Liability insurance premium income forecast based on improved GM (1,1) - A case of Shandong Province

Xiaojun Yang 1 and Mingqing Zhao 2*

1 College of Mathematics and Systems, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China
2 College of Mathematics and Systems, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China
*Corresponding author’s e-mail: yxjzzczx0628@sdust.edu.cn

Abstract. Liability insurance premium income is not only an important source of insurance company profits, but also an important aspect used to evaluate the development level of the insurance industry. Therefore, it has important practical significance for accurate prediction. This paper uses improved GM (1,1) to predict the premium income of liability insurance in Shandong Province, and the results show that the fitting effect is very good.

1. Introduction

Liability insurance is a compensatory insurance, it is a type of insurance based on which the insured may cause losses to the interests of others during the validity period of the insurance. Liability insurance based business can be divided into employers, public, third party, product and professional liability insurance business five categories, it is widely used and can be used for all individuals, families, or various units that may cause personal injury or property damage to others. In 2019, Shandong Province's liability insurance premium income was RMB 4.034 million, which is close to 22.7 times of 2006 premium income and is growing rapidly. In the context of normal rapid economic development and continuous structural optimization, focusing on the development of liability insurance is a necessary choice for the rapid development of my country's insurance industry. Therefore, predicting liability insurance premium income has very important practical significance.

There are many ways to forecast the premium income. Diao Li et al. constructed X12-LSTM model based on (Long Short-Term Memory, LSTM) neural network to predict the property insurance premium income of the original CIRC website from July 2004 to November 2018 in [1]. The results show that this model is more suitable for some data with seasonal characteristics. Liang Laicun et al. used ARIMA model and selected China's monthly premium income from January 1999 to July 2005 to forecast [2]. Shao Sheng et al. used SARIMA model and selected the data of China accident insurance from January 2006 to November 2015 for prediction [3]. The results show that the fitting effect of the model is better, but the trend growth and seasonal contrast of the future prediction data are obvious. In response to the above problems, Zhou Hua et al. based on the TEI@I method, constructed a nonlinear integrated forecasting model with correction errors and machine learning capabilities, and selected monthly data of my country's premium income as the research object for forecasting [4]. The results show that this model compared with the SARIMA model and the ordinary linear addition mixed model, its prediction accuracy is higher. Wu Kaibing et al. used the modified Logistic model to predict premium income indirectly, and the results showed that the compound growth rate of premium income was approximately
10.006% [5]. Liu Hongliang separated the four influencing factors of premium income based on the X12 seasonal adjustment method, and then respectively used the ARIMA model to predict, and finally combined them to obtain the final forecast value [6]. He Shujing et al. used the BP neural network to predict the life insurance premium income of my country from 1989 to 2005 [7]. The results show that the model has a higher prediction accuracy than the econometric model, which highlights the strong application value of the model. GM (1,1) proposed by Deng Julong is an important forecasting model, which uses some known and related data generation rules to predict the future development trend of things when there are few data or related information [8]. There have been many successful examples of using GM (1,1) for prediction [9-11], but there are also some cases where the deviation of the prediction results is too large. The background value is an important parameter that affects the prediction effect of the model. In order to improve the prediction effect of the model, scholars have conducted in-depth research on the construction of the background value from different angles, and have achieved some valuable results [12-15]. This paper uses the improved GM (1,1) in literature [13] to predict the premium income of liability insurance in Shandong Province.

The data processing in this paper is realized by MATLAB programming.

2. Improve the prediction principle and steps of GM (1,1)

2.1. Model establishment

There are original data columns

\[ X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(N)) \]

A new sequence can be obtained by accumulating \( X^{(0)} \)

\[ X^{(1)} = (X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(N)) = (X^{(0)}(1), X^{(0)}(1) + X^{(0)}(2), \cdots) \]

Among

\[ X^{(1)}(k) = \sum_{i=1}^{k} X^{(0)}(i), k = 1,2, \cdots, n \]

GM (1,1) corresponding gray differential equation \( X^{(0)}(k) + aZ^{(1)}(k) = b \)

Corresponding albinism differential equation \[ \frac{dX^{(1)}}{dt} + aX^{(1)} = b \] (1)

Among them, \( Z^{(1)}(k) \) is the background value sequence constructed by \( X^{(1)}(k) \), and the coefficients \( a \) and \( b \) are the development gray number and gray effect.

The solution of formula (1) can be obtained as

\[ \hat{X}^{(1)}(k + 1) = X^{(0)}(1) - \frac{b}{a} e^{-ak} + \frac{b}{a} \]

Thus subtract consecutively generates predicted value

\[ \hat{X}^{(0)}(k + 1) = \hat{X}^{(1)}(k + 1) - \hat{X}^{(1)}(k), k = 1,2, \cdots, n \]

Least square estimation \( \hat{u} = (B^T B)^{-1} B^T Y \) of the coefficient vector \( u = (a, b)^T \), among

\[ B = \begin{bmatrix} -Z^{(1)}(2) & 1 \\ \vdots & \vdots \\ -Z^{(1)}(n) & 1 \end{bmatrix}, Y = (X^{(0)}(2), X^{(0)}(3), \cdots, X^{(0)}(n))^T \]

By integrating formula (1), obtain
\[ \int_k^{k+1} \frac{dX^{(i)}}{dt} dt + a \int_k^{k+1} X^{(i)} dt = b \]

Namely
\[ X^{(i)}(k+1) - X^{(i)}(k) + a \int_k^{k+1} X^{(i)} dt = b \]
\[ X^{(0)}(k+1) + a \int_k^{k+1} X^{(i)} dt = b \]

Thereupon
\[ Z^{(i)}(k+1) = \int_k^{k+1} X^{(i)} dt \]

Treat it as follows:

(1) Lagrange interpolation polynomials are constructed by using three points
\[ (k, X^{(1)}(k)), (k+1, X^{(1)}(k+1)), (k+2, X^{(1)}(k+2)) \]
\[ P_2(t) = X^{(1)}(k) \frac{(t-k-1)(t-k-2)}{(-1) \cdot (-2)} + X^{(1)}(k+1) \frac{(t-k)(t-k-2)}{1 \cdot (-1)} + X^{(1)}(k+2) \frac{(t-k)(t-k-1)}{1 \cdot 2} \]

Make
\[ P_2(t) \approx X^{(1)}(t), t \in [k, k+1] \]

(2) Divide the interval into three equal parts to obtain three subintervals
\[ (k+\frac{1}{3}, X^{(1)}(k+\frac{1}{3})), (k+\frac{2}{3}, X^{(1)}(k+\frac{2}{3})), (k+1, X^{(1)}(k+1)) \]

This moment
\[ X^{(1)}(k+i\frac{1}{3}) \approx P_2(k+i\frac{1}{3}), i = 1,2 \]

Then
\[ X^{(1)}(k+\frac{1}{3}) \approx P_2(k+\frac{1}{3}) = \frac{5}{9} X^{(1)}(k) + \frac{5}{9} X^{(1)}(k+1) - \frac{1}{9} X^{(1)}(k+2) \]
\[ X^{(1)}(k+\frac{2}{3}) \approx P_2(k+\frac{2}{3}) = \frac{2}{9} X^{(1)}(k) + \frac{8}{9} X^{(1)}(k+1) - \frac{1}{9} X^{(1)}(k+2) \]

(3) Obtain the piecewise interpolation function on the interval \([k, k+1]\)
\[ S_k(t) = \begin{cases} X^{(1)}(k) + \frac{2}{3} X^{(1)}(k+1) - \frac{1}{3} X^{(1)}(k+2), & k \leq t \leq k+\frac{1}{3} \\ X^{(1)}(k+\frac{2}{3}) + \frac{2}{3} X^{(1)}(k+1) - \frac{1}{3} X^{(1)}(k+2), & k+\frac{1}{3} \leq t \leq k+\frac{2}{3} \\ X^{(1)}(k+1) + X^{(1)}(k+\frac{2}{3}) - \frac{2}{3} X^{(1)}(k+2), & k+\frac{2}{3} \leq t \leq k+1 \end{cases} \]

Thus
\[ Z^{(i)}(k+1) = \int_k^{k+1} X^{(i)} dt \approx \int_k^{k+1} S_k(t) dt = \int_k^{k+\frac{1}{3}} S_k(t) dt + \int_{k+\frac{1}{3}}^{k+\frac{2}{3}} S_k(t) dt + \int_{k+\frac{2}{3}}^{k+1} S_k(t) dt \]
\[ = \frac{1}{6} X^{(1)}(k+1) + \frac{1}{3} X^{(1)}(k+\frac{2}{3}) + \frac{1}{3} X^{(1)}(k+\frac{1}{3}) + \frac{1}{6} X^{(1)}(k) \]

Thus
\[ Z^{(i)}(k+1) = \int_k^{k+1} X^{(i)} dt \approx \int_k^{k+1} S_k(t) dt = \frac{23}{54} X^{(1)}(k) + \frac{35}{54} X^{(1)}(k+1) - \frac{2}{27} X^{(1)}(k+2) \]
2.2. Model test
(1) Residual test
Residual \( e^{(0)}(k) = X^{(0)}(k) - \hat{X}^{(0)}(k) \)
Relative error \( \Delta(k) = \frac{e^{(0)}(k)}{X^{(0)}(k)} \times 100\% \)
Average relative error \( \bar{\Delta} = \frac{1}{n} \sum_{k=1}^{n} \Delta(k) \times 100\% \)
Grey prediction model residual test, one is given \( \alpha \) (in general prediction \( 0.01 \leq \alpha \leq 0.05 \)), when \( \bar{\Delta} < \alpha \) is established, the model residual test is judged to be qualified, and the other is to judge the accuracy level of the corresponding model based on the value.

(2) Posterior error test
Record \( \bar{X} = \frac{1}{n} \sum_{k=1}^{n} X^{(0)}(k), S_1^2 = \frac{1}{n} \sum_{k=1}^{n} (X^{(0)}(k) - \bar{X}), \bar{\varepsilon} = \frac{1}{n} \sum_{k=1}^{n} e^{(0)}(k), S_2^2 = \frac{1}{n} \sum_{k=1}^{n} (e^{(0)}(k) - \bar{\varepsilon}) \)
The posterior ratio \( C = \frac{S_2}{S_1} \) is used to judge the accuracy level of the model based on the value.

(3) Small error probability test
Record \( P = P \{ | e^{(0)}(k) - \bar{\varepsilon} | < 0.6745S_1 \} \)
Determine the accuracy level of the corresponding model based on the value of \( P \).
The accuracy test of GM (1,1) is shown in Table 1.

| Level | \( \bar{\Delta} \) | \( C \) | \( P \) |
|-------|-----------------|------|------|
| Level 1 | < 0.01 | < 0.35 | > 0.95 |
| Level 2 | < 0.05 | < 0.50 | < 0.80 |
| Level 3 | < 0.10 | < 0.65 | < 0.70 |
| Level 4 | > 0.20 | > 0.80 | < 0.60 |

2.3. Prediction steps
The steps to improve GM (1,1) for prediction are as follows:
(1) Accumulative generating sequence \( X^{(1)} \);
(2) Calculate the background value \( Z^{(1)}(k+1) = \frac{23}{54} X^{(1)}(k) + \frac{35}{54} X^{(1)}(k+1) - \frac{2}{27} X^{(1)}(k+2) \)
(3) Seeking the least squares estimate \( \hat{u} = (B^T B)^{-1} B^T Y \)
(4) Calculation \( \hat{X}^{(1)}(k+1) = \left[ X^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{\hat{b}}{a} \), further calculate the cumulative subtraction to generate the predicted value \( \hat{X}^{(0)}(k+1) = \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k), k = 1, 2, \cdots, n \)
(5) Test and judge the accuracy level of the model.

3. Liability insurance premium income forecast
From 2010 to 2019, the premium income of liability insurance in Shandong Province is shown in Table 2 and Figure 1. It can be seen that the premium income of liability insurance in Shandong Province is increasing steadily year by year. In 2019, it has reached 4.03456 billion yuan, nearly 8 times the amount in 2010. It shows that people are paying more and more attention to liability insurance.
Table 2 2010-2019 Liability insurance premium income of Shandong Province (unit: RMB 100000)

| Years | 2010     | 2011     | 2012     | 2013     | 2014     | 2015     | 2016     | 2017     | 2018     | 2019     |
|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Premium | 5611.1   | 6852.1   | 8734.9   | 10968.2  | 13080.7  | 16745.8  | 21426.9  | 25346.6  | 32853.6  | 40345.6  |

Data source: 2010-2019 Shandong Province statistical yearbook

Figure 1 Premium income of liability insurance in Shandong Province (unit: RMB 100000)

We use the traditional GM (1,1) and the improved GM (1,1) to get two groups of forecast data of liability premium income in Shandong Province from 2010 to 2019 (see Table 3), and compare the average relative error $\bar{\Delta}$, posterior ratio $C$ and small error probability $P$ (see Table 4). It can be seen from Table 4 that both the average relative error and the posterior ratio of the improved GM (1,1) are smaller than those of the traditional GM (1,1), and the prediction accuracy and accuracy are greatly improved.

Table 3 Prediction and comparison of premium income of liability insurance in Shandong Province from 2010 to 2019 (unit: RMB 100000)

| Years | 2010     | 2011     | 2012     | 2013     | 2014     | 2015     | 2016     | 2017     | 2018     | 2019     |
|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Original data | 5611.1   | 6852.1   | 8734.9   | 10968.2  | 13080.7  | 16745.8  | 21426.9  | 25346.6  | 32853.6  | 40345.6  |
| GM (1,1) | 5611.1   | 6908     | 8603.5   | 10715    | 13345    | 16621    | 20700    | 25781    | 32109    | 49805    |
| Improved GM (1,1) | 5611.1   | 6901.7   | 8612.2   | 10747    | 13410    | 16733    | 20881    | 26056    | 32513    | 40571    |

Table 4 Comparison of prediction effects between traditional GM (1,1) and improved GM (1,1)

|                | Traditional GM (1,1) | Improved GM (1,1) |
|----------------|----------------------|--------------------|
| Average relative error $\bar{\Delta}$ | 1.57%              | 1.37%              |
| Posterior ratio $C$          | 0.0328              | 0.0306             |
| Small error probability $P$    | 1                   | 1                  |

The premium income forecast of improved GM (1,1) is fitted with the original data as shown in Figure 2, which shows that the forecast effect is very good. According to Table 4, the average relative error of the improved GM (1,1) is $\bar{\Delta} = 1.37% < 5\%$, the posterior ratio is $C = 0.0306 < 0.35$, and the small error probability is $P = 1 > 0.95$. Therefore, the accuracy of the model is first level.

Figure 2 Fitting of improved GM (1,1) prediction results with original data

Using the improved GM (1,1) to predict the liability insurance premium income of Shandong Province in the next 6 years, the results are shown in Table 5. It can be seen that between 2020 and 2025,
Shandong's liability insurance premium income will still grow at a relatively rapid rate. At present, our country is in a period of social and economic transition, and the corresponding social public safety is facing severe tests. Infringements such as environmental pollution, disputes between doctors and patients, tourism, production, and food safety occur frequently. Therefore, As people pay more and more attention to liability insurance, liability insurance premium income will continue to increase accordingly.

Table 5 2020-2025 Shandong Province liability insurance premium income forecast (unit: RMB 100000)

| Years | 2020  | 2021  | 2022  | 2023  | 2024  | 2025  |
|-------|-------|-------|-------|-------|-------|-------|
| Predictive value | 50626 | 63173 | 78830 | 98367 | 122750 | 153170 |

4. Conclusion

This article uses improved GM (1,1) to predict the premium income of liability insurance in Shandong Province, which has high accuracy and credibility, and has important guiding significance for promoting the stable development of the insurance industry. The improved GM (1,1) is suitable for the data with large changes in the short term, but for the data with small changes in the short term, the fitting effect is not particularly significant. Further consideration can be given to the combination of improved initial value, optimized background value and residual correction for prediction, and the fitting effect may be more significant.

References

[1] Diao, L., Wang, N. (2020) Research on premium income prediction based on X12-LSTM model[J]. Computer Science, 47(S1): 512-516.
[2] Liang, LC., Pi, YJ. (2006) The ARIMA model and forecast of my country’s premium income[J]. Statistics and Decision, (07): 25-26.
[3] Shao, S., Shen, Y., Shen, SC., Fu, GL., Chen, P. (2016) The prediction research of SARIMA model on accident insurance in China[J]. Journal of Qiqihar University (Natural Science Edition), 32(05): 88-94.
[4] Zhou, H., Lu, ZY., Zheng, M. (2020) Prediction of China's insurance industry premium income based on the TEI@I method[J]. Management Review, 32(07): 166-179.
[5] Wu, KB., Xing,W., Yu, ZY. (2000) The changing law of insurance depth and the prediction of my country's insurance market[J]. Forecast, (03): 53-56.
[6] Liu, HL. (2013) Analysis of Short-term Premium Income Forecast Model in Chongqing——Based on X12 Seasonal Adjustment Method and ARIMA Model[J]. Shang, (16): 130-131.
[7] He, SJ., Xiang, XD. (2006) Research on the Application of BP Neural Network in my country's Personal Insurance Premium Income[J]. Value Engineering, (11): 96-98.
[8] Deng, JL. (1982) Gray control system [J]. Journal of Huazhong Institute of Technology, (03): 9-18.
[9] Wang, XJ. (2012) Research on forecasting theory and method based on data analysis [D]. Hefei: Hefei University of Technology.
[10] Feng, ST. (2018) Co-integration analysis of life insurance premium income and GM(1,1) gray forecast [D]. Fujian: Fujian Normal University.
[11] Xu, LL., Liang, GG., Wang, JJ. (2010) The application of GM (1,1) model in insurance premium income prediction[J]. Modern Business and Trade Industry, 22(11): 24.
[12] Yang, XL., Zhou, M., Zeng, B. (2018) A new method for the construction of the background value of the gray prediction model[J]. Statistics and Decision, 34(19): 14-18.
[13] Tian, ZC., Ji, G., Liu, M. (2021) Analysis and prediction of Xinjiang population based on improved grey GM(1,1) model[J]. Mathematics in Practice and Knowledge, 51(05): 258-264.
[14] Tan, GJ., Tan, JY., Wang, JY. (2015) Research on background value reconstruction of grey system prediction model GM (1,1) [J]. Mathematics in Practice and Knowledge, 45(15): 267-273.
[15] Chen, PY., Deng, HW. (2019) Improvement status and application of GM (1,1) model[J]. Statistics and Decision, 35(06):80-82.