Fuzzy Least Squares Regression Analysis for Social Judgment Study

Kazuhisa Takemura

Department of Psychology, Waseda University
1-24-1 Toyama, Shinjuku-Ku, Tokyo 162-8644, Japan
E-mail: kazupsy@waseda.jp

[Received October 12, 2004; accepted June 6, 2005]

Social judgment data was analyzed using fuzzy least squares regression analysis based on the extension principle. The proposed analysis is new fuzzy least squares regression analysis in which input data, output data, and coefficients are represented by L-R fuzzy numbers. To evaluate data fitness, we propose a fuzzy version of a squared multiple correlation ($R^2$) and conducted an experiment to determine the effect of partial attribute information on the overall evaluation of desirability for a marital partner and personality of a person using fuzzy rating to measure vagueness in social judgment. Participants in the experiment were 90 university students. The result of marital partner judgment indicated that the fuzzy weight of consideration was higher than physical attraction on overall desirability evaluation for a marriage partner. The result of personality judgment indicated that the fuzzy weight of kindness was slightly higher than responsibility in overall desirability for a marriage candidate.

Keywords: fuzzy regression analysis, least squares method, fuzzy rating, fuzzy data, social judgment

1. Introduction

Judgment in social situations often involves vagueness concerning confidence. People may not be able to make judgments without using confidence intervals. Fuzzy rating has been proposed and developed to measure vagueness in human judgment [2]. In fuzzy rating, respondents select a representative rating on a scale and indicate lower or upper rating points depending on the relative vagueness of their judgment (Fig.1). Fuzzy rating would be useful, for example, in measuring perceived temperature indicating the representative value and lower or upper values. This rating allows for asymmetry and solves the problem, identified by Smithson [6], that researchers have in arbitrarily deciding the most representative value from a range of scores. By making certain simplifying assumptions (not uncommon within fuzzy set theory), the rating is viewed as an L-R fuzzy number (Fig.2), enabling the use of fuzzy set theory [3, 4]. Note in Fig.2 that representations of variables are abbreviated as follows: $x_{ij}^L$ for $x_{ij(0)}^L$, $x_{ij}^M$ for $x_{ij(0)}^M$, $x_{ij}^R$ for $x_{ij(1)}^R$. To analyze fuzzy rating data, it would be more useful to apply fuzzy linear regression analysis in which observed values and estimated values are assumed to have fuzziness, rather than ordinal linear regression analysis in which all values are assumed to be crisp. Although the original version of fuzzy linear regression analysis proposed by Tanaka et al. [12] assumed that while output data is fuzzy numbers, input data is not fuzzy numbers, Sakawa and Yano [5] formulated three types of multiobjective programming problems for obtaining fuzzy linear regression models and developed possibilistic linear regression analysis in which both input and output data are fuzzy numbers. Possibilistic linear regression analysis, which is a version of multiobjective fuzzy linear analysis, is considered to be a generalized method of previous
fuzzy regression models.
This could be very effective for human sciences such as psychology, sociology, and ergonomics, because most input and output data for such sciences is considered fuzzy.

Unfortunately, some studies using fuzzy rating data indicated that the predicted variable for probabilistic linear analysis for fuzzy input-output data had too great a spread of fuzzy number for meaningful interpretation [7–9]. Practically speaking, it takes many resources to calculate the solution for multiobjective fuzzy linear models by computer.

We use alternative fuzzy regression analysis using the least squares method developed by Diamond [1] and Takemura [9–11] to determine the effect of partial information on interpersonal attributes on overall evaluation for social judgment using fuzzy rating to measure vague-

2. Fuzzy Least Squares Regression Analysis

A set of fuzzy input-output data for i-th observation is defined by:

\( (Y_i; X_{i0}, X_{i1}, X_{i2}, \ldots , X_{in}), \quad i = 1, 2, \ldots , n \) \( (1) \)

where \( Y_i \) is a fuzzy dependent variable and \( X_{in} \) is a fuzzy independent variable represented by L-R fuzzy numbers, \( X_{i0} \) is 1. For simplicity, we assume that \( Y_i \) and \( X_{ij} \) are positive for any membership value, \( \alpha \in (0,1) \).

The fuzzy linear regression model in which both input and output data is fuzzy numbers is represented as follows:

\[ \bar{Y}_i = A_0 + A_1 \times X_{i1} + \cdots + A_m \times X_{im} \ldots \ldots (2) \]

where \( \bar{Y}_i \) is a fuzzy predicted value, \( X_{im} \) is a fuzzy independent variable, \( X_{i0} = 1 \), and \( A_j (j = 0, 1, \ldots , m) \) is a fuzzy regression parameter represented by L-R fuzzy number, and \( \otimes \) is the product operator based on Zadeh’s extension principle [13].

Note that although the explicit form of the membership function of \( \bar{Y}_i \) cannot be directly obtained, the \( \alpha \)-cut of \( \bar{Y}_i \) is obtained from the result of Nguyen’s theorem [4].

Let \( \underline{z}_{i(\alpha)}^l \) be the lower value of the \( \alpha \)-cut \( \bar{Y}_i \), and \( \underline{z}_{i(\alpha)}^R \) be a upper value of the \( \alpha \)-cut of \( \bar{Y}_i \). Then,

\[ Z_i = [\underline{z}_{i(\alpha)}^l, \underline{z}_{i(\alpha)}^R], \quad \alpha \in (0, 1) \ldots \ldots (3) \]

where

\[ \underline{z}_{i(\alpha)}^l = \sum_{j=0}^{m} \left\{ \min \left( a_{ij(\alpha)}^L x_{ij(\alpha)}, a_{ij(\alpha)}^R x_{ij(\alpha)} \right) \right\} \ldots (4) \]

\[ \underline{z}_{i(\alpha)}^R = \sum_{j=0}^{m} \left\{ \max \left( a_{ij(\alpha)}^L x_{ij(\alpha)}, a_{ij(\alpha)}^R x_{ij(\alpha)} \right) \right\} \ldots (5) \]

In Eq.(4) above, \( a_{ij(\alpha)}^L x_{ij(\alpha)} \) is a product between a lower value of the \( \alpha \)-cut of fuzzy coefficient for \( j \)-th attribute and a lower value of the \( \alpha \)-cut of fuzzy input data \( X_{ij} \), \( a_{ij(\alpha)}^L x_{ij(\alpha)}^L \) or \( a_{ij(\alpha)}^R x_{ij(\alpha)}^R \) is defined the same way.

To define dissimilarity between a predicted value and an observed value of the dependent variable, we adopt the following indicator \( D_i (\alpha)^2 \).

\[ D_i (\alpha)^2 = (\bar{Y}_i - \underline{z}_{i(\alpha)}^l)^2 + (\bar{Y}_i - \underline{z}_{i(\alpha)}^R)^2 \ldots (6-1) \]

Definition by Eq.(6-1) is applied to interval data and L-R fuzzy numbers, i.e., Eq.(6-1) represents a sum of squares for distance between interval data.

To generalize, a dissimilarity indicator representing a square of distance for L-R fuzzy numbers is written as follows:

\[ D_i^2 = \sum_{j=0}^{k} w_j (\underline{z}_{i(\alpha)}^l)^2 + (\bar{Y}_i - \underline{z}_{i(\alpha)}^l)^2 + (\bar{Y}_i - \underline{z}_{i(\alpha)}^R)^2 \ldots \ldots \ldots \ldots (6-2) \]

where \( \alpha_j = j/k, \quad j = 0, \ldots , k, \) and \( w_j \) is a weight for the \( j \)-th level.

In the case of a triangular fuzzy number with \( w_j = 1 \), the above equation is approximate as:

\[ D_i^2 = (\bar{Y}_i - \underline{z}_{i(\alpha)}^l)^2 + (\bar{Y}_i - \underline{z}_{i(\alpha)}^l)^2 + (\bar{Y}_i - \underline{z}_{i(\alpha)}^R)^2 \ldots \ldots \ldots \ldots (6-3) \]

If all fuzzy parameters are crisp, the above definition of the dissimilarity indicator in Eq.(6-3) is equivalent to the indicator by Diamond [1]. Fuzzy parameters are, however, generally assumed to be fuzzy in this equation.

We propose estimating fuzzy coefficients using the minimization of sum of \( D_i^2 \) respecting i. Assuming \( D_i^2 \) in Eq.(6-3), we propose solving the following optimizing problem:

Objective function:

\[ \text{Min} \sum_{i=1}^{n} D_i^2 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (7) \]

Subject to:

\[ a_{j(\alpha)}^L \leq 0, \quad j \in J_1 \ldots \ldots \ldots \ldots (8) \]

\[ a_{j(\alpha)}^R \leq 0, \quad j \in J_2 \ldots \ldots \ldots \ldots (9) \]

\[ a_{j(\alpha)}^L \leq 0, \quad j \in J_3 \ldots \ldots \ldots \ldots (10) \]

\[ -a_{j(\alpha)}^R \geq 0 \ldots \ldots \ldots \ldots (11) \]

where

\[ j \in \{ 0, \ldots , m \} = J_1 \cup J_2 \cup J_3, \]

\[ J_1 \cap J_2 = \phi, \quad J_2 \cap J_3 = \phi, \quad J_3 \cap J_1 = \phi, \]

\[ \underline{z}_{i(\alpha)}^L = \sum_{j_{i1} \in J_1} a_{ij(\alpha)}^L x_{ij(\alpha)}^L + \sum_{j_{i1} \in J_2} a_{ij(\alpha)}^R x_{ij(\alpha)}^L \ldots \ldots \ldots \ldots (12) \]

Since the sign of the parameter of linear regression models is not often known in advance, \( J_1, J_2, \) and \( J_3 \) are variable, not constant.

Estimated coefficients are derived through quadratic programming. The proposed fuzzy least squares method is shown in Fig.3. Note that representations of variables are abbreviated as follows: \( x^L \) for \( x_{ij(0)}^L \), \( x^R \) for \( x_{ij(0)}^R \), \( x^M \) for \( x_{ij(1)}^R = x_{ij(1)}^R \).