A Study on the Decision of Insurance Order Quantity in Disaster-prone Areas Based on Grey Prediction and Multivariate Linear Regression

Zhuohang Du*

Aberdeen Institute of Data Science and Artificial Intelligence, South China Normal University, Foshan, China, 528234

*Corresponding author: u02zd21@abdn.ac.uk

Abstract. In recent years, earthquakes, floods, sandstorms, extreme heat and other natural disasters have occurred with increasing frequency in all corners of the Earth, and these disasters have caused huge economic losses. In this paper, the data of the disaster-prone state of Florida is taken as an example. In order to make the local insurance companies better make insurance decisions, the article first estimate the disaster frequency of the region in the next five years based on Grey Prediction, and then obtain the average expected weather-related disaster frequency of Sky and Land disasters (431.24 and 428.86). At the same time, the article carries out Multivariate Linear Regression on insurance sales (dependent variable) and disaster frequency (independent variable), obtaining the linear regression equation of the state, and calculates the expected value of insurance sales in the next five years (14.2372 million). Then, paper uses the concept of marginal benefit in economics and curve fitting toolbox to fit the marginal insurance cost curve and marginal insurance revenue curve, and take the intersection point of the curve (20.2692 million) as the number of underwriting orders that make the insurance company most profitable. Through comparison, it is found that the underwriters in Florida should continue to conduct business.

Keywords: Grey Prediction Model, Linear Regression, Marginal benefit.

1. Introduction

In recent years, the increased frequency and intensity of extreme weather events have posed unprecedented challenges to the global insurance industry. From storms and floods to extreme heatwaves, these weather phenomena have had direct impacts on human life and significantly affected the risk assessment and financial stability of the insurance sector. In response, insurance companies need to adopt more flexible and innovative approaches to adapt to this evolving risk landscape. However, past literature research mainly focused on insurance business models and risk management methods in conventional areas [1-3], and few studies focused on insurance business in areas with frequent disasters. In these disaster-prone areas, insurance companies can not only stay in the past insurance strategy formulation thinking, must be adjusted according to the local situation. Specifically, firstly, frequent disaster events increase the uncertainty of risk assessment, so that traditional risk models and pricing strategies may not be applicable anymore. Secondly, risk concentration is high in areas with frequent disasters, which may lead to a large number of claims and bring huge financial pressure to insurance companies once disasters occur [4-5]. Therefore, it is an important and pressing issue for insurance companies to conduct business in these regions and remain profitable. With this background, the paper aims to help local companies adjust their decision to underwrite insurances by calculating the number of insurance sales over the next five years and the number of insurances underwrote that will be most profitable for insurers in disaster-prone areas in this paper. By comparing these two indicators, the adjustment direction of insurance strategies in disaster prone areas can be obtained. In addition, the loss of disaster data come from National Centers for Environmental Information (NCEI) (noaa.gov) Insurance data come from Insurance Information Institute.

2. Grey prediction model

2.1. Pole ratio test

Since the obtained data is a small amount of data, and the Grey Prediction Model can calculate and speculate according to the small amount of information available [1-2], article uses the modified model to forecast. In order to predict the number of sky disasters and land disasters in Florida in the next five years, Grey Prediction Model is performed based on the annual data of the above two types of weather disasters in Florida from 2014 to 2023. First of all, the polar ratio test should be carried out on the data of the three disasters, which is located in the interval $\left(e\hat{\ }(-2/(n+1)), e\hat{\ }(2/(n+1))\right)$.The data is suitable for model construction. If the data does not pass the level ratio test, the sequence is translated so that the sequence meets the level ratio test after translation conversion.

Through testing, it is found that all values of the translated sequence were within the interval (0.834, 1.199), indicating that the translated sequence was suitable for constructing the gray prediction model

2.2. Establishment and results of Grey prediction model

Grey differential equation is obtained by using OLS estimation：

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{1}$$

Winterization equation:

$$\frac{dx^{(1)}(t)}{dt} = -\hat{a}\,x^{(1)}(t) + \hat{b} \tag{2}$$

Residual test

After the results are obtained, residual test is conducted to judge the degree of fit of the model, and the residual test is the following formula

Absolute residuals：

$$\varepsilon(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 2,3,...,n \tag{3}$$

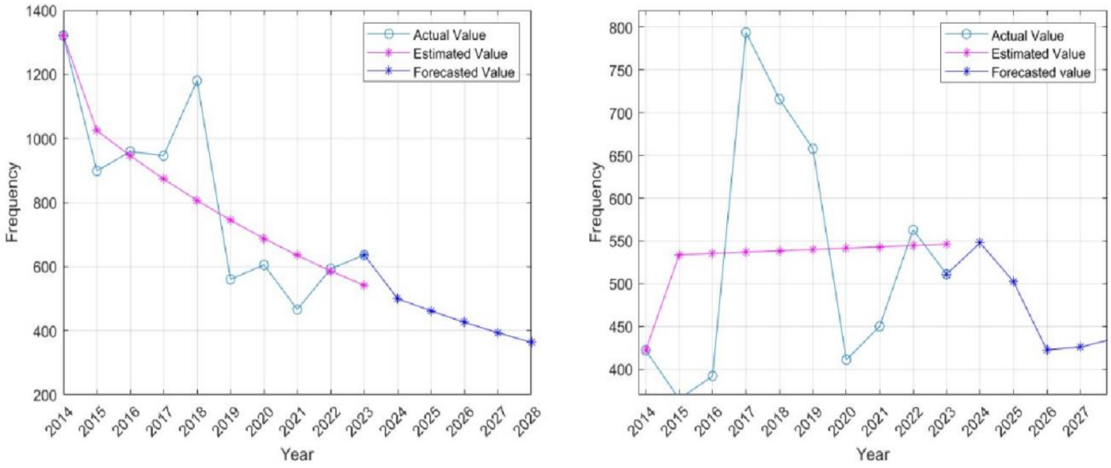
Relative residual:

$$\varepsilon_r(k) = \frac{|x^{(0)}(k) - \hat{x}^{(0)}(k)|}{x^{(0)}(k)} \times 100\%, k = 2,3,...,n \tag{4}$$

Average relative residual:

$$\bar{\varepsilon}_{\mathrm{r}} = \frac{1}{n-1}\sum_{k=2}^{n}|\varepsilon_r(k)| \tag{5}$$

The visualization of predicted results are shown in Figure 1:



(a) Sky disaster focast (a) land disaster focast

Figure 1 Grey Prediction Model result

The predicted frequency of five years future disasters are shown in Table.1.

Table.1. Frequency of future five years disasters in Florida

Prediction order	Sky disaster prediction	Land disaster prediction
1	547.760	499.812
2	502.489	461.515
3	422.624	426.147
4	425.662	393.489
5	435.663	363.333

Plug in the data to get the results and analysis:

Sky disaster: the average relative residual is 15.891%, less than 20%, which proves the degree of fit is good

Land hazards: The average relative residual is 18.759%, less than 20%, which proves the degree of fit is good

3. Regression Analysis

3.1. Normal distribution test of data

The article uses the Q-Q chart to show the distribution of the acquired data. To use the Q-Q plot map to identify whether the sample data is approximately normal distribution, just look at whether the points on the QQ map are approximately near a straight line [3]. Through Figure 2. it is clear to see that the data conform to the normal distribution, so we can think that there are no outliers in the acquired data, which can be used for the following analysis

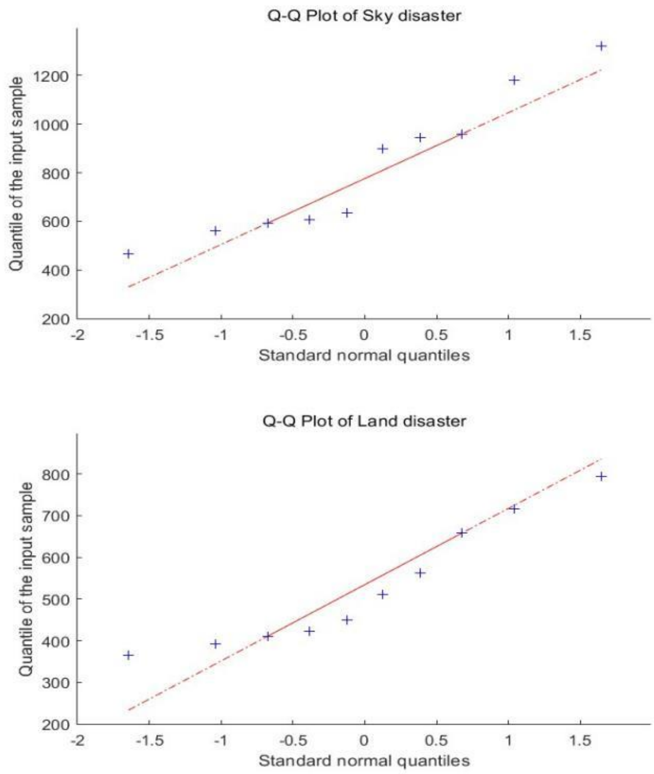


Figure 2. Q-Q plot of disaster data

The article visualizes the data of sky disasters, land disasters and the total number of disasters and the sales for insurance, drew a line chart shown as Figure 3, and found that the frequency of disasters seemed to have a certain correlation with the sales for insurance. Based on this, the paper tries to establish a regression equation to obtain the relationship between them.

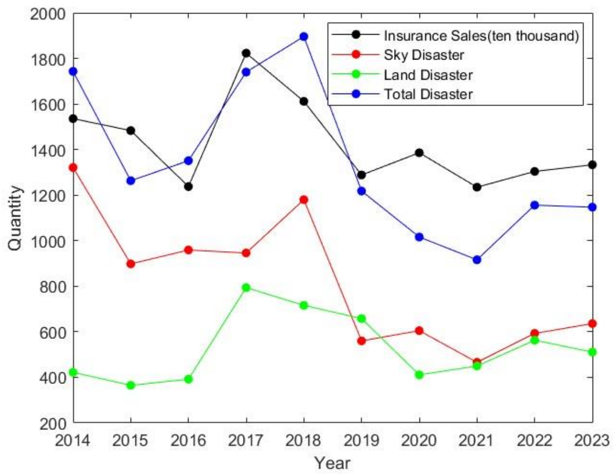


Figure 3. Line chart of disaster and insurance demands data

3.2. Linear model hypothesis

The types and contents of variables are shown in Table.2.

Table.2. Variable type

Variable Type	Variable	Name
Quantitative index	Dependent variable	Insurance sales
	Independent variable	Sky disaster
		Land disaster

From the perspective of economics [4], we improve the initial regression equation:

In order to weaken the heteroscedasticity of the data, logarithms of the independent variables is taken.

Assuming that the greater the number of disasters, the greater the loss caused, then the demand for insurance that conforms to behavioral insurance will also increase with the law of the elastic market, so the article needs to take the logarithm of the independent variable

At the same time, since the Insurance sales cannot be less than zero, the article takes $\ln(1+f_i)$

Above all for independent variable $F = \{f_1, f_2\}$={Sky disaster, Land disaster }, Assume S = Insurance sales, and the following linear relationship is satisfied：

$$S = \beta_0 + \sum_{i=1}^{2} \beta_i ln(1+f_i) + \epsilon \tag{6}$$

Where ϵ is a disturbance term that is difficult to observe and completely random.

3.3. Regression result

First, the article exams regression coefficient β_i of f_i and gets $p = 0.0209 < 0.05$ conducting a joint significance test. Therefore, it can be argued that under the 95% confidence interval, the original assumption is rejected that the null hypothesis and think that β_i is not all 0.

Table.3. Regression result

S	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
f_1	1703202	606311.1	2.81	0.026	269503.9	3136900
f_2	1675181	781283.4	2.14	0.069	-172261.1	3522622
_cons	-2.93E+07	1.24E+07	-2.36	0.05	-5.87E+07	6.19E+04

According to Table.3. all independent variables f_i are very significant, so the final regression equation can be obtained as follows:

$$S_{Florida} = -29{,}390{,}000 + 3{,}410{,}000 ln(1+f_1) + 3{,}357{,}000 ln(1+f_2) \tag{7}$$

Through the estimated number of disasters obtained by the previous gray prediction model and the regression equation, the article can obtain the insurance sales of the insurance market in Florida in the next five years. By taking the average value, the paper determines that this is the expected insurance sales of the region in the next five years E_S

$$E_S = \frac{\sum_{i=1}^{5} S_i'}{5} \tag{8}$$

Finally, according to the data obtained by the Grey Prediction Model, it can be plugged into the regression equation to obtain the average insurance underwrote in the next five years: 14.2372 million.

4. Marginal model

4.1. Marginal insurance cost

The article assumes that the insurance industry is dominated by real estate. In order to estimate the loss scope of insurers, the article considers how much contribution WI the insurance industry can make to the social industry, and then judge the approximate economic loss suffered by the insurance industry in the region. Then, to further refine the loss, in order to withstand the impact of extreme weather disasters, the disaster resistance index of the house S will determine whether the building can be damaged as little as possible in the disaster, which will further affect the cost of reconstruction and maintenance [5-7], and ultimately affect the loss of the insurer. In addition, DC is the damage loss cost by disasters. Therefore, such a relationship can be established:

$$MIC = DC \times WI \times S \tag{9}$$

At the same time, the firmness of the house S_W and the material used in the house S_M should be the variables that can best reflect the disaster resistance index of the house [8-9], and the better this index is, the smaller the loss should be, so the final expression is shown as follows:

$$S = \frac{1}{e^{S_W \times S_M}} \tag{10}$$

$$MIC = DC \times WI \times \frac{1}{e^{S_W \times S_M}} \tag{11}$$

4.2. Marginal insurance revenue

Assuming that the revenue of the insurance industry only considers the insured income related to real estate, the article considers the marginal profit as the contribution WI of the insurance industry to the local output value R

$$MIR = R \times WI \tag{12}$$

Thus, based on past data, the MIC and MIR of the insurance industry in the selected region each year can be computed. Finally, through fitting, the article obtains the MIC curve and MIR curve, and finally calculate that the insurance sales corresponding to the intersection of the curve is the optimal underwriting singular number [10].

5. Results

Based on the data of Florida, the article obtaines the following results through calculation.

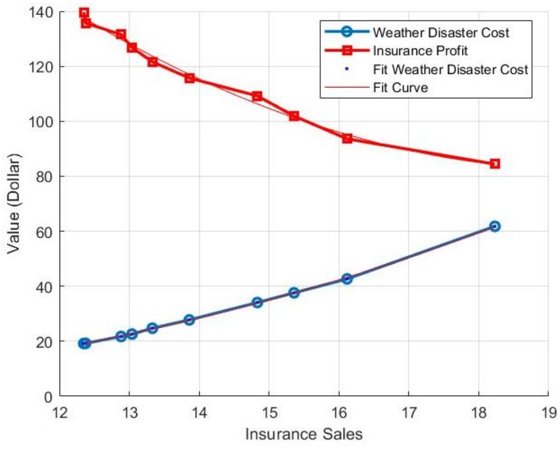


Figure 4 Marginal cost curve and Marginal revenue curve

Using the Matlab curve fitting toolbox, Marginal insurance cost curve and Marginal insurance revenue curve can got as well as the intersection point namely the optimal value. From this, the article obtains the optimal underwriting number of Florida insurance market is 20.2692 million.

6. Conclusions

Through Grey Prediction, the article obtains the average frequency of Sky disasters and Land disasters in Florida in the next five years as 431.24 and 428.86. At the same time, the paper obtains the Multivariate Linear Regression equation between insurance sales and disaster frequency in Florida, and based on the Grey Prediction value, got a five-year forecast of 14.2372 million insurance sales in Florida. Finally, based on the concept of marginal benefit, using the curve fitting toolbox, the Marginal insurance cost curve and the Marginal insurance revenue curve are clear, and the intersection point of the curve, 20.2692 million is found. By comparison, it can be found that according to the past trend, the expected value of insurance underwrote in this area is lower than the optimal insured amount. Therefore, the article recommends that the local underwriters in Florida further expand the insurance business to maximize its benefits.

References

[1] Chen Yuxi. (2024). A study of the problem of insurance company product liability insurance risk management master's degree thesis, east China normal university). https://link.cnki.net/doi/10.27149/d.cnki.ghdsu.2024.000021

[2] Zheng X. (2023). How to effectively prevent economic risks in the insurance industry. Chinese Business Community (07),114-115.

[3] Hu Hongwei, Duan Xiao-Le.(2023). Research on the business model of Life insurance and commercial long-term care insurance liability transformation. China's medical insurance (12), 108-117.

[4] Sun Jing, Yang Yan, Guan Sisi, Zhang Muzhan, Ren Lingzhi. Implementation of flood insurance system in typical countries and enlightenment from experience. Research on Water Conservancy Development 1-8.

[5] He Xiao-wei, Tong Jin-Lin, Liu Yi-Xin, Cui Zi-hang.(2023). Flood disaster insurance in China: Status, dilemma and optimization. Chinese Journal of Emergency Management (11),15-30.

[6] Yang Y F. (2024). Residual value test of water quality total nitrogen calibration curve based on mathematical statistics. Yunnan Chemical Industry (01),82-87.

[7] Deng L X, Yang F & Zhao L L. (2023). Application of grey system theory to drought and flood disaster prediction. Water Resources Development and Management (04),66-70.

[8] Li Y M. (2015). Normality test method of data. Journal of Huaihua University (11),81-82.

[9] Jeffrey M. Woodridge. Introduction to Econometrics Modern View. Beijing: China Renmin University Press, 2023.05

[10] Yang B. (2023). Discussion on the application of structural strengthening technology in building construction. Tile (12), 151-153 + 156.

[11] Liu W Y. (2023). Research on the design and application of building structure reinforcement and construction technology. Engineering seismic and reinforcement reconstruction (06),180.

[12] Wang Lezhe.(2023). Analysis of structural strengthening technology in building construction. Housing Industry (10),87-89+93.

[13] Lv Xiao, Jia Guanghao, Zhang Xin, Wang Shun.(2023). Research on the mechanical behavior of prestressed concrete steel tube truss superposition soleplate. Journal of Xi 'an University of Architecture and Technology (Natural Science Edition)(06),791-800.

[14] Wang Yuji, Huang Chenyu, Shan Jia-zhan, Yu Hua and Su Jinrong. A machine learning prediction method for earthquake damage of existing buildings based on multivariate features. 1-13. Journal of building structures.

[15] Ma Weijing, Zhang Wenshun, Li Chengyi, Kou Jingwen, Wang Meimei and Xue Bing.(2023). Spatiotemporal characteristics and driving factors of allometric growth of economic scale and marginal benefit of water resources in the Yellow River Basin. Journal of Natural Resources (12),3116-3134.