WHERE, WHEN, AND WHY FUZZY REGRESSION ANALYSIS
O. BASKAN, S. SENOL, S. PEKER
Department of Statistics
Ege University
Ege University Faculty of Science, Department of Statistics, 35100 Bornova, Izmir
TURKEY

Abstract: Where fuzzy regression can be applied and, in which conditions fuzzy regression method more appropriate tool for the investigations are identified in this paper. The contrast between fuzzy regression and ordinary regression analysis and three approach of fuzzy regression are summarized.

Key-Words: fuzzy regression, vaguness, fuzziness, randomness, insufficient data, assumptions.

1 Introduction
As you can regression is a popular methodology for expressing the functional relationship on two or more related variables. With the mathematical form, the value of one variable can be predicted from the values of the others. It is function that makes regression analysis one of the most useful techniques in OR applications [19].

In the research of complex systems, such as the systems existing in biology agriculture, engineering technique and social economy, we frequently cannot get the exact numerical data for the information of systems because of the complexity of systems themselves, the vagueness in people’s thinking and judgement and the influence of various uncertain factors existing in boundary environment around the systems. For this situation, the traditional least squares regression may not be applicable. How to deal with this kind of curve fitting problems containing fuzzy information has been attracting the attention of many researchers.

In 1982, Tanaka et al used linear programming techniques to develop a fuzzy linear regression model. Resembling traditional least squares regression Diomand established fuzzy linear least squares models by defining the metric on triangular fuzzy number spaces [1]. Subsequently, there have been many other studies which can be roughly divided into two approaches: linear-programming-based methods (possibilistic approach), fuzzy least-squares methods (least squares approach) [5]. Yang M., Liu H. H., proposed new types of fuzzy least-squares algorithms with a noise cluster for interactive fuzzy linear regression models. These algorithms are robust for the estimation of fuzzy linear regression models, especially when outliers are present [12]. Lee and Chen introduced a generalized fuzzy linear regression model and proposed a nonlinear programming model to identify the fuzzy parameters and their vagueness for the generalized regression model. Yi H. and Hong D. show that proposed nonlinear programming model is invalid [21]. Hong D., Hwang C., Ahn C., describe a ridge estimation of fuzzy multiple linear regression model, using the ridge learning algorithm in the Lagrangian dual space [22]. D’Urso P., suggest regression models with crisp or fuzzy inputs and crisp or fuzzy output by taking into account a least-squares approach [23]. Wu H., propose a method for obtaining the fuzzy estimates of regression parameters with the help of “Resolution Identity” in fuzzy sets theory [11]. Tran L., Duckstein L., develop a multiobjective fuzzy regression model (MOFR). The MOFR model combines central tendency and possibilistic properties of statistical and fuzzy regressions and overcomes several shortcomings of these approaches [13]. Wu B., Tseng N., construct a fuzzy regression model (least squares approach) by fuzzy parameters estimation by using the fuzzy samples [14]. Nather W., Wünsche A., obtain formulas which are analogous to the classical structure in special cases and by use of Hukuhara’s difference between fuzzy sets [15]. Yang M. and Lin T. propose two estimation methods a long with a fuzzy least-squares approach for fuzzy inputs and output [16].

In both approaches, the notion of “best fit” incorporates the optimisation of a functional associated with the problem. When the observations are defined by possibility distributions or fuzzy member functions rather than the classical probability distributions, the associated regression models are called possibility or fuzzy regression model [19]. In particular, in the possibilistic approach, “this functional takes the form of a measure of the spreads of the estimated output, either as a weighted linear sum involving the estimated coefficients in linear regression, or as quadratic form in the case of exponential possibilistic regression”. In the
least-squares approach “the functional to be minimized is an L2 distance between the observed and estimated outputs. This reduces to a class of quadratic optimization problems and constrained quadratic optimization”[23]. Thus, Tanaka’s approach to possibility regression analysis, instead of the measure of best fitting by residuals, uses linear programming inclusion relations. However, the fuzzy least-squares approach to possibility regression analysis, does not consider inclusion relations, directly uses the best fitting measure by residuals and information included in the input-output data under fuzzy consideration [20].

2. Problem Formulation:

In the real world, the data sometimes cannot be recorded or collected precisely. For instance, the water level of a river cannot be measured in an exact way because of the fluctuation, and the temperature in a room is also not able to be measured precisely because of the similar reason. Therefore fuzzy sets theory is naturally to be appropriate tool in modeling the statistical models when the fuzzy data have been observed. The more appropriate way to describe the water level is to say that the water level is around 30 m. The phrase “around 30 m” can be regarded as a fuzzy number 30. Since Zadeh introduced the concept of fuzzy sets, the applications of considering fuzzy data to the regression models have been proposed in the literature [11].Vague or fuzzy data find application in several fields, such as psychometry, reliability marketing, quality control, ballistic, ergonomy, image recognition, artificial intelligence etc. A typical problem where vague data arise is that of assigning numbers to subjective perceptions or to linguistic variables (such as “enough”, “good”, “sufficient”, etc.) In fact, there are many cases where observations cannot be known or quantified exactly, and, thus, we can only provide an approximate description of them, or intervals as close them. For instance, “in measuring the influence of character size on the reading ability a video display terminal [...] the reading ability of the subject, which is essentially the experimental output, depends upon his/her eyesight, age, the environment, individual responses, and even how tired is the individual. Some of these factors cannot be described accurately and [...] the best description of these kinds of output is that they are fuzzy outputs”[4].

In the empirical study assumptions may hardly be realized, since there are many observations that experience linguistic or vague data inside the classical type[18]. And previous research has shown that Fuzzy regression might perform better than statistical regression in the following cases:

1) When the data set is insufficient to support statistical regression analysis.
2) When statistical distribution assumption cannot be justified.
3) If the representativity of the regression model is poor.
4) When the human judgements are involved. (i.e. inputs/outputs are fuzzy numbers)
5) If the errors are associated with the indefiniteness of the model structure and with the vagueness of human perception of the model (in contrast with the statistical case where the errors are associated with observations) [3].

The contrast between ordinary statistical regression and fuzzy regression analysis

There are many situations where observations cannot be described accurately as, for instance, when they depend on environment conditions or individual responses. In such cases, we can only provide an approximate description of them, or an interval to enclose them. Notice that we are concerned with a kind of uncertainty which is different from randomness and that is sometimes referred to as vagueness. For example, when “we are measuring a current by a digital measuring apparatus, we receive data which are theoretically continuous but which, in reality, are discrete. The reason of that is that measurement proceedings are not precise [12].

In conventional regression analysis, deviations between the observed values and the estimates are assumed to be due to random errors. Thus, statistical techniques are applied to perform estimation and inference in regression analysis. However, the deviations are sometimes due to the indefiniteness of the structure of the system or imprecise observations. The uncertainty in this type of regression model becomes fuzziness, not randomness. Since Zadeh proposed fuzzy sets, fuzziness has received more attention[20] Both fuzzy regression and ordinary regression only consider part of the totality uncertainty. In fact, randomness and fuzziness are two different kinds of uncertainty [6].

In the ordinary regression analysis the performance and validity of least squares method is degraded if its assumptions, such as independence and homogeneity of error terms are violated. Fuzzy linear regression aims to model vague and imprecise phenomena using fuzzy model parameters. Fuzzy regression is different from ordinary least squares in the sense that it is a nonstatistical method [2].

The fuzzy linear regression

Different fuzzy regression models are obtained the
fitting criterion used.

**Fuzzy regression using minimum fuzziness criterion**

Tanaka et al. proposed the first linear regression analysis with a fuzzy model.

\[ \tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X_1 + ... + \tilde{A}_N X_N \]  

(1)

where \( \tilde{A}_0 \) is a fuzzy intercept coefficient, and \( \tilde{A}_i \) is a slope fuzzy slope coefficient. Each fuzzy parameter \( \tilde{A}_i = (m_i, c_i) \) is expressed as symmetrical triangular membership function, which consist of fuzzy center \( m_i \) and fuzzy half-width \( c_i \). Other membership function forms can be used as well.

The term \( h \) is referred to as a measure of goodness of fit or a measure of compatibility between data and a regression model.

Each observed data sets, which can be fuzzy \( \tilde{Y}_i \) or crisp datum \( Y_i \), must fall within estimated \( \tilde{Y} \) at \( h \) level.

Tanaka et al. formulated the fuzzy regression objective as the following linear programming problem.

**minimize** \[ S = nc_0 + c_1 \sum_{i=1}^{n} |X_i| \]  

subject to \( c_0 \geq 0, c_1 \geq 0, \)

\[ \sum_{j=0}^{i} m_i X_{ij} + (1-h) \sum_{j=0}^{i} c_i |X_{ij}| \]

\[ \geq Y_i + (1-h)e_i \quad \text{for } i = 1 \text{ to } n, \]  

(3)

\[ \sum_{j=0}^{i} m_i X_{ij} - (1-h) \sum_{j=0}^{i} c_i |X_{ij}| \]

\[ \geq Y_i - (1-h)e_i \quad \text{for } i = 1 \text{ to } n, \]  

(4)

where \( S \) is the total fuzziness of the regression model.

Eq. (3) and (4) can deal with observed fuzzy datum \( \tilde{Y}_i = (Y_i, e_i) \), where \( Y_i \) is the fuzzy center, and \( e_i \) is the fuzzy half-width [6].

**Fuzzy least-squares regression**

1) **Fuzzy least-squares regression using the maximum compatibility criterion**

Celmins proposed an approach for fuzzy least squares regression, based on compatibility measure between data and a fitted model. Defining \( \mu_{\tilde{A}}(X) \) and \( \mu_{\tilde{B}}(X) \) as the membership functions of two fuzzy quantities \( \tilde{A} \) and \( \tilde{B} \), Clemins suggested compatibility measure between \( \tilde{A} \) and \( \tilde{B} \) as \( \gamma(\tilde{A}, \tilde{B}) \).

Let \( \gamma_i \) be measure of the compatibility between each datum and the fitted model, the objective function for the data is to minimize the sum of the squares of the deviations (W) as

\[ W = \sum_{i=1}^{m} (1 - \gamma_i)^2 \]

The final formula for the fuzzy least-squares regression using maximum compatibility is as follows:

\[ \tilde{Y} = \tilde{A}_0 + \tilde{A}_1 X \]

\[ = m_0 + m_1 X \pm \sqrt{c_0^2 + 2c_1 X + c_1^2 X^2} \]  

(5)

The coefficients \( m_0 \) and \( m_1 \) are obtained by a weighted least-squares regression, and the term \( 1/(\text{datum fuzziness})^2 \) is used as the weighted assigned for each datum [6].

2) **Fuzzy least-square regression using the minimum fuzziness criterion:**

Savic and Pedrycz formulated the fuzzy regression method the least-squares principle and minimum fuzziness criterion. The method is performed in two consecutive steps. The first step uses ordinary least-squares regression to find fuzzy center values of fuzzy regression coefficients. The second step uses the minimum fuzziness criterion to find the fuzzy widhts of fuzzy regression coefficients [6].

**Interval Regression**

According to the this method, the fuzzy data and fuzzy regression coefficients are treated as interval numbers. The fuzzy regression coefficients are determined such that all fuzzy outputs are within a fuzzy regression model. The following linear programming is used to solve for the fuzzy regression coefficients \( \tilde{A}_0 = (m_0, c_0) \) and \( \tilde{A}_i = (m_i, c_i) \):

**minimize** \[ S = nc_0 + c_1 \sum_{i=1}^{n} |Y_i| \]  

subject to \( c_0 \geq 0, c_1 \geq 0, \)

\[ (m_0 - c_0) + (m_i - c_i) \leq Y_{i,L} \]

\[ \text{for } i = 1 \text{ to } n, \]  

(7)

\[ (m_0 - c_0) + (m_i - c_i) \leq Y_{i,U} \]

\[ \text{for } i = 1 \text{ to } n, \]  

(8)

where \( Y_{i,L} \) and \( Y_{i,U} \) are the lower and upper limits for each fuzzy datum, respectively [6].

**3. Conclusion**

In the empirical study assumptions may hardly be realized, since there are many observations that limited and experience linguistic or vague data , and variables that interacting in an uncertain, qualitative, and fuzzy way. In this situations nonstatistical method, fuzzy regression is an alternative method of ordinary
regression analysis. Fuzzy regression analysis is dealing with fuzziness while ordinary regression analysis is dealing with randomness.

References: