \sum_{i=1}^{34} Q_i + \beta_2 \sum_{i=1}^{34} DR_i + \beta_3 \sum_{i=1}^{34} SDR_i - 0.5 \left[ C_0 + C_1 \sum_{i=1}^{34} Q_i + C_2 \sum_{i=1}^{34} DR_i + C_3 \sum_{i=1}^{34} SDR_i \right] &\geq P_i
\end{align*}
\]

Solving the problem using winQSB software, regression equation for \( h = 0.5 \) will be as follows:

\[ \tilde{P} = (0, 1562/2) + (0/16, 0/38)Q + (1534/5, 650)DR + (0, 0.27)SDR \]

In this way, regarding the effect of qualitative variable on cement price, the effect of qualitative variable has been assumed to be continuous:

\[ \text{Fig 12: Process of True Cement Price in Last 34 Years} \]
As Fig. 12 shows, cement price process was almost fixed till 1991. Afterwards, when manufacturers began cement pricing, the chart was convulsed for a short period and pattern gradient was reversed. Convulsion in cement pricing process in the respective years seems to be out of pricing by manufacturers; thus, the effect of this continuous qualitative variable on pricing is considered as a fuzzy system. With regard to the fact that the vagueness of the model estimation can be caused by variability in qualitative variable thresholds or how this variable affects dependent variable process; thus, coefficients of the estimated model in a definite range are variable. But, regarding pattern dependent variable process, continuous qualitative variable membership function can be defined as follows:

\[
\mu_i(x) = \begin{cases} 
0 & ; x < 20 \\
\frac{(x-20)}{3} & ; 20 \leq x < 23 \\
-0.09x+3.07 & ; 23 \leq x < 27 \\
0.64 & ; 27 \leq x
\end{cases}
\]

Having continuous qualitative variable membership function determined, the model can be solved using normal Least Squares Method:

Replacing DR and estimating pattern using normal Least Squares Method we will have:

\[
P = 1494.14 - 0.09Q - 1917.94DR + 0.17SDR
\]

\[
\begin{align*}
t1: & (26.37) \ (-15.05) \ (-10.21) \ (13.99) \\
R^2: & 0.89 \quad F-Statistics: 82.21 \quad D.W=1.98
\end{align*}
\]

This is a special form of fuzzy estimation model in which membership function is achieved using the experience of a skilled person.

Comparing the Results of Estimated Models:

The most important parameters made from pattern estimation are entered in the following table for comparison purpose:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Classical Pattern</th>
<th>Fuzzy Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.788</td>
<td>0.890</td>
</tr>
<tr>
<td>$F$</td>
<td>37.27</td>
<td>82.21</td>
</tr>
<tr>
<td>$t_1$</td>
<td>-10.11</td>
<td>15.05</td>
</tr>
<tr>
<td>$t_2$</td>
<td>-5.08</td>
<td>-10.21</td>
</tr>
<tr>
<td>$t_3$</td>
<td>8.82</td>
<td>13.99</td>
</tr>
<tr>
<td>$t_4$</td>
<td>18.43</td>
<td>26.37</td>
</tr>
<tr>
<td>D.W</td>
<td>2.21</td>
<td>1.99</td>
</tr>
</tbody>
</table>

As table 1 shows, application of fuzzy system in the place of virtual variables 0 and 1 has remarkably increased the significance of pattern parameters and the entire pattern. Changes of statistics $F$ and $t$ prove this claim. In this example, D.W has noticeably improved. Furthermore, application of fuzzy numbers in the place of 0 and 1 virtual variables has improved explanation power of model ($R^2$) from 0.788 to 0.89.
Although $R^2$ is not the only criterion for choosing the best model; but in cases where there is no criterion except $R^2$ for comparison purpose, $R^2$ can be used for this purpose.

The above example shows the application of continuous qualitative variables in econometric patterns with a fuzzy system approach. In general, estimation of different models with qualitative variables has shown that, regarding the effect of qualitative variable on the pattern, in discrete state, the best patter is classic virtual variable with 0 and 1 values and in continuous state, the best pattern is fuzzy virtual variable.

Conclusion Remarks:

Whereas the application of qualitative variables in econometric patterns estimation is usually based on classical mathematics and amount with two values, this paper has aimed at introducing fuzzy sets, numbers and function and to show their application (in continuous state) in patterns estimation and to drift approaches from two-value amounts to multi-value and fuzzy amounts. As was stated earlier, this article is not to replace 0 and 1 values with fuzzy amounts. In fact, this paper holds that selection of classical or fuzzy virtual variable should be done based on the effect of qualitative variable (in either discrete or continuous forms). Application of virtual variables in fuzzy form (multi values) or classical form (two values) can be determined with regard to the effect of qualitative variable (discrete or continuous) on pattern dependent variable. Multi-value variables (fuzzy) can be used in cases where the effect of qualitative variable (continuous) results in convolution in the pattern in a specific period, using fuzzy average and median concepts. In this paper, application of continuous qualitative variables in regression analyses was introduced as a fuzzy system. The fuzzy system introduced here, regarding the effect of virtual variable on the pattern and also qualitative variable effect thresholds, can have several inputs. Designing a fuzzy processing system (in the form of fuzzy if-then rules), one can obtain the best assessment pattern from the output.

REFERENCES

Shapiro, A., 2005. Fuzzy Regression Model , Penn State University.