Payments per claim model of outstanding claims reserve based on fuzzy linear regression

Yan, C., Liu, Q., Liu, J., Liu, W., Li, M. and Qi, M.

https://doi.org/10.1007/s40815-019-00617-x

Payments Per Claim Model of Outstanding Claims Reserve Based on Fuzzy Linear Regression

Chun Yan[1] · Qian Liu[1] · Jiahui Liu[1] · Wei Liu[2] · Meixuan Li[1] · Man Qi[3]

Abstract There are uncertainties in factors such as inflation. Historical data and variable values are ambiguous. They lead to ambiguity in the assessment of outstanding claims reserves. The payments per claim model can only perform point estimation. But the fuzzy linear regression is based on fuzzy theory and can directly deal with uncertainty in data. Therefore, this paper proposes a payments per claim model based on fuzzy linear regression. The linear regression method and fuzzy least square method are used to estimate the parameters of the fuzzy regression equation. And the estimated results are introduced into the payments per claim model. Then, the predicted value of each accident reserve is obtained. This result is compared with that of the traditional payments per claim model. And we find that the payments per claim model of estimating the fuzzy linear regression parameters based on the linear programming method is more effective. The model gives the width of the compensation amount for each accident year. In addition, this model solves the problem that the traditional payments per claim model cannot measure the dynamic changes in reserves.

Keywords Outstanding claims reserve · Payments per claim model · Fuzzy linear regression · Linear programming · Least square method

✉ Wei Liu
liuwei_doctor@yeah.net

1 College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, China

2 College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China

3 Department of Computing, Canterbury Christ Church University, Canterbury CT1 1QU, UK

1 Introduction

The extraction of accurate reserves can guarantee the insurance company's ability [1] to assume future liability or payment obligations. It also can protect the rights of policyholders and beneficiaries effectively [2]. The estimates of reserves are subject to many exogenous variables. Therefore, there exists uncertainties. In the deterministic method of assessing reserves, in general, the payments per claim model is superior to other methods, because it applies the report claims to average the claims. But it can only make point estimates and cannot measure the uncertainties of the reserves. However, fuzzy regression is based on fuzzy uncertainty. And it can directly process spatial information and measure uncertainty in data. So this paper introduces the fuzzy linear regression method into the payments per claim model to improve the estimation accuracy and help the company to provide more accurate reserves.

The fuzzy regression concept has been widely used in various fields. Liu et al. proposed a fuzzy regression model based on extreme learning machine, designed the training algorithm and analyzed the computational complexity. The simulation results show that the fuzzy system can effectively approximate the fuzzy input and output system [3]. Zuo et al. proposed the original fuzzy regression transfer learning method based on fuzzy rule. The results show that the proposed method has better performance than the existing model in solving the regression problem [4]. Baser et al. used the method of fuzzy regression function and support vector machine to estimate the annual average horizontal global solar radiation level [5]. Ren and Guo proposed the estimation method of the least squares order of fuzzy regression coefficients. They proved that this method is simpler. And the method avoids the problem that the ambiguity of the estimated value of the output variable becomes larger with the increase in the observed data [6]. Guo established an emergency material demand forecasting model based on multivariate fuzzy linear regression. It was found that the accuracy of emergency material demand forecasting can be improved by using symmetrical triangular fuzzy numbers to represent fuzzy attributes [7]. Xiang et al. used fuzzy linear regression model to evaluate the level of flood disaster events. They verified the feasibility of the parameter estimation method based on the probability mean-standard deviation distance fuzzy multivariate regression model [8]. Gong et al. proposed a fuzzy linear regression method based on preference least squares method in decision-making based on the continuous order weighted average (COWA) operator expectation distance measure. And they estimated model parameters and analyzed error in this way. By comparison, it is shown that the fitting effect of this method is better and feasible [9].

The payments per claim model is a more common model for determining the certainty of reserves. Its random model [10, 11] has been studied. Li et al. introduced outliers in the case of indemnity and proposed robust payments per claim model that can correct the selection of progress factors and settlement rate data. This method can effectively identify and adjust outliers, and can get more smooth final claim amount estimate [12]. Li and Meng studied its effects on the Poisson–Gamma and Tweedie models by analyzing the negative correlation between the frequency of claims and the claims. They empirically show that the Poisson–Gamma model is slightly more robust [13]. Yan and Zhang introduced the mathematical programming method in the payments per claim model, introduced the weighting factor and established the corresponding model. The results show that the method can reduce the influence of outliers on the progress factor and the settlement rate [14].

In practice, the assessment of outstanding claims reserves will be affected by factors such as inflation [15]. But these factors are uncertain. And historical data collection and statistics are generally not comprehensive and accurate due to the constraints of real conditions [16]. Therefore, the evaluation results are generally ambiguous. There is also a large amount of uncertainty. The uncertainty makes the future value of related variables often undetermined. The payments per claim model is a deterministic model. It can only make point estimation and cannot measure the uncertainty. So the error is relatively large if the payments per claim model is used to estimate reserves. But fuzzy linear regression is based on fuzzy theory [17] and can directly deal with uncertainty in data. Therefore, this paper introduces fuzzy linear regression in the payments per claim model and proposes a payments per claim model based on fuzzy linear regression. The model increases the accuracy of reserve forecasting. This paper uses fuzzy least square method and linear programming method to estimate the fuzzy linear regression parameters. Then, we compare the reserve estimation results with the reserve evaluation results obtained by the traditional payments per claim model. And it is found that the reserve estimation results obtained by the fuzzy linear regression method is close to the reserve prediction value obtained by the traditional payments per claim model. Its effect is better than that of the payments per claim model based on fuzzy least square method. It also verifies the feasibility of the random method. The payments per claim model based on the linear programming method also gives the width corresponding to the total amount of compensation in the current year for each accident year. And it measures the random dynamic changes of reserve uncertainty.

This paper is organized as follows. Sect. 2 introduces the traditional payments per claim model. In Sect. 3, the principle of fuzzy linear regression is introduced. Sect. 4 proposes the payments per claim model based on fuzzy linear regression. An empirical analysis is given in Sect. 5. The conclusions are given in Sect. 6.

2 The Traditional Payments Per Claim Model

In an unstable economic situation, for example, when the inflation rate is high and fluctuate violently, if only the chain ladder method is used, the accurate and effective reserves cannot be obtained [18]. However, the payments per claim model can better solve this problem. So this article is based on the payments per claim model. According to the number of paid claims and the number of reported claims, the payments per claim model can be divided into two models. One is the reported payments per claim model based on reported claims data and the number of reported claims. The other is the closed payments per claim model based on paid claims data and paid claims. The closed payments per claim model involved in this paper is introduced as follows.

2.1 The Payments Per Claim Finalized Model

The following traffic triangle data are known to be used for the reserve estimate using the Payments Per Claim Finalized (PPCF) model. $N_{i,j}$ is the run-off triangle that accumulated the number of reported cases, $D_{i,j}$ is the run-off triangle that accumulated the number of closed cases, and $P_{i,j}$ is the run-off triangle that described the incremental closed claims.

To obtain the original data, first use the chain ladder method to predict the lower triangular portion of the run-off triangle that is called the cumulative number of reported cases. Second, the number of closed cases is calculated. The final number of closed cases is equal to the difference between the cumulative number of reported cases and the number of outstanding cases. And then, the settlement rate can be calculated based on the cumulative number of closed cases and the cumulative number of reported cases.

$$V_{i,j} = \frac{D_{i,j}}{N_{i,j}} \tag{1}$$

Next the average of the settlement rate is selected to predict the lower triangular portion of the run-off triangle of the cumulative number of closed cases. Then, the triangles of the indemnity flow of the closed cases in each accident year are found. And the chain ladder method is used to predict the lower triangular part. Finally, the future outstanding claims reserve is predicted by the closed claims. The final outstanding claims are the sum of the data in the lower triangle of the run-off triangle for the reserve forecast.

3 Fuzzy Linear Regression

Fuzzy regression is based on the possibility. It can get the interval estimate while performing the point estimation, and deal with the uncertainty in the data. Fuzzy regression uses fuzzy functions to represent coefficients. It treats the deviation between the observed value and the estimated value as the ambiguity of the system itself. It is assumed that the value of the estimated value has a range, i.e., this estimate value can take any value in the range. The effective width is called width. Fuzzy regression is divided into fuzzy linear regression and fuzzy nonlinear regression. The goal of fuzzy linear regression is to estimate the fuzzy parameters of the regression model.

Fuzzy variables y_i are assumed to be functions about fuzzy independent variables $x_1, x_2, \ldots, x_n$. $f: R^n \to E$, $y_i = f(x_{i1}, x_{i2}, \ldots, x_{in})$, where i is the number of the observation and n is the number of independent variables. Then the fuzzy linear regression can be expressed as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \mu_i \tag{2}$$

Fuzzy nonlinear regression is a parameter for estimating nonlinear models. The specific formula is shown in formula (3).

$$y_i = \beta_0 e^{\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in}} + \mu_i \tag{3}$$

The fuzzy nonlinear regression model can be transformed into fuzzy linear regression by using the logarithm method based on formula (4).

$$\ln y_i = \ln(\beta_0) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \ln \mu_i \tag{4}$$

The above parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are all fuzzy numbers.

3.1 Fuzzy Numbers

Fuzzy numbers are the basis for fuzzy data analysis. The set of real numbers is called R. The fuzzy set $\widetilde{A}$ on the real number field is called the fuzzy number. And $\widetilde{A}$ has a membership function, $\mu: R \to [0, 1]$. According to the above definition, the symmetrical triangular fuzzy number is recorded as $\widetilde{A} = (c, \omega)_L$, where c is the center point and ω represents the width. It conforms to the following operation rules.

$$(c_1, \omega_1)_L + (c_2, \omega_2)_L = (c_1 + c_2, \omega_1 + \omega_2)_L \tag{5}$$

$$\lambda(c, \omega)_L = (\lambda c, |\lambda| \omega)_L \tag{6}$$

According to the defined triangular fuzzy number and the operation of symmetrical fuzzy number, the fuzzy linear regression model can be expressed as

$$\begin{aligned} &\widetilde{Y}(x_p) \\ &\quad = \widetilde{A_0} + \widetilde{A_1} X_{p1} + \cdots + \widetilde{A_n} X_{pn} = (c(x_p), \omega(x_p))_L \end{aligned} \tag{7}$$

where

$$\begin{aligned} c(x_p) &= c_0 + c_1 x_{p1} + \cdots + c_n x_{pn} \\ \omega(x_p) &= \omega_0 + \omega_1 |x_{p1}| + \cdots + \omega_n |x_{pn}| \end{aligned}$$

3.2 Parameter Estimation Method of Fuzzy Linear Regression Model

There are two main methods for estimating fuzzy regression parameters in fuzzy linear regression models. One is linear programming, and the other is fuzzy least square. The principles are described below.

3.2.1 Principle of Linear Programming

The linear programming method proposed by Tanaka [19] is based on the fuzzy set theory proposed by Zadeh [20]. The regression problem of fuzzy data is studied. The method of fuzzy parameters in fuzzy regression is obtained by linear programming problem. It is believed that the regression coefficient in fuzzy regression is ambiguous and can be represented by the number of intervals with membership grade. And the goal of this linear programming is to minimize the ambiguity of the system.

For example, historical statistical data groups are

$(x_{1,1}, x_{1,2}, \ldots, x_{1,n}, y_1)^{\mathrm{T}}$
$(x_{2,1}, x_{2,2}, \ldots, x_{2,n}, y_2)^{\mathrm{T}}$
$\ldots$
$(x_{m,1}, x_{m,2}, \ldots, x_{m,n}, y_m)^{\mathrm{T}}$

where m represents the number of observation data sets, n represents the number of independent variables, $x_{i,j}(i = 1, 2, \ldots, m)$ represent historical observation data of the jth independent variable, and y_i represent the historical observation data of the dependent variable. Then the linear programming problem of the fuzzy regression coefficient of the above fuzzy regression model can be expressed as

$$\min\left(m\omega_0 + \sum_{j=1}^{n}\sum_{i=1}^{m}\omega_j x_{ij}\right)$$

The constraints are

$$\begin{cases} y_i \leq \left(c_0 + \sum_{j=1}^{n} c_j x_{ij}\right) + (1-h)\left(\omega_0 + \sum_{j=1}^{n}\omega_j |x_{ij}|\right) \\ y_i \geq \left(c_0 + \sum_{j=1}^{n} c_j x_{ij}\right) - (1-h)\left(\omega_0 + \sum_{j=1}^{n}\omega_j |x_{ij}|\right) \\ (\omega_j \geq 0 \quad i = 1, 2, \ldots, m \quad j = 0, 1, 2, \ldots, n) \end{cases} \tag{8}$$

If $h = 0.5, c_j(j = 1, 2, \ldots, n)$ are the coefficients that can be obtained by the conventional multiple linear regression models. The linear programming problem of the above formula is solved. And c_j, $\omega_j(j = 1, 2, \ldots, n)$ are found. And then $\widetilde{A} = (c, \omega)_L$ is calculated. $\widetilde{A} = (c, \omega)_L$ is a fuzzy regression coefficient of the established fuzzy regression model.

3.2.2 Principle of Fuzzy Least Square Method Based on Euclidean Distance

The concept of fuzzy least square method was proposed in 1988. And it is similar to the traditional least square method. The first step is to define the distance between two fuzzy numbers [21]. Then, the regression coefficient problem is transformed into a problem that minimizes the sum of squared deviations [22]. According to the definition of distance, the current least square method can be divided into least square methods based on three different distances. The first one is the Euclidean distance. The second one is $Y - K$ distance. And the third one is D_k distance. The literature [23] studied these three least square methods and error terms. It can be seen that the least square estimates of the regression coefficients obtained by the three methods are the same. But the error term of the least square method based on the Euclidean distance is smaller than the other two methods. In view of this, the least square method based on Euclidean distance is used to estimate the fuzzy regression coefficients in the fuzzy regression model.

We assume that the fuzzy number $\tilde{A} = (a_m, a_l, a_r)_L$ and $\tilde{B} = (b_m, b_l, b_r)_L$. Then, the Euclidean distance between two fuzzy numbers is defined as

$$D_E^2(\tilde{A}, \tilde{B}) = (a_m - b_m)^2\omega_m + (a_l - b_l)^2\omega_l + (a_r - b_r)^2\omega_r \tag{9}$$

where $\omega_m > 0$, $\omega_l > 0$, $\omega_r > 0$, and they are arbitrary weights.

For fuzzy regression models

$$\tilde{Y}(x_i) = \tilde{A}_0 + \tilde{A}_1 X_{i1} + \cdots + \tilde{A}_p X_{ip}, i = 1, 2, \ldots, n \tag{10}$$

where X_{ip} are the determined real numbers, $\tilde{Y}(x_i)$ and $\tilde{A}_m(m = 1, 2, \ldots, p)$ are triangular fuzzy numbers. For the convenience of calculation, we let $\tilde{Y}(x_i)$ and $\tilde{A}_m$ be symmetrical triangular fuzzy numbers with the same membership function. Then, the formula (10) can be expressed as

$$(c_i, s_i) = (a_0, r_0) + (a_1, r_1)x_{i1} + \cdots + (a_p, r_p)x_{ip} \tag{11}$$

The sum of squared errors can be expressed as

$$\begin{aligned} D_E^2 = \sum^{n} &[(c_i - (a_0 + a_1 x_{i1} + \cdots + a_p x_{ip}))^2 \\ &+ (s_i - (r_0 + r_1 x_{i1} + \cdots + r_p x_{ip}))^2] \end{aligned} \tag{12}$$

where $a = (a_0, a_1, \ldots, a_p)'$, $r = (r_0, r_1, \ldots, r_p)'$, $c = (c_0, c_1, \ldots, c_n)'$, $s = (s_0, s_1, \ldots, s_n)'$,

$$X = \begin{Bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{Bmatrix} \tag{13}$$

Then, the sum of squared errors is expressed as $D_E^2 = (Xa - c)'(Xa - c) + (Xr - s)'(Xr - s)$. The partial derivatives of the sum of the squared errors are found. And these partial derivatives are the partial derivatives with respect to the vectors $a = (a_0, a_1, \cdots, a_p)'$ and $r = (r_0, r_1, \cdots, r_p)'$. The partial derivatives are set equal to 0. Then, the fuzzy regression coefficients are gotten.

$$\begin{cases} \hat{a} = (X'X)^{-1}X'c \\ \hat{r} = (X'X)^{-1}X's \end{cases}$$

4 Payments Per Claim Model Based on Fuzzy Linear Regression

The fuzzy linear regression model is combined with the payments per claim model. And we propose the payments per claim model based on fuzzy linear regression. The principles and calculation steps of the two models are introduced below. One model is the payments per claim model based on the linear programming method. The other model is the payments per claim model based on the least square method.

4.1 Estimating Fuzzy Regression Parameters Using Linear Programming

The fuzzy regression model can be used to find the outstanding claims reserve based on the average claims data. The steps are as follows.

(1) Calculate the number of paid claims based on the accumulated paid amount and the cumulative paid indemnities.

$$S_{i,j} = \frac{Z_{i,j}}{N_i}, i = 1, 2, \ldots, n \tag{14}$$

where $Z_{i,j}$ is the total amount of claim for the year i as at development year j, N_i is the number of cases in which the accident year is i, and $S_{i,j}$ is the average amount of a claim for the year i as at development year j.

(2) Take the logarithms of $S_{i,j}$, λ_{i+j},and γ. And then a fuzzy linear regression model for $\ln S_{i,j}$ and regression coefficients $\ln \hat{\gamma}_j$, $\ln \hat{\lambda}_{i+j}$ is established. The model can be expressed as

$$\ln S_{i,j} = \ln \hat{\lambda}_{i+j} + \ln \hat{\gamma}_j + \varepsilon_{i,j}, i + j \leq n \tag{15}$$

where n represents the number of independent variables, $\lambda_{i,j}$ can be interpreted as exogenous variables such as inflation, and γ_j is interpreted as the final number of claims in the development year j.

The model as a matrix is expressed as

$$Y = X\beta + \varepsilon \tag{16}$$

where $Y = (\ln S_{0,0}, \ldots, \ln S_{0,n}, \ln S_{1,0}, \ldots, \ln S_{1,n-1}, \ldots, \ln S_{n,0})^{\mathrm{T}}$, and Y is a vector containing $[(n+1)(n+2)/2]$ components. $\beta = (\ln \gamma_0, \ln \gamma_1, \ldots, \ln \gamma_n, \ln \lambda_0, \ldots, \ln \lambda_n)^{\mathrm{T}}$, and β is composed of $2n+2$ unknown parameters. $\varepsilon = (\varepsilon_{0,0}, \ldots, \varepsilon_{0,n}, \varepsilon_{1,0}, \ldots, \varepsilon_{1,n-1}, \ldots, \varepsilon_{n,0})^{\mathrm{T}}$, and ε is the corresponding error vector.

$$X = \begin{bmatrix} I_{(n+1,n+1)} & \cdots & \cdots & I_{(n+1,n+1)} \\ I_{(n,n)} & 0_{(n,1)} & 0_{(n,1)} & I_{(n,n)} \\ \vdots & \vdots & \vdots & \vdots \\ I_{(n-i+1,n-i+1)} & 0_{(n-i+1,i)} & 0_{(n-i+1,i)} & I_{(n-i+1,n-i+1)} \\ \vdots & \vdots & \vdots & \vdots \\ I_{(1,1)} & 0_{(1,n)} & 0_{(1,n)} & I_{(1,1)} \end{bmatrix} \tag{17}$$

Matrix X is the matrix of order $[(n+1)(n+2)/2] * (2n+2)$ matrix, where $I_{(i,j)}$ is the identity matrix of order $i * j$ and $0_{(i,j)}$ is the zero matrix of order $i * j$. Then, the fuzzy regression parameters can be gotten based on optimal solution obtained by the linear programming method.

(3) Blur the regression coefficient. We need to assume that the error is normal [24] and has an independent distribution [25], and then we can write the fuzzy range of the estimated coefficient by the obtained error range.

(4) Estimate the average number of indemnities in the lower triangular part of the logarithmic run-off triangle.

$$\begin{aligned} \ln \hat{\tilde{S}}_{i,j} &= \left(\ln \hat{S}_{i,j}, c_{\ln \hat{S}_{i,j}}\right) \\ &= \ln \hat{\tilde{\gamma}}_j + \ln \hat{\tilde{\lambda}}_{i+j} \\ &= \left(\ln \hat{\gamma}_j, c_{\ln \hat{\gamma}_j}\right) + \left(\ln \hat{\lambda}_{i,j}, c_{\ln \hat{\lambda}_{i+j}}\right) \end{aligned} \tag{18}$$

(5) According to the lower triangle obtained in step (4), the lower triangular part of the run-off triangle of the case is found. If the weighting function of the fuzzy number is used, the future claim for each accident year can be evaluated as a clear real number. The following is used to calculate the average amount of compensation.

$$\hat{S}_{i,j} = e^{\ln \hat{S}_{i,j}} \left(\frac{e^{c_{\ln \hat{S}_{i,j}}} + e^{-c_{\ln \hat{S}_{i,j}}} - 2}{c^2_{\ln \hat{S}_{i,j}}} \right) \tag{19}$$

(6) Calculate the accumulated outstanding claims reserve for each progress year based on the predicted number of claims.

$$\hat{Z}_{i,j} = \hat{S}_{i,j} N_i \tag{20}$$

(7) Finally, the final compensation reserve is obtained as

$$\hat{R} = \sum_{j=n-i+1}^{n} \hat{Z}_{i,j}, i = 1, 2, \ldots, n \tag{21}$$

4.2 Estimating Fuzzy Regression Parameters by Least Squares Method

R software is used to solve the regression coefficient point estimate β of the regression model and substitute the fuzzy regression coefficient into the fuzzy regression linear model. Then, the weight function of the fuzzy number is used to obtain the predicted value of the average amount of compensation for each accident year in the current year. And a 95% confidence interval for the fuzzy regression coefficient is constructed.

4.2.1 Solving Confidence Intervals of Fuzzy Regression Coefficients

The significance level $\alpha = 0.05$. The definition of the membership function of the fuzzy regression parameter can be determined by the confidence interval. Since the fuzzy regression parameter is a fuzzy number, we need to construct a confidence interval for the fuzzy function. First, we must assume that the error is subject to a normal distribution. Then, both $\frac{\ln\hat{\gamma}_j - \ln\gamma_j}{\hat{\sigma}_{\ln\hat{\gamma}_j}}$ and $\frac{\ln\hat{\lambda}_{i+j} - \ln\lambda_{i+j}}{\hat{\sigma}_{\ln\hat{\lambda}_{i+j}}}$ obey the $t(n-2)$ distribution, so the $(1-\alpha)100\%$ confidence intervals for the regression parameters $\ln\gamma_j$ and $\ln\lambda_{i+j}$ are shown in formula (22) and formula (23).

$$P\left(\ln\hat{\gamma}_j - t_{\alpha/2}(n-2)\hat{\sigma}_{\ln\hat{\gamma}_j} \le \ln\hat{\gamma}_j + t_{\alpha/2}(n-2)\hat{\sigma}_{\ln\hat{\gamma}_j}\right) = 1-\alpha \tag{22}$$

$$P\left(\ln\hat{\lambda}_{i+j} - t_{\alpha/2}(n-2)\hat{\sigma}_{\ln\hat{\lambda}_{i+j}} \le \ln\hat{\lambda}_{i+j} + t_{\alpha/2}(n-2)\hat{\sigma}_{\ln\hat{\lambda}_{i+j}}\right) = 1-\alpha \tag{23}$$

where $1-\alpha$ is the confidence level, t is the value of the statistic, $\hat{\sigma}_{\ln\hat{\gamma}_i}$ is the standard deviation of $\ln\gamma_j$, $\hat{\sigma}_{\ln\hat{\lambda}_{i+j}}$ is the standard deviation of $\ln\hat{\lambda}_{i+j}$, and $t_{\alpha/2}(n-2)$ is the value of the student statistic.

Through the above two formulas, we can establish the triangular fuzzy number parameter $\ln\gamma_j$ and the parameter $\ln\lambda_{i+j}$. The estimated confidence intervals for parameters $\ln\gamma_j$ and $\ln\lambda_{i+j}$ allow us to reflect the standard deviation contained in the data set to the fuzzy prediction of the reserve. And the fuzzy regression parameter is a symmetrical triangular fuzzy number, which can be expressed as $\ln\hat{\tilde{\gamma}}_j = \left(\ln\hat{\gamma}_j, c_{\ln\hat{\gamma}_j}\right)$ and $\ln\hat{\tilde{\lambda}}_{i,j} = \left(\ln\hat{\lambda}_{i,j}, c_{\ln\hat{\lambda}_{i+j}}\right)$. Where $\ln\hat{\lambda}_{i+j}$ and $\ln\hat{\gamma}_j$ are the centers of the fuzzy regression coefficients, and $c_{\ln\hat{\gamma}_j}$ and $c_{\ln\hat{\lambda}_{i+j}}$ are the spreads of the error range. In view of this, the range of the extended parameter of the fuzzy regression parameter is determined as

$$c_{\ln\hat{\gamma}_j} = t_{\alpha/2}(n-2)\cdot\hat{\sigma}_{\ln\hat{\gamma}_j} \tag{24}$$

$$c_{\ln\hat{\lambda}_{i+j}} = t_{\alpha/2}(n-2)\cdot\hat{\sigma}_{\ln\hat{\lambda}_{i+j}} \tag{25}$$

4.2.2 Fuzzy Time Trend of External Influence

In order to predict the cost of an unknown reserve claim, we need to predict the value of $\ln\hat{\tilde{\lambda}}_{i,j}, (i+j = n+1,\ldots,2n)$. Therefore, we adjust the time trend of external influences. We can use the linear function of the period and then estimate it with mixed fuzzy least squares regression $\ln\hat{\tilde{\lambda}}_{i,j}, (i+j = n+1,\ldots,2n)$. So the fuzzy regression model can be expressed as

$$\ln\hat{\tilde{\lambda}}_{i,j} = \tilde{A} + \tilde{B}(i+j), i+j \le n \tag{26}$$

where $\tilde{A} = (a, c_a)$, $\tilde{B} = (b, c_b)$.

Therefore, each predicted value can be expressed as a symmetrical triangular fuzzy number.

$$\ln\hat{\tilde{\lambda}}_{i,j} = \left(\ln\hat{\lambda}_{i,j}, c_{\ln\hat{\lambda}_{i+j}}\right) = (a, c_a) + (b, c_b)(i+j), \quad i+j \le n \tag{27}$$

$\tilde{A} = (a, c_a)$ and $\tilde{B} = (b, c_b)$ are calculated based on the following two formulas.

$$\begin{cases} a(n+1) + b\sum(i+j) = \sum\ln\hat{\lambda}_{i+j} \\ a\sum(i+j) + b\sum(i+j)^2 = \sum(i+j)\ln\hat{\lambda}_{i+j} \end{cases} \tag{28}$$

$$\begin{cases} c_a(n+1) + c_b\sum(i+j) = \sum c_{\ln\hat{\lambda}_{i+j}} \\ c_a\sum(i+j) + c_b\sum(i+j)^2 = \sum(i+j)c_{\ln\hat{\lambda}_{i+j}} \end{cases} \tag{29}$$

The remaining steps are the same as the linear programming method for solving the fuzzy regression parameters, as shown in formulas (18)–(21).

5 Empirical Analysis

The following is an empirical analysis of the payments per claim finalized model based on fuzzy linear regression through a set of classic data. The raw data are from "Non-Life Insurance Actuarial" [26].

The cumulative run-off triangle N_{ij} of the number of closed cases is shown in Table 1.

Table 1 Cumulative closed case flow table

Year of accident	Development progress				
	0	1	2	3	4
2004	275	375	426	466	479
2005	300	408	460	499	
2006	326	440	500		
2007	340	464			
2008	350				

The run-off triangle of the settlement triangle $P_{i,j}$ of the indemnity flow is shown in Table 2.

The run-off triangle of the settled traffic run-off triangle is shown in Table 3.

5.1 Introducing the Fuzzy Average Regression Method to Estimate the Outstanding Claims Reserve

R software is used to calculate the logarithm of the payments per claims, and the logarithm is shown in Table 4.

A fuzzy regression model is established by using the logarithm of the claimed indemnity.

$$Y = X\beta + \varepsilon$$

$$Y = \left(\ln S_{0,0}, \ldots, \ln S_{0,4}, \ln S_{1,0}, \ldots, \ln S_{1,3}, \ldots, \ln S_{4,0}\right)^{\mathrm{T}} = (y_1, \ldots, y_{15})^{\mathrm{T}}$$

i.e.,

$$Y = (1.293, 1.15879, 1.7342, 1.8618, 1.9411, 1.3173, 1.6446, 1.7915, 1.9189, 1.3638, 1.7060, 1.8672, 1.4769, 1.820444, 1.5980)^{\mathrm{T}}$$

5.1.1 *Estimating Fuzzy Regression Parameters of Fuzzy Regression Models Using Linear Programming*

According to the introduction of the linear programming method, the constant term value is zero. And the parameter solution of the established fuzzy regression model can be transformed into the following formula.

$$min \sum_{j=1}^{10} \sum_{i=1}^{15} \omega_j x_{ij}$$

Its linear constraints are

$$\begin{cases} y_i \le \sum_{j=1}^{10} c_j x_{ij} + 0.5\left(\omega_0 + \sum_{j=1}^{10} \omega_j x_{ij}\right) \\ y_i \ge \sum_{j=1}^{10} c_j x_{ij} - 0.5\left(\omega_0 + \sum_{j=1}^{10} \omega_j x_{ij}\right) \\ \left(\omega_j \ge 0 \quad i = 1, 2, \ldots, 15 \quad j = 0, 1, 2, \ldots, 10\right) \end{cases}$$

R software is used to solve the linear programming. The optimal solution is as follows when the system ambiguity $min \sum_{j=1}^{10} \sum_{i=1}^{15} \omega_j x_{ij}$ takes the minimum value of 1.4668.

$$\omega = (\omega_1, \ldots, \omega_{10})^{\mathrm{T}} = (0.0066, 0.0020, 0.0448, 0.0574, 0, 0, 0.0504, 0.0498, 0, 0)^{\mathrm{T}}$$

$$c = (c_1, \ldots, c_{10})^{\mathrm{T}} = (1.2963, 1.5231, 1.5912, 1.6492, 1.6428, 0, 0.0495, 0.0957, 0.1839, 0.2984)^{\mathrm{T}}$$

The resulting fuzzy regression coefficient $\widetilde{A} = (c, \omega)_L$ is

Table 2 Closed indemnity flow statement

Year of accident	Development progress				
	0	1	2	3	4
2004	1003	1855	2413	2999	3337
2005	1120	2113	2776	3400	
2006	1275	2423	3235		
2007	1489	2865			
2008	1730				

Table 3 Closed case indemnity run-off triangle

Year of accident	Development progress				
	0	1	2	3	4
2004	3.6473	4.9467	5.6643	6.4356	6.9666
2005	3.7333	5.1789	6.0348	6.8136	
2006	3.911	5.5068	6.47		
2007	4.3794	6.1746			
2008	4.9429				

Table 4 Closed account logarithm run-off triangle

Year of accident	Development progress				
	0	1	2	3	4
2004	1.2930	1.5987	1.7342	1.8618	1.9411
2005	1.3173	1.6446	1.7975	1.9189	
2006	1.3638	1.7060	1.8672		
2007	1.4769	1.8204			
2008	1.5980				

$\widetilde{A}_1 = (c_1, \omega_1)_L = (1.2963, 0.0066)$
$\widetilde{A}_2 = (c_2, \omega_2)_L = (1.5231, 0.0020)$
$\widetilde{A}_3 = (c_3, \omega_3)_L = (1.5912, 0.0448)$
$\widetilde{A}_4 = (c_4, \omega_4)_L = (1.6492, 0.0574)$
$\widetilde{A}_5 = (c_5, \omega_5)_L = (1.6428, 0)$
$\widetilde{A}_6 = (c_6, \omega_6)_L = (0, 0)$
$\widetilde{A}_7 = (c_7, \omega_7)_L = (0.0495, 0.0503)$
$\widetilde{A}_8 = (c_8, \omega_8)_L = (0.0957, 0.0498)$
$\widetilde{A}_9 = (c_9, \omega_9)_L = (0.1839, 0)$
$\widetilde{A}_{10} = (c_{10}, \omega_{10})_L = (0.2984, 0)$

Then, the regression model is

$$Y = (1.2963, 0.0066)X_1 + \cdots + (0.2984, 0)X_{10}$$

According to the algorithm of fuzzy numbers and the number of claims in the current year, the total amount $Z_{i,j}$ of compensation for each accident year in the current year can be calculated as 2359.8721, 3084.1598, 308.3260, 3253.5355 and 2438.2922. The corresponding widths are 350.0000, 491.4048, 502.2465, 501.2420 and 482.1624.

5.1.2 *Estimating Fuzzy Regression Parameters of Fuzzy Regression Linear Model by Least Squares Method*

R software is used to solve regression coefficients of statistical regression model, and the point estimation of regression coefficients is

$$\beta = (0.2535, 0.3248, 0.3721, 0.3776, 0.0383, 0.0951, 0.1800, 0.2706)$$

The corresponding widths are 0.045, 0.0505, 0.0587, 0.0774, 0.0814, 0.0785, 0.0774 and 0.0774. The fuzzy regression coefficient is introduced into the fuzzy regression linear model, and then the weighting function of the fuzzy number is used. In this way, we can obtain the predicted value of the average amount of compensation for each accident year in the current year, as shown in Table 5.

Table 5 Case average compensation number forecast value traffic triangle

Year of accident	Development progress				
	0	1	2	3	4
2004					4.78374
2005				6.1712	
2006			6.6283		
2007		6.9517			
2008	6.9944				

The sum of the claims for the case is multiplied to the number of claims in the current year for each year of the accident. And the total amount of compensation for the current year of each accident year can be calculated. These are 2291.3894 thousand yuan, 3079.4338 thousand yuan, 3314.1635 thousand yuan, 3225.6106 thousand yuan and 2448.0530 thousand yuan.

5.2 Result Analysis

According to the previous calculation, we can get three sets of data. The first are estimated results of claim reserves obtained under the traditional payments per claim model. The second are estimated claim reserves for fuzzy linear regression models based on linear programming. And the third are estimated claim reserves based on fuzzy linear regression model obtained by fuzzy least squares. The three sets of data are compared, and the results are shown in Table 6.

Table 6 shows that there is an error between the estimated value of the claim reserve requested by the two methods and the forecast value of the traditional case compensation method reserve, but the two values are very close. It verifies the feasibility of a random method. In addition, the payments per claim model based on the linear programming method to estimate the fuzzy linear regression coefficient gives the width corresponding to the total amount of compensation in the current year for each accident year. They are 350.0000, 491.4048, 502.2465, 501.2420, 482.1624. They measure the dynamic changes in the outstanding claims reserve under the influence of various factors.

According to the reserve evaluation results in Table 6, the absolute errors corresponding to the two random methods are obtained. Specifically, as shown in Table 7.

According to Table 7, the absolute error of the payments per claim model based on linear programming method is smaller than that of the payments per claim model based on fuzzy least square method. It can be seen that the effect of using the linear programming method to solve the parameters and evaluate the outstanding claims reserve in this example is slightly better than the effect of using the fuzzy least square method.

6 Summary

Based on the data of the average indemnity, this paper establishes an evaluation model of the payments per claim reserve based on fuzzy linear regression. The fuzzy regression coefficients of the model are estimated by linear

Table 6 Reserve evaluation result comparison table

Year of accident	Reserve evaluation results			
	Traditional payments per claim method	Linear programming method	The width of the linear programming method reserve	Fuzzy least squares
0	3336.9810	2359.8721	350.0000	2291.3894
1	3424.2586	3084.1598	491.4048	3079.4338
2	3191.6489	3308.3260	502.2465	3314.1635
3	2816.7948	3253.5355	501.2420	3225.6106
4	1730.0150	2438.2922	482.1624	2448.0530

Table 7 Absolute error comparison table corresponding to two evaluation methods based on fuzzy linear regression model

Year of accident	Absolute error corresponding to two evaluation methods based on fuzzy linear regression model	
	Absolute error of linear programming method	Absolute error of fuzzy least square method
0	977.1089	1045.5916
1	340.0988	344.8248
2	− 116.6771	− 122.5146
3	− 436.7407	− 408.8158
4	− 708.2772	− 718.038

programming method and fuzzy least square method respectively. Then, the fuzzy regression coefficients obtained by the two methods are introduced back into the model and applied to the payments per claim model. Next, the estimated reserve values for each accident year are found. At last, the results of reserve estimation are compared with that of traditional the payments per claim model. The results show that the payments per claim model of estimating the fuzzy linear regression coefficient based on the linear programming method is better and more feasible. The width corresponding to the total compensation amount of compensation in the current year for each accident year achieves a measure of the dynamic changes in uncertainty of outstanding claims reserves. It also helps insurance companies use the payments per claim model to extract more accurate reserves.

Acknowledgements This work was financially supported by the Project of National Natural Science Foundation of China(61502280,61472228).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Yan, C., Wang, L., Liu, W., Qi, M.: Financial early warning of non-life insurance company based on RBF neural network optimized by genetic algorithm. Concurr. Comput. Pract. Exp. **30**(23), 1–11 (2017)
2. Yan, C., Sun, H.T., Liu, W., Chen, J.: An integrated method based on hesitant fuzzy theory and RFM model to insurance customers segmentation and lifetime value determination. J. Intell. Fuzzy Syst. **35**(1), 159–169 (2018)
3. Liu, H.T., Wang, J., He, Y.L., et al.: Extreme learning machine with fuzzy input and fuzzy output for fuzzy regression. Neural Comput. Appl. **28**(11), 3465–3476 (2017)
4. Zuo, H., Zhang, G., Pedrycz, W., et al.: Fuzzy regression transfer learning in Takagi–Sugeno fuzzy models. IEEE Trans. Fuzzy Syst. **25**(6), 1795–1807 (2017)
5. Baser, F., Demirhan, H.: A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. Energy **123**, 229–240 (2017)
6. Ren, Y., Guo, S.Z.: Fuzzy linear regression based on structural element least squares order. Fuzzy Syst. Math. **29**(01), 126–133 (2015)
7. Guo, Z.X., Han, R., Qi, M.R.: Emergency material demand forecasting model based on multiple fuzzy regression. J. Hebei Univ. (Nat. Sci. Edit.) **37**(04), 337–342 (2017)
8. Gong, Y.B., Xiang, L., Liu, G.F.: Research on flood disaster classification evaluation method based on fuzzy regression model. Stat. Inf. Forum **2018**(5), 65–74 (2018)
9. Gong, Y.B., Dai, L.L., Hu, N.: Parameter estimation of fuzzy linear regression model based on COWA operator expectation. Stat. Decis. **2018**(2), 15–18 (2018)
10. Xiao, G.Q., Li, K.L., Li, K.Q.: Reporting l most influential objects in uncertain databases based on probabilistic reverse top-k queries. Inf. Sci. **405**, 207–226 (2017)
11. Xiao, G.Q., Li, K.L., Li, K.Q., Zhou, X.: Efficient top-(k, l) Range query processing for uncertain data based on multicore architectures. Distrib. Parallel Databases **33**(3), 381–483 (2015)
12. Yan, C., Li, Y.X., Sun, X.H., Liu, Z.B.: Non-life insurance reserve case for indemnity. Insur. Stud. **2015**(11), 15–24 (2015)
13. Li, Z.X., Meng, S.W.: Automobile insurance pricing model under dependent risk conditions. Insur. Stud. **2016**(7), 68–77 (2016)
14. Yan, C., Zhang, J., Su, R.: Case-based compensation method based on mathematical programming method considering outliers. Theory Pract. Finance Econ. **38**(3), 46–51 (2017)

15. Yan, C., Sun, H.T., Liu, W.: Study of fuzzy association rules and cross-selling toward property insurance customers based on FARMA. J. Intell. Fuzzy Syst. **31**(06), 2789–2794 (2016)
16. Xiao, G.Q., Li, K.L., Zhou, X., Li, K.Q.: Efficient monochromatic and bichromatic probabilistic reverse top-k query processing for uncertain big data. J. Comput. Syst. Sci. **89**, 92–113 (2017)
17. Xiao, G.Q., Wu, F., Zhou, X., Li, K.Q.: Probabilistic top-k range query processing for uncertain databases. J. Intell. Fuzzy Syst. **31**(2), 1109–1120 (2016)
18. Li, Y.Q., Yan, C., Liu, W., Li, M.Z.: A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. Appl. Soft Comput. **70**, 1000–1009 (2018)
19. Tanaka, H., Uejima, S., Asai, K.: A liner regression analysis with fuzzy functions. IEEE Trans. Syst. Man Cybern. **25**(2), 162–174 (1982)
20. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
21. Chen, J.G., Li, K.L., Tang, Z., Yu, S., Li, K.Q.: A parallel random forest algorithm for big data in Spark cloud computing environment. IEEE Trans. Parallel Distrib. Syst. **28**(4), 919–933 (2017)
22. Zhou X., Li K.L., Yang Z.B., Xiao G.Q., Li K.Q.: Progressive approaches for Pareto optimal groups computation. IEEE Trans. Knowl. Data Eng. (in press) https://doi.org/10.1109/TKDE.2018.2837117 (2018)
23. Zhang, A.W.: Comparison of fuzzy linear least squares regression based on different distances. J. Jiangsu Univ. Sci. Technol. (Nat. Sci. Edit.) **26**(5), 509–513 (2012)
24. Li, K.L., Yang, W.D., Li, K.Q.: Performance analysis and optimization for SpMV on GPU using probabilistic modeling. IEEE Trans. Parallel Distrib. Syst. **26**(1), 196–205 (2015)
25. Chen, Y.D., Li, K.L., Yang, W.D., Xiao, G.Q., Xie, X.H., Li, T.: Performance-aware model for sparse matrix–matrix multiplication on the sunway TaihuLight supercomputer. IEEE Trans. Parallel Distrib. Syst. (in press) https://doi.org/10.1109/TPDS.2018.2871189 (2018)
26. Han, T.X.: Non-life Insurance Actuarial, pp. 154–200. Chinese financial & Economic Publishing House, Beijing (2010)