Fuzzy Regression Prediction and Application Based on Multi-Dimensional Factors of Freight Volume

Mengting Xiao, Cheng Li*

School of Air Transportation/Flying, Shanghai University of Engineering Science, Shanghai 201620, China

*Corresponding author e-mail: licheng@sues.edu.cn

Abstract. Based on the reality of the development of air cargo, the multi-dimensional fuzzy regression method is used to determine the influencing factors, and the three most important influencing factors of GDP, total fixed assets investment and regular flight route mileage are determined. The system's viewpoints and analogy methods, the use of fuzzy numbers and multiple regression methods to predict the civil aviation cargo volume. In comparison with the 13th Five-Year Plan for China's Civil Aviation Development (2016-2020), it is proved that this method can effectively improve the accuracy of forecasting and reduce the risk of forecasting. It is proved that this model predicts civil aviation freight volume of the feasibility, has a high practical significance and practical operation.

1. Introduction
In the theoretical study of civil aviation freight forecasting, more time series method, regression analysis method, input-output method, gray prediction method and Markov chain forecasting method have been used in many theoretical studies, but there are also with many deficiencies:

1) Requires a lot of data, the data is less difficult to find statistical laws.
2) Requires that the sample be subject to a typical probability distribution, requiring a linear relationship between the factor data (socioeconomic indicators) and the system characteristic data (demand).

Therefore, this paper applies the fuzzy regression theory to study the forecasting accuracy of the random variables with large volatility on the basis of the credibility of the Chinese civil aviation in recent years, and to abandon the complexity of the pursuit strategy Nature and advanced nature of the method to solve the practical problems, but also in the case of less information to obtain more accurate results.

2. Influencing Factors of Civil Aviation Freight Transport and Construction of Fuzzy Regression Model

1) Analysis of influencing factors of civil aviation freight volume

There are many factors such as fuel price, exchange rate change, residents' disposable income, traffic expenditure, tourist number, fiscal revenue, population proportion, road freight volume, railway freight volume and postal service volume. Based on SPSS and Literature research, extract the most
important three influencing factors - gross domestic product (GDP), the whole society fixed asset investment, regular flight route mileage.

① Gross domestic product (GDP). Through the study found that GDP can directly or indirectly affect the demand for civil aviation transport. Based on time series analysis, the impact of GDP on civil aviation demand is inverted U-shaped, the impact on the freight volume is more significant.

② Total investment in fixed assets. The future investment in fixed assets has greatly contributed to the development of the transport industry. And civil aviation special construction requirements also determines its high investment in fixed assets for the whole society.

③ Regular flight mileage. With the flight and route mileage increase, will bring more convenience, but also save more time, more companies will choose the most efficient air cargo, so the regular flight route mileage on the impact of civil aviation cargo, is also very important.

(2) The construction of fuzzy regression model
According to the above factors, the fuzzy regression analysis of the original data, the regression analysis and forecasting of the GDP of the whole society, and the inclusion of the lagging variables in the regression equation, and the historical data and the recent forecast the mathematical model is introduced to improve the accuracy of long-term data forecasting. The mileage of scheduled flight routes is predicted by regression analysis method and the method of mean value is used to predict and determine the method. Finally, the fuzzy regression model is used to predict and analyze the future data, including the sampling test of the model in the original data, and then the test is applied to the forecast data to make the model more accurate. The main theoretical model is: 

\[ A_g = \sum_{i=1}^{m} b_i \bar{x}_{ig} + b_0 \]

g-level prediction factor of the sample average \( \bar{x}_{ig} \) into the formula, you can calculate \( A_g (g = 1, 2, ..., G) \). The fuzzy model (ie, fuzzy membership function) of N samples can be created in turn:

\[
y_1(X) = \begin{cases} 
1, & A_1 \geq Y_t \\
A_i / Y_t, & A_1 < Y_t 
\end{cases} \\
y_2(X) = \begin{cases} 
Y_i / A_2, & A_2 \geq Y_t \\
A_2 / Y_t, & A_2 < Y_t 
\end{cases} \\
\vdots \\
y_G(X) = \begin{cases} 
Y_i / A_G, & A_G \geq Y_t \\
A_G / Y_t, & A_G < Y_t 
\end{cases} 
\]

(1)

According to the fuzzy model, the membership degree of each level can be calculated separately: \( y_g (X) \) \((g = 1, 2, ..., G)\), fuzzy identification is carried out according to the maximum degree of membership degree, that is \( y_L (X) = \max \{ y_1 (X), y_2 (X), ..., y_G (X) \} \), if , the sample is assigned to the Lth level (1≤L≤G).

(3) Based on fuzzy regression model to predict the basic realization
Because the data may be abnormal value, the reliability of the model will be affected, and the center value \( m_j \) is reduced in the basic fuzzy regression linear model. Therefore, in order to improve the robustness of the model, Huber estimation, M-estimation And so on. On the other hand, to extract the data, from which to remove the abnormal value, and then "pure" data to do linear planning. Under the
condition that each \( Y_j \) membership is greater than \( H \in [0,1] \), the total ambiguity \( S = \sum_{j=1}^{m} W_j \) of the parameter is minimized, and the linear programming problem is obtained:

\[
\begin{align*}
\min S &= \sum_{j=1}^{m} W_j \\
\text{s.t.} \quad (1 - H)W^T |X_i| - |y_i - X_i^T a| &\geq 0 \\
X_i &\neq 0, i = 1,2,\ldots,n \\
W_j &\geq 0, j = 1,2,\ldots,n
\end{align*}
\]

(2)

Huber estimates, the application of minimum multiplication and M-estimation. The use of the three methods to improve the usability of the data, here referred to herein as the data "purity" to make the model results more reasonable and effective.

3. Construction of Civil Aviation Freight Volume Model Based on Fuzzy Regression

3.1. Data collection

(1) Data source

According to the "China Statistical Yearbook" to collect data from 1996 to 2015, as the original data, selected a dozen related factors and extracted the most important three related factors, to predict.

| years  | x1 (GDP billion yuan) | x2 (Total investment in fixed assets Billion) | x3 (Regular flight route mileage - ten thousand kilometers) | y (Freight volume - Million tons) |
|--------|----------------------|---------------------------------------------|------------------------------------------------------------|----------------------------------|
| 1996   | 71813.6              | 22913.5                                     | 116.65                                                     | 115.0                            |
| 1997   | 79715.0              | 24941.1                                     | 142.50                                                     | 124.7                            |
| 1998   | 85195.5              | 28406.2                                     | 150.58                                                     | 140.1                            |
| 1999   | 90564.4              | 29854.7                                     | 152.22                                                     | 170.4                            |
| 2000   | 100280.1             | 32917.7                                     | 150.29                                                     | 196.7                            |
| 2001   | 110863.1             | 37213.5                                     | 155.36                                                     | 171.0                            |
| 2002   | 121717.4             | 43499.9                                     | 163.77                                                     | 202.1                            |
| 2003   | 137422.0             | 55566.6                                     | 174.95                                                     | 219.0                            |
| 2004   | 161840.2             | 70477.4                                     | 204.94                                                     | 276.7                            |
| 2005   | 187318.9             | 88773.6                                     | 199.85                                                     | 306.7                            |
| 2006   | 219438.5             | 109998.2                                    | 211.35                                                     | 349.4                            |
| 2007   | 270232.3             | 137323.9                                    | 234.30                                                     | 401.8                            |
| 2008   | 319515.5             | 172828.4                                    | 246.18                                                     | 407.6                            |
| 2009   | 349081.4             | 224598.8                                    | 234.51                                                     | 445.5                            |
| 2010   | 413030.3             | 278121.9                                    | 276.51                                                     | 563.0                            |
| 2011   | 489300.6             | 311485.1                                    | 349.06                                                     | 557.5                            |
| 2012   | 540367.4             | 374694.7                                    | 328.01                                                     | 545.0                            |
| 2013   | 595244.4             | 446294.1                                    | 410.60                                                     | 561.3                            |
| 2014   | 643974.0             | 512020.7                                    | 463.72                                                     | 594.1                            |
| 2015   | 685505.8             | 561999.8                                    | 531.72                                                     | 629.3                            |
(2) Forecast object classification
According to the growth of freight volume, after several attempts to classify, apply the model to predict, and compare the forecast results with the original data to determine the optimal level. In this paper, the forecast object (y) is divided into five levels, the specific classification criteria are shown in Table 2 below.

| Level | Grading standards | Traffic | Corresponding year |
|-------|-------------------|---------|--------------------|
| 1     | $y \leq 170$      | less    | 1996 1997 1998     |
| 2     | $170 < y \leq 210$| less    | 1999 2000 2001 2002|
| 3     | $210 < y \leq 400$| medium  | 2003 2004 2005 2006|
| 4     | $400 < y \leq 560$| More    | 2007 2008 2009 2012|
| 5     | $y > 560$         | many    | 2010 2011 2013 2014 2015|

3.2. Multiple Regression and Fuzzy Model Construction
According to Table 3 to predict the object level, respectively, the average value of each sample: $\bar{x}_{ig} (i = 1, 2, 3; g = 1.2.3.4.5)$

| Level | $\bar{x}_1$ | $\bar{x}_2$ | $\bar{x}_3$ |
|-------|-------------|-------------|-------------|
| 1     | 78908.03    | 25420.27    | 136.58      |
| 2     | 105856.25   | 35871.45    | 155.41      |
| 3     | 176504.90   | 81203.95    | 197.77      |
| 4     | 369799.15   | 227361.45   | 260.75      |
| 5     | 565411.02   | 421984.32   | 406.32      |

Establishment of multiple regression models:

$$Y_i = 11.354067 + 0.002181x_{1i} - 0.001630x_{2i} + 0.038775x_{3i}$$ (3)

(1) According to the multiple linear regression model, the observation value of each forecast factor is substituted into $Y_t$, and the estimated value of freight volume $Y_t (t = 1, 2, ..., 20)$ is calculated (Table 4)

| Serial number | years | Estimated value $Y_t$ | Serial number | years | Estimated value $Y_t$ |
|---------------|-------|-----------------------|---------------|-------|-----------------------|
| 1             | 1996  | 135.1323              | 11            | 2006  | 318.7658              |
| 2             | 1997  | 150.0605              | 12            | 2007  | 385.8764              |
| 3             | 1998  | 156.6763              | 13            | 2008  | 435.9286              |
| 4             | 1999  | 166.0869              | 14            | 2009  | 415.5473              |
| 5             | 2000  | 198.4789              | 15            | 2010  | 597.7611              |
| 6             | 2001  | 182.2065              | 16            | 2011  | 584.1233              |
| 7             | 2002  | 212.2270              | 17            | 2012  | 469.3730              |
| 8             | 2003  | 227.2360              | 18            | 2013  | 591.6176              |
| 9             | 2004  | 257.3401              | 19            | 2014  | 598.9305              |
| 10            | 2005  | 282.8772              | 20            | 2015  | 610.6547              |
\[ A_g = \sum_{i=1}^{m} b_i \bar{x}_{ig} + b_0 \], the average value of the forecast factors at each level is substituted into Equation 3 to obtain the value of each sample, and Table 5 is shown below.

Table 5. values

| level | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|-------|--------|--------|--------|--------|--------|
| \( A_g \) | 147.3132 | 189.7821 | 271.6174 | 457.3974 | 572.4362 |

According to the formula (1), the establishment of \( A_g \) membership function,

\[
y_1(X) = \begin{cases} 
1, & 147.3132 \geq Y_i \\
147.3132/Y_i, & 147.3132 < Y_i 
\end{cases}
\]

\[
y_2(X) = \begin{cases} 
Y_i/189.7821, & 189.7821 \geq Y_i \\
189.7821/Y_i, & 189.7821 < Y_i 
\end{cases} \quad (t = 1, 2, \ldots, N)
\]

\[
y_3(X) = \begin{cases} 
Y_i/271.6174, & 271.6174 \geq Y_i \\
271.6174/Y_i, & 271.6174 < Y_i 
\end{cases}
\]

\[
y_4(X) = \begin{cases} 
Y_i/457.3974, & 457.3974 \geq Y_i \\
457.3974/Y_i, & 457.3974 < Y_i 
\end{cases}
\]

\[
y_5(X) = \begin{cases} 
Y_i/572.4362, & 572.4362 \geq Y_i \\
572.4362/Y_i, & 572.4362 < Y_i 
\end{cases}
\]

(2) From 1990 to 2015 randomly selected 10 data as a sample for regression testing. The year in which the original data was taken was 1996, 1998, 2000, 2001, 2002, 2005, 2007, 2009, 2013 and 2015. According to Table 4 to get the corresponding forecast.

Table 6. Raw data and forecast values

| years | GDP (billion yuan) | Total investment in fixed assets (Billion) | Regular flight route mileage (ten thousand kilometers) | \( Y \) | \( Y_t \) |
|-------|-------------------|------------------------------------------|-----------------------------------------------|-------|-------|
| 1996  | 71813.6           | 22913.5                                  | 116.65                                        | 115.0 | 135.1323 |
| 1998  | 85195.5           | 28406.2                                  | 150.58                                        | 150.29 | 156.6763 |
| 2000  | 100280.1          | 32917.7                                  | 150.29                                        | 150.29 | 196.7 | 198.4789 |
| 2001  | 110863.1          | 37213.5                                  | 155.36                                        | 155.36 | 171.0 | 182.2065 |
| 2002  | 121717.4          | 43499.9                                  | 163.77                                        | 163.77 | 202.1 | 212.2270 |
| 2005  | 187318.9          | 88773.6                                  | 199.85                                        | 199.85 | 306.7 | 282.8772 |
| 2007  | 270232.3          | 137323.9                                 | 234.30                                        | 234.30 | 401.8 | 385.8764 |
| 2009  | 349081.4          | 224598.8                                 | 234.51                                        | 234.51 | 445.5 | 415.5473 |
| 2013  | 595244.4          | 446294.1                                 | 410.60                                        | 410.60 | 561.3 | 591.6176 |
| 2015  | 685505.8          | 561999.8                                 | 531.72                                        | 531.72 | 629.3 | 610.6547 |

The following annual forecast values are substituted into the membership function of \( A_g \), and the following return verification data, the comparison of the forecast level and the actual registration are calculated.
Table 7. Predicted level vs. actual level

| years | y1       | y2       | y3       | y4       | y5       | prediction level | Poor with actual level |
|-------|----------|----------|----------|----------|----------|-----------------|-----------------------|
| 1996  | 1        | 0.71204  | 0.49751  | 0.29544  | 0.23607  | 1               | 0                     |
| 1998  | 0.94024  | 0.82556  | 0.57683  | 0.34254  | 0.27370  | 1               | 0                     |
| 2000  | 0.74221  | 0.95618  | 0.73073  | 0.43393  | 0.34673  | 2               | 0                     |
| 2001  | 0.80850  | 0.96008  | 0.67082  | 0.39836  | 0.31830  | 2               | 0                     |
| 2002  | 0.69413  | 0.89424  | 0.78135  | 0.46399  | 0.37074  | 2               | 0                     |
| 2005  | 0.52077  | 0.67090  | 0.96020  | 0.61845  | 0.49416  | 3               | 0                     |
| 2007  | 0.38176  | 0.49182  | 0.70390  | 0.84364  | 0.67410  | 4               | 0                     |
| 2009  | 0.35450  | 0.45670  | 0.65364  | 0.90850  | 0.72593  | 4               | 0                     |
| 2013  | 0.24900  | 0.32079  | 0.45911  | 0.77313  | 0.96758  | 5               | 0                     |
| 2015  | 0.24124  | 0.31079  | 0.44480  | 0.74903  | 0.93741  | 5               | 0                     |

From the above table can be very intuitive to see according to the forecast to verify the degree of fit up to 100%, high reliability forecast, the model has a very high reliability and practical significance, can be applied to the future data prediction.

4. Fuzzy Forecasting of Civil Aviation Freight Volume

(1) GDP total forecast

In this paper, we try to predict the regression model of the total amount of GDP by using the regression model of lagged one, lagging two, lag three and lag four periods. Finally, we use the lagged two phases to be closer to the actual value. \( Y = 1999.3x^2 - 8443.9x + 85389 \), its goodness of fit \( R^2 = 0.9960 \).

(2) The whole society fixed asset investment forecast

In this paper, we try to forecast the regression model of the fixed assets investment in the whole society by using the regression model of the lagged period, lagging two, lagging three and lagged four periods. The \( y = 43.9x^3 + 833.0x^2 - 7132.1x + 36713 \), the goodness of fit \( R^2 = 0.9982 \).

(3) Scheduled flight route mileage forecast

Through the study, it is found that the regression method and the average method are used to compare the predicted data, and the average number of the chain growth rate is used to predict the future data more accurate and reliable. Calculate the mean 0.09 from the chain growth rate, calculated from the average of the chain growth rate to the future 2025 to the regular flight route mileage data.

Table 8. Estimates of consumption levels

| Serial number | year | Consumption level estimates | Serial number | years | Consumption level estimates |
|---------------|------|----------------------------|---------------|-------|----------------------------|
| 1             | 2016 | 579.57                     | 6             | 2021  | 891.75                     |
| 2             | 2017 | 631.74                     | 7             | 2022  | 972.00                     |
| 3             | 2018 | 688.59                     | 8             | 2023  | 1059.49                    |
| 4             | 2019 | 750.57                     | 9             | 2024  | 1154.84                    |
| 5             | 2020 | 818.12                     | 10            | 2025  | 1258.77                    |

(4) According to Equation 3, calculate the value of the future civil aviation cargo volume is extremely forecasting grade, as shown in Table 9 below:
**Table 9.** Forecast value of civil aviation freight in 2016-2020

| years | Civil aviation cargo volume (Million tons) | Prediction level | years | Civil aviation cargo volume (Million tons) | Prediction level |
|-------|------------------------------------------|-----------------|-------|------------------------------------------|-----------------|
| 2016  | 679.1048                                 | 5               | 2019  | 803.2105                                 | 5               |
| 2017  | 719.2027                                 | 5               | 2020  | 858.2963                                 | 5               |
| 2018  | 766.0419                                 | 5               |       |                                          |                 |

5. Conclusion

Based on the study of fuzzy regression theory and the study of civil aviation freight industry, this paper applies the fuzzy regression model to the forecast of air cargo volume, and predicts the traffic volume of civil aviation under the condition of small amount of data. The results show that the fuzzy regression model provides an excellent forecasting model for the forecast of air cargo volume. According to the requirements of the thirteenth five-year plan (2016-2020) of China's civil aviation development, 8.5 million tons and predicted by the model of 8.58 million tons of the basic match, the growth rate of 6.03% with the "thirteen" target 6.2% basically consistent. It can be seen that the fuzzy regression theory model established in this paper has high accuracy and foresight, and the prediction accuracy is high, which has high practical significance and practical operation.

References

[1] Dai Hong. Traffic flow forecasting model based on fuzzy regression [J]. Journal of Jilin Institute of Engineering Technology, 2011, 27 (6): 77 ~ 80.
[2] Ding Shifei. Multivariate fuzzy regression prediction model and its application [J]. Fuzzy Systems and Mathematics, 2000, 14 (3): 74 ~ 98.
[3] Cheng Dingyuan. Optimization of GM (1, 1) Model Based on Fuzzy Regression Theory [J]. Journal of Hefei University of Technology (Natural Science Edition), 2009, 32 (5): 754 ~ 758.
[4] Chen Yizeng. Fuzzy House Decision Model Based on Fuzzy Regression [J]. Information and Control, 2003, 32 (2): 118 ~ 121.
[5] William L. Cooper. Models for Air-Cargo Allotment Management and Booking Control [J]. Kannapha Amaruchkul, 2007.
[6] Dujian. Application of fuzzy multiple regression model to forecast the transportation elasticity coefficient [J]. Journal of Wuhan University of Technology (Transportation Science and Engineering Edition), 2009, 2.
[7] He Meiling. Tourism traffic demand forecasting method [J]. Logistics Science and Technology, 2006.5: 83 ~ 85.
[8] Wen Wen. Fuzzy variable coefficient regression model of the improvement and application of [D]. East China University of Science and Technology, 2013.
[9] Zhao Jianyou, Zhou Sunfeng, Cui Xiaojuan, Wang Gaoqing. Measurement of Highway Freight Volume Based on Fuzzy Linear Regression Model [J]. Journal of Traffic and Transportation Engineering, 2012, 12 (03): 80-85.
[10] Yang Wenquan. Review of fuzzy regression analysis [D]. Jilin University, 2009.
[11] Yang Zhiming, Deng Naiyang. Virtual Linear Support Vector Regression [J]. Computer Engineering and Applications, 2004, (36): 54-57.
[12] Liu Ximing. Measurement and Verification Analysis of Highway Logistics Based on Mathematical Model [J]. Logistics Technology, 2014, 33 (09): 252-254.
[13] Hu Wequn. Based on freight strength changes in the law of freight forecasting method [J]. Transportation Science and Technology, 2011, (02): 107-109.