A Compositional regression model based on fuzzy weighted evaluation method

Mengqi Gong, S Cheng*

School of Economics and Management, China University of Geosciences, Lumo Road 388, 430074, Wuhan, China

Received 1 October 2014, www.cmnt.lv

Abstract

This paper adopts 6 indexes from regression model evaluation criterion, regards the membership grade of the optimal index set of the given equation distance as the weight, linearly weights the corresponding equation and establishes the compositional regression model with the fuzzy evaluation method. The practical application indicates this method gets better fitting and predicting effect compared with the traditional regression model and it is a relative and perfect optimization method.

Keywords: fuzzy weight; membership grade; compositional regression model

1 Introduction

Model specification is the key to the application of the regression equation. In case of error in model specification, the parameter estimation obtained with the ordinary least square (OLS) method has no minimum variance and unbiasedness. The traditional model specification selection considers from the angle of variable selection whether the irrelevant variable added or the missing significant variables results in the error in model specification [1-3]. Next, the methods such as Ramsey Regression Equation Specification Error Test (RESET), DW test, Lagrange multiplier test are adapted to test whether the error exists in the model specification. Then the instrument variable method is adopted to approximately fit the objective model. However, the largest disadvantage of the instrument variable method is that the estimated result is very sensitive to the sample size and the instrument variables selected [4-5]. Based on the regression model evaluation criterion, this paper adopts the thought of fuzzy evaluation, regards the membership grade of the optimal index set of distance as the weight and linearly combines and sets the model and establishes the compositional regression model. It is indicated by the practical application results that the fuzzy weighted evaluation method proposed by this paper has better fitting and predicting effects compared with the traditional and single regression model specification.

2 The approach to fuzzy evaluation and weighted optimization

Given there are n evaluation indexes for evaluating m schemes, the characteristic quantity matrix X of the original indexes can be obtained:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = (x_{ij})^{m \times n} \qquad (2.1)$$

$i = 1,2,.....m; j = 1,2,.....n.$ Where x_{ij} is the index characteristic quantity of the factor of evaluation for its scheme, which includes the

* ***Corresponding author's*** e-mail: chengsheng@cug.edu.cn

greater the priority type index and the smaller the priority type index. The membership grade of a single factor of evaluation to the corresponding element of the optimal index set can be calculated with different calculation formulas respectively. The formula of grade of relative membership is described as follows:

To the greater the priority type index:

$$r_{ij} = x_{ij} / \max(x_{.j}) \qquad (2.2)$$

To the smaller the priority type index:

$$r_{ij} = \min(x_{.j}) / x_{ij} \qquad (2.3)$$

Where, x_j is the set of x_{ij} (i=1, 2,..., m), in this way, there is no dimension and the adjusted index data are of comparability. Consequently, Matrix X is changed into the matrix of index membership grade R.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} = (r_{ij})^{m\times n} \qquad (2.4)$$

$i = 1,2,.....m; j = 1,2,.....n.$ The matrix of index membership grade R is the basic information of fuzzy compositional evaluation and the basis for optimization. m indexes of the optimal index set should be the max. Value of the corresponding index membership grade of all schemes, denoted with the vector as:

$$G = (\max(r_1), \max(r_2), ...\max(r_j)), j = 1,2,.....n)$$

(2.5)Considering the different functions of each factor of evaluation in the actuality, given the weighted vector of nth index is: $w = (w_1, w_2, ...w_n)$ .Then the distance between the ith scheme and the optimal index set can be denoted with generalized euclidean distance u_i:

$$u_i = \|w(G - r_i)\| = \sqrt{\sum w_j (G_j - r_{ij})^2} \qquad (2.6)$$

As to regression prediction, it is deemed that as long as a certain scheme is of priority in a certain index, a certain combination of various schemes may have the synthetic advantages, which is the existence of weighted optimization [6-8]. Based on this, it is held by us that the distance u_i between the ith scheme and the optimal index set can be deemed that this scheme lives under the optimal index set in form of membership grade u_i, so it can be regarded as the weight of each equation for linear combination to realize the better prediction effect.

3 A fuzzy weighted compositional regression model

In the classical regression analysis, the significance test has some disadvantages. Firstly, as to two F values with small difference, the significance test of regression effect is becoming uncertain. Secondly, the type of the regression model is artificially selected based on the experience and the distribution trend of the scatter diagram, so it is of great subjectivity. Thirdly, the test criterion just takes into account a certain factor, for example, the sum squared error, so the test is very rough and it cannot satisfy the actual need.

Meanwhile, one of the most important objectives for data analysis and establishment of the regression equation is prediction and control. As a result, the fitting precision is becoming especially important. This paper applies the above-mentioned fuzzy compositional evaluation and weighted optimization thought to the regression model to improve the fitting precision.

3.1 EATABLISHMENT OF ORIGINAL REGRESSION EQUATIONS

The common regression model includes the following types: hyperbolic type, power function type, exponential function type, log curve type, s-curve type, multinomial type and trigonometric function type for selection. Once the regression equation type is selected, the related technique such as the least square method and the orthogonal polynomial method can be adopted for curve fitting of the scatter points and the fitting equation is:

$$y = f_i(x). \quad (3.1)$$

3.2 EASTABLISHMENT OF COMPOSITIONAL EVALUATION MODEL OF DEGREE OF FITTING

The disadvantages of the classical regression analysis have been discussed in the foregoing paragraphs. In order to overcome these unreasonable phenomena, this paper adopts the following 6 kinds of indexes for comprehensive evaluation of degree of fitting. (1) Variance, namely $\sum(\hat{y}_i - y_i)^2 / n$, which reflects the proximity of the fitting curve and the scatter point as a whole; (2) Maximum error, namely $\max\left|\hat{y}_i - y_i\right|$, which reflects the proximity of the range between the theoretical value and the measured value of the fitting curve; (3) Least error, namely $\min\left|\hat{y}_i - y_i\right|$, which reflects the precision that the fitting curve may reach up; (4) Absolute value of average error, namely $\left|\Sigma(\hat{y}_i - y_i)/n\right|$, which reflects the proximity of the fitting curve and the theoretical value as a whole, however, different from the sum squared error, it holds that the point errors are "equivalent"; (5) Average absolute relative error, namely $\left|\Sigma(\hat{y}_i - y_i)/ny_i\right|$, which reflects the fitting precision of the measured value and the theoretical value of the regression equation; (6) Positive-negative rate of change, namely the absolute value of the differences in number of the data at both sides of the fitting curve, which reflects the degree of uniformity of the positive-negative distribution of the error value of the regression value. In order to guarantee there is no negative correlation in the calculation of the degree of membership, the absolute values of the average error, average relative error and positive-negative rate of change are adopted.

The foregoing 6 kinds of indexes are the basis of fuzzy compositional evaluation. Since different indexes account for different proportions in different regression curves, in order to reflect the differences in the requirements, the non-equal weight structure is adopted in the model and the weight can be calculated in combination with the actuality with analytic hierarchy process or binary pairing method, in this way, the mathematical model established is featured in wide adaptability and finally the weight vector $w = (w_1, w_2, ..., w_6)$ can be obtained.

3.3 CALCULATION OF DEGREE OF FITTING OF THE ORIGINAL REGRESSION EQUATION

Firstly these 6 evaluation indexes of the original regression equations are calculated, where the indexes belong to the smaller the priority type. Then the matrix of index membership grade R is adjusted, the optimal index set G can be obtained based on R and the membership grade of the original regression equations can be calculated through the compositional evaluation model of degree of fitting.

3.4 WEIGHTED ADJUSTMENT OF ORIGINAL REGRESSION EQUATIONS

Regard the membership grade calculated in Step 3 as the original weighted regression equation, and then the new equation can be obtained to establish a new optimal index set together with the original equation and then the membership grade of the corresponding equation can be obtained. The equation with the max. corresponding membership grade is the optimal fitting and prediction model under the given compositional evaluation model. Since the new optimal index set is comprehensively obtained based on the original regression equation and the new equation, the new optimal index set is surely not inferior to the original optimal index set; meanwhile, the original regression equation also participates in the optimization under the new optimal index set, in such a perfect condition, based on the given compositional evaluation weight, the new optimized equation is surely not inferior to the original regression equation, so the better fitting and prediction effect can be realized.

4 Case Representation

The detailed data about the sales volume and the prediction index of a certain material enterprise are shown in Table 1. According to the scatter diagram, suppose the relationship between y and x has the following four regression equation types to be determined:

$$\hat{y}_1 = \alpha + \beta x \tag{4.1}$$

$$\hat{y}_2 = \alpha e^{\beta x} \ (\beta > 0) \tag{4.2}$$

$$\hat{y}_3 = \alpha + \beta \ln x \tag{4.3}$$

$$\hat{y}_4 = \alpha x^{\beta} \tag{4.4}$$

Based on the regression analysis, the corresponding regression equation can be obtained as follows:

$$\hat{y}_1 = -2.341 + 0.41x \tag{4.5}$$

$$\hat{y}_2 = 4.712 e^{0.02076x} \tag{4.6}$$

$$\hat{y}_3 = -50.394 + 41.052 \ln x \tag{4.7}$$

$$\hat{y}_4 = 0.239 x^{1.12} \tag{4.8}$$

TABLE 1 Relationship between sales volume and the prediction index of a certain material enterprise

Year	Prediction index x	Measured value y	$\hat{y}_1$	$\hat{y}_2$	$\hat{y}_3$	$\hat{y}_4$	$\hat{y}$
1	99	39	37.1	37.9	35.4	34.8	36.9
2	45	12	16.7	12.9	16.1	21	15.1
3	102	33	36.2	36.3	34.6	34.4	35.8
4	85	29	31.8	28.8	30.3	32.1	30.1
5	91	33	36.6	37.1	35.0	34.6	36.3
6	109	57	43.6	53.6	41.8	37.8	46.8
7	29	10	7.3	7.8	7.7	8.1	7.6
8	19	8	4.1	6.6	5.0	0.3	5.3
9	52	16	16.7	12.9	16.1	21.0	15.1
10	39	10	12.6	10.4	12.4	16.4	11.7
11	46	11	14.7	11.6	14.2	18.8	13.4
12	18	5	3.3	6.3	4.3	-2.2	4.7
13	32	14	6.9	7.7	7.4	7.3	7.3
14	41	12	11.8	9.9	11.7	15.3	11.1
15	31	12	8.6	8.4	8.8	10.4	8.6
16	87	35	32.2	29.3	30.7	32.3	30.6
17	104	49	36.6	37.1	35.0	34.6	36.3

Thus 6 kinds of indexes and their membership grades of the regression equation can be obtained, as shown in Table 1

It can be easily found from the table that each type has its own recommendation, that is to say, at least one index in every type is superior to others, so every type participates in the weighting, with its membership grade as the weight, then the weighted estimated value:

$$\hat{y} = (0.83\hat{y}1 + 0.99\hat{y}2 + 0.83\hat{y}3 + 0.55\hat{y}4) / (0.83 + 0.99 + 0.83 + 0.55) \quad (4.9)$$

can be obtained, the optimal index set before weighting is (21.98, 11.07, 0.08, 3.52, 0.19, 1) and the one after weighting is (21.98, 11.07, 0.08, 0.26, 0.03, 1). Obviously, the new optimal index set is better; meanwhile, regardless of the original optimal index set or the new optimal index set, the weighted membership grade is larger than the original optimal scheme (exponent type), it can be obtained in the perfect condition that the compositional fitting precision of the new weighted scheme is higher and it is more advantageous for the prediction and control in the practice. Equations and the latter is obtained based on the optimal index set produced from the four original equations and its weighted equations, among which the first weighted membership grade is the so-called “membership grade” of the weighted index calculated based on the first optimal index set and its result is 2.69, exceeding 1, validating the fitting effect of the weighting equation again and indicating the completeness.

TABLE 2 Indexes and their membership grades of the equations

Type	Variance	Max. error	Min. error	Absolute value of average error	Average absolute relative error	Positive-negative rate of change	Membership grade
Linear type	26.5	11.33	0.27	3.99	0.2	1	0.83,0.66
Exponent type	21.98	11.07	0.08	3.52	0.21	1	0.99,0.82
Log type	45.92	14.25	0.1	5.75	0.42	5	0.83,0.66
Power type	26.59	12.95	0.21	3.8	0.19	1	0.55,0.43
Weighting type	22.21	11.69	0.35	0.26	0.03	1	2.69,0.91

Note: the former of the membership grade is obtained based on the optimal index set produced from the four original

5 Conclusions

This paper proposes one evaluation criterion based on the regression model, adopts the fuzzy evaluation method, regards the membership grade of the optimal index set of the given equation distance as the weight, linearly weights the corresponding equations and establishes the method of compositional regression model. The main conclusions of this paper are summarized as follows: 1. it adopts the thoughts of fuzzy evaluation and regards the membership grade living under the optimal scheme as the criterion of model evaluation, which effectively overcomes the dimension effects from the judgment of the regression model; 2. It regards the membership grade of the optimal index set of the given equation distance as the weight, weights the existing model and establishes the new compositional regression model, in this way, the fitting and prediction effects of the model are greatly improved; 3. The compositional optimal index set includes the sets of the original model specification and participates in the comprehensive judgment of the new index set with the membership grade, as a result, it guarantees the new optimal equation is surely not inferior to the original regression model specification and is a relatively perfect optimization method.

Acknowledgement

This research was supported by the National Science Foundation of China (No.71373246) the Fundamental Research Founds for National University, China University of Geosciences (Wuhan) (No.CUG120120).

References

[1] Shi G H, Han S F, Xu S N 2011 QFD Objective Level Decision-making Model Based on Fuzzy Non-linear Regression *Statistics and Decision* **16**: 47-49.

[2] Li Z Y, Liu W Y, Wang T J 2009 Fuzzy Bilinear Regression of Yield Series *Statistical Research* (2): 68-73.

[3] Li Z Y, Zhang C 2008 Structure Analysis of Regression Model with Fuzzy Data *Statistical Research* (8): 74-78.

[4] Liu S A, Zheng B L, Wang M G 1998 A Study on Improved Fuzzy Linear Regression and Solution *Prediction* (3):61-64.

[5] Chan K Y., Kwong C K, Law M C 2014. A fuzzy ordinary regression method for modeling customer preference in tea maker design. *Neurocomputing* **142**: 147-154.

[6] Liu Y, Zhou J, Chen Y 2014. Using fuzzy non-linear regression to identify the degree of compensation among customer requirements in QFD. *Neurocomputing* **142**: 115-124.

[7] Otadi M 2014 Fully fuzzy polynomial regression with fuzzy neural networks *Neurocomputing* **142**: 486-493.

[8] Zhang D, Agterberg F, Cheng Q, Zuo R 2014 Response to comment by Helmut Schaeben on "A Comparison of Modified Fuzzy Weights of Evidence, Fuzzy Weights of Evidence, and Logistic Regression for Mapping Mineral Prospectivity" *Mathematical Geosciences* **46**(7): 895-900.

Authors

Mengqi Gong, born in May. 1992, Honghu, Hubei, China

Current position, grades: post graduate of Statistics in China University of Geosciences.
University studies: China University of Geosciences.
Scientific interest: She is very interested in problems of econometrics, and has done quite a lot of researches on these kinds of issues.

Sheng Cheng, born in June 1974, Xishui, Hubei, China

Current position, grades: Associate professor of Statistic department in China University of Geosciences, Vice dean of Statistic department in China University of Geosciences.
Scientific interest: He is an expert in rural energy economy and econometrics, he has studies rural energy economy for decades, and has got plenty of achievements about that.
Experience: He is an associate professor in China university of Geosciences now, and he was economic counselor of the World Bank and Shangdong province.