Prediction model of insurance profit and loss under extreme weather based on PCA and ARIMA Models

Xinyao Ding[1, *], Jincheng Bai[2]

[1]Portland Institute, Nanjing University of Posts and Telecommunications, Nanjing, China, 210023

[2]College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, China, 210023

* Corresponding Author Email: sherry1012dxy@163.com

Abstract. Because global catastrophe insurance gap widening, insurance profit forecast is important for ensuring the operation of insurance companies under extreme weather. In this paper, on the basis of analyzing the profit structure of insurance, the dimensionality reduction capability of PCA and the flexibility of ARIMA are combined to establish a prediction model of insurance profit and loss under the extreme weather. On the basis of using PCA (Principal Component Analysis) to extract the important features of extreme weather data and to conduct correlation analysis, the influences of each factor on the expression of different principal components are clarified, and the calculation formula of the principal component is obtained. ARIMA (Autoregressive Integrated Moving Average Model) is used to fit historical catastrophe insurance data and capture trends, to predict the likely future compensation ratio in this region. This method can not only reduce the dimension of complex weather data to simplify the analysis, but also flexibly predict the risk degree of future losses of insurance companies. The results show that with the increase of the frequency of extreme weather and the maximum rainfall, the operating risk of insurance companies will rise. Finally, in order to better develop regional catastrophe insurance, it’s a good choice to carry out long-term and short-term planning according to the prediction model.

Keywords: PCA, ARIMA, Catastrophe Risks, Prediction Model.

1. Introduction

China is one of the countries with the most serious natural disasters in the world, with an average annual direct economic loss of 329.3 billion yuan in the past 10 years [1]. With the increasing frequency of extreme weather, the human and property losses caused by it are becoming more and more serious. As an important tool for dispersing natural disaster risks, insurance is indispensable for enhancing the resilience of economic and social development against risks [2]. However, it is important to realize that extreme weather can cause extraordinary effects and injuries, so it is essential to assess and predict the future coverage rates of local insurance. Anita M, Snejana G, and Michael M also pointed out that natural disasters are a significant financial burden for the insurance industry, which can affect the profitability and financial stability of companies [3].

Domestic and foreign scholars have carried out extensive research on such issues. Patrick L used the learning vector quantization method and backpropagation method to predict the ability of insurance companies to reimburse users [4]. This method can’t avoid the complexity and sensitivity of its calculation, when the data is relatively large, the calculation is limited. Hou Xuhua and Peng Juan considered the financial risk early warning of insurance companies based on the report data of insurance companies based on the entropy method and the efficiency coefficient method [5]. The entropy weight method does not reduce the dimensionality of the data, so the processing of high-dimensional data may be affected by the dimensionality, resulting in errors. Moreover, the data sources in this paper are not rigorous enough. Mikosch used the extreme value theory to discuss the bankruptcy probability and reinsurance of insurance companies with a large amount of compensation [6]. Extreme value theory has a high requirement on the selection of threshold value, and improper selection is easy to result in error. Samuel R, Peter J.R, and Wouter J.W.B used the logit model to estimate factors associated with the purchase of natural disaster insurance and the adoption of flood risk mitigation measures [7]. Logit model has high data requirements, and it is difficult to deal with multi-level structured data. Hou Xuhua and Xu Xiang calculated and analyzed the development level of green insurance of insurance companies by using TOPSIS comprehensive evaluation method empowered by the combination of entropy weight and AHP [8]. This model is difficult to balance subjectivity and objectivity, and AHP is greatly affected by subjective preferences, which will interfere with the results.

At present, the accuracy of traditional methods to study the profit risk of insurance companies under the influence of extreme weather is not high enough, and it is difficult to discover larger potential influencing factors from limited data [9]. In this paper, PCA and ARIMA models are combined to predict the profitability of insurance companies in areas affected by extreme weather. It is estimated that the forecast results of insurance profitability in this paper directly affect the input and pricing of catastrophe insurance, which is important for companies in dynamic adjustment, and improving economic benefits.

2. The establishment and solution of the model

PCA (Principal component analysis) is applied to combine several indexes linearly and reduce the dimensionality of data to evaluate the degree of insurance risk in different regions. Due to the irregular nature of extreme weather data, using ARIMA (Autoregressive Integrated Moving Average model) to predict future risk level changes seems to be a right method.

2.1. Insurance Profit Model

For the income and expenditure of insurance companies, the relevant impact factors are set up respectively to better reflect the impact of factors on their profitability. The operation process of an insurance company is shown in Figure 1.

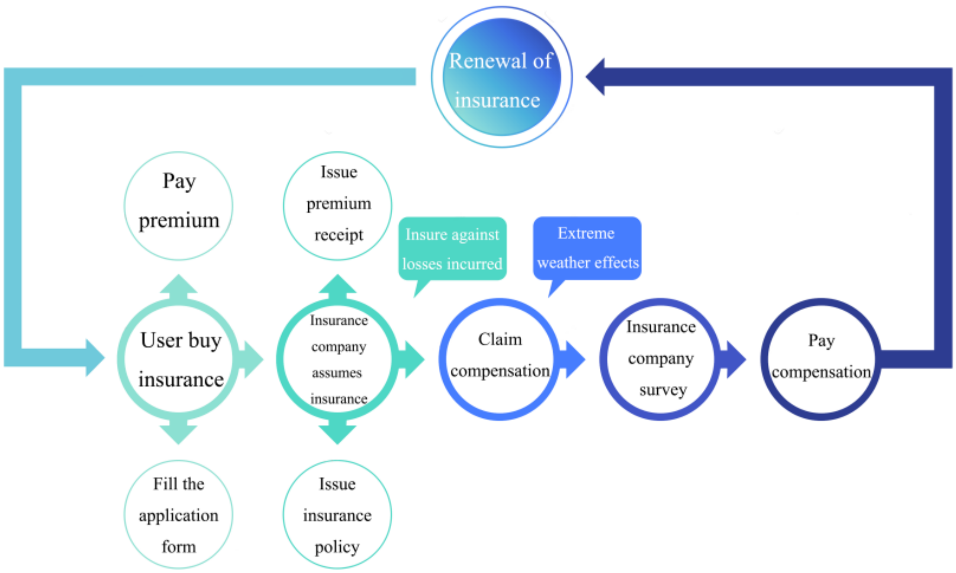

Figure 1. Operation process of insurance company

It is clearly seen from Figure 1 that the degree of profit and loss of an insurance company in a specific period of time mainly depends on the insurance income, compensation amount and compensation probability in this period.

2.1.1. Expenditure

(1) Claim probability

Claim frequency $N(t)$ usually refers to the number of times an insurance company makes a payment in a specific time period. The higher the claim frequency of within a certain period of time, the more frequent the disaster in the region.

It is stated in the essay that since the process of the claim number $N(t)$ of natural disasters is a Poisson process, and the probability value has a corresponding relationship with the historical frequency [10]:

$$P(N(t)=k)=\frac{\lambda^k e^{-\lambda}}{k!},k=0,1,2... \tag{1}$$

where, the mean of the Poisson distribution is λ .

(2) Total claims

The insurer can seek protection by transferring part of its risk exposure to the reinsurance, in order to reduce the underwriting risk [11]. Here, the variable R_r is introduced as the proportion of reinsurance, so the total amount of claims is:

$$Total=(\sum Total_i \times N(t)_i)\times(1-R_r) \tag{2}$$

$Total_i$ is the amount of claims for type i risk, because the insurance companies classify policyholders' types of insurance into different risk levels.

2.1.2. Compensation ratio

The compensation ratio R (The ratio of an insurance company's total catastrophe insurance program operating costs to the company's total catastrophe insurance premium income) is introduced to analyse whether the insurance company's income and expenditure are balanced:

R<1 means that the insurance company can make a profit.

R=1 means that insurance companies are breaking even on programs that cover disasters affected by extreme weather.

R>1 means that it isn't conducive to the future development of the insurance company.

2.2. Standardized treatment

Since the units and orders of magnitude of each influencing factor are different, the previously collected data should be standardized before data analysis.

Positive indicator processing:

$$x^{'}=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{3}$$

Negative indicator processing:

$$x^{'}=\frac{\max(x)-x}{\max(x)-\min(x)} \tag{4}$$

2.3. Principal component analysis

By using y_n to represent the n-th principal component, U_{nj} to represent the n-th component of the j-th eigenvector, as follows:

$$y_n=u_{1n}\bar{x}_1+u_{2n}\bar{x}_2+\cdots+u_{nn}\bar{x}_n \tag{5}$$

The variable b_j is introduced as the contribution rate of the principal component y_j, and calculate the cumulative contribution rate α_p:

$$b_j=\frac{\lambda_j}{\sum_{k=1}^{n}\lambda_k}(j=1,2,\ldots,n) \tag{6}$$

$$\alpha_p=\frac{\sum_{k=1}^{p}\lambda_k}{\sum_{k=1}^{n}\lambda_k}(p\le n) \tag{7}$$

Generally, the eigenvalue with cumulative contribution rate greater than 80% is the principal component, which can determine the number of principal components by cumulative contribution rate α_p.

Through the weight of each principal component, the comprehensive score of the insurance risk level in each region can be obtained. The higher the score, the higher the risk of the insurance company's investment here, the investment is not recommended.

2.4. Time Series Analysis (ARIMA)

ARIMA model is used to predict the degree of future insurance risks.

The general form of the ARIMA model is as follows:

$$Y_t = c + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \cdots + \gamma_q \varepsilon_{t-q} \tag{8}$$

The specific modeling process of the ARIMA model can be divided into the following several steps:

Firstly, the stationarity of time series must be tested. Secondly, determining the order of the model. Then, doing the parameter estimation and diagnostic test. Checking the validity of the model and the significance of the model parameters are necessary. If they can pass the test, it means that the model setting is basically correct. Otherwise, the form of the model and this step must be re-determined and repeated. Finally, the prediction is made by the established ARIMA model.

3. Results

3.1. Analysis of experimental results

According to the data of Ministry of Emergency Management of the People's Republic of China, the data of Zhengzhou from 2003 to 2011 is integrated and processed to fit the predicted data for 2012-2022.

The importance of the hidden variables in each principal component can be analyzed by the load matrix heat map, as shown in the Figure 2.

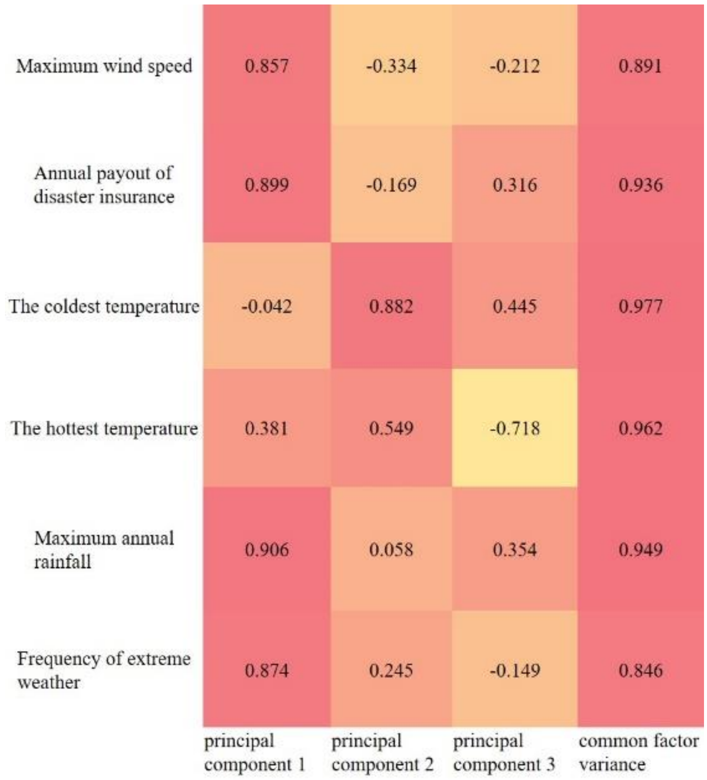

Figure 2. Correlation coefficient heat map of Zhengzhou

It can be easily discovered from Figure 2 that the frequency of extreme weather, the maximum rainfall and the annual disaster insurance payout are the three main factors that determine whether local catastrophe insurance is profitable, because the darker the heat map color, the greater the correlation.

PCA converts the original variables (the number of extreme weather, the highest annual rainfall, the highest annual temperature, the lowest annual temperature, the economic loss, and the maximum wind speed) into three comprehensive indicators, namely the principal component, through dimensionality reduction technology.

After calculation, the component matrix table is obtained, indicating the factor score coefficient (principal component load) contained in each component, as shown in the Table 1.

Table 1. Factor score coefficient

Name	Principal Component 1	Principal Component 2	Principal Component 3
Frequency of extreme weather	0.267	0.191	-0.148
Annual maximum rainfall	0.277	0.045	0.352
Annual hottest temperature	0.116	0.428	-0.714
Annual coldest temperature	-0.013	0.688	0.443
Economic losses	0.275	-0.132	0.314
Maximum wind speed	0.262	-0.26	-0.211

As can be seen from Table 1, principal component 1 mainly explains information including the number of extreme weather events, the highest annual rainfall, economic losses, and maximum wind speed. These factors all have high loads, indicating that they have a strong positive correlation on principal component 1, which together reflect the overall impact of extreme weather events and their economic consequences. In particular, the positive loads of maximum annual rainfall and the number of extreme weather events are similar, indicating that both are of similar importance in describing the frequency and intensity of extreme climate events.

Principal component 2 shows the difference of temperature extremes represented by annual minimum and annual maximum temperature. This may reflect the degree of temperature fluctuation in climate change, which has a potential impact on the regional ecological environment and the economic benefits of insurance companies, so it needs to be paid attention to.

Principal component 3 not only reflects the inverse relationship between annual maximum temperature and economic loss, but also presents positive and negative correlation between annual maximum and annual minimum temperature. This indicates that for this region, extreme heat events have a large negative economic impact.

It can be clearly concluded that the calculation formulas of the three principal components.

For principal component 1:

$$F1 = 0.267 * Frequency\ of\ extreme\ weather + 0.277 * Annual\ maximum\ rainfall + 0.116 * Annual\ hottest\ temperature - 0.013 * Annual\ coldest\ temperature + 0.275 * Economic\ losses + 0.262 * Maximum\ wind\ speed \tag{9}$$

For principal component 2:

$$F2 = 0.191 * Frequency\ of\ extreme\ weather + 0.045 * Annual\ maximum\ rainfall + 0.428 * Annual\ hottest\ temperature + 0.688 * Annual\ coldest\ temperature - 0.132 * Economic\ losses - 0.26 * Maximum\ wind\ speed \tag{10}$$

For principal component 3:

$$F3 = -0.148 * Frequency\ of\ extreme\ weather + 0.352 * Annual\ maximum\ rainfall - 0.714 * Annual\ hottest\ temperature + \tag{11}$$

$$0.443 * Annual\ coldest\ temperature + 0.314 *\ Economic\ losses - 0.211 * Maximum\ wind\ speed$$

It can be obtained from above:

$$F = (0.546/0.927) * F1 + (0.214/0.927) * F2 + (0.168/0.927) * F3 \quad (12)$$

Then the risk score of the region can be obtained to judge the degree of risk of the region being affected by extreme weather.

The compensation ratio was introduced to analyse whether the insurance company's income and expenditure are balanced.

The fitting value of the data in recent years is basically consistent with the true value, as shown in Figure 3.

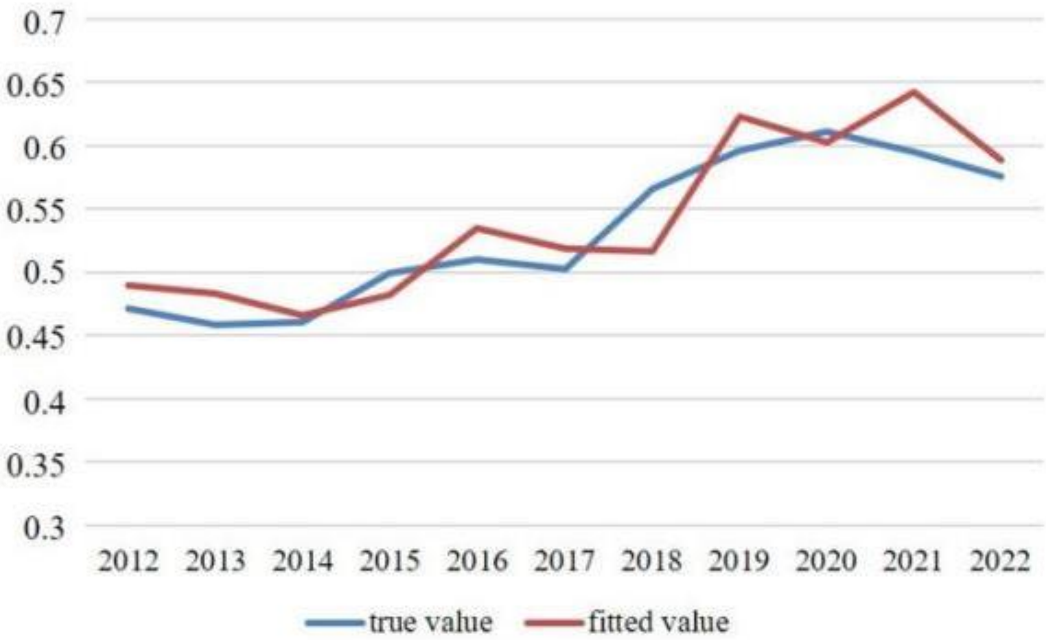

Figure 3. Time series fitting graph of Zhengzhou

The Figure 3 shows forecast data of insurance compensation ratio for Zhengzhou City from 2012 to 2022, and processing of disaster and insurance data based on this area from 2003 to 2011.The fitting result of the model is good.

As can be seen from the Figure 3, the compensation ratio of insurance companies in Zhengzhou is low, and this area belongs to the medium risk area. At present, the probability of natural disasters in Zhengzhou is not particularly high, but extreme time may also occur in the future. Insurance in extreme weather can be profitable in Zhengzhou.

4. Conclusions

The wide range and high loss of extreme weather impact on the profit protection of insurance companies. In this paper, the dimensionality reduction capability of PCA and the flexibility of ARIMA are combined to establish an insurance profit forecasting model under the influence of extreme weather. The experimental results show that the adverse effects of catastrophe insurance on insurance companies increase with the increase of risk intensity and frequency. Predicting the future insurance profit and loss situation of the region is of great significance to increasing insurance companies' ability to withstand the impact of extreme weather. Insurance companies can reduce the risk of future operations by dynamically adjusting premiums through forecasting, developing more effective claims policies and planning reserve funds. The prediction model can better develop regional catastrophe insurance and guide the insurance company to plan whether to carry out catastrophe insurance program in this region at different times.

References

[1] Dai Yuxuan. A Preliminary Study on China's Diversified Catastrophe Insurance Model [J]. Journal of Disaster Prevention and Mitigation Engineering, 2024, 44(03):745-750.

[2] Zhang T, Ma X, Wang W J. Natural Disaster Insurance in China: Model Reference, Development Challenges and Suggestions[J]. International Journal of Natural Resource Ecology and Management, 2022, 7(2):93-98.

[3] Anita M, Snejana G, Michael M. Impact of Natural Disasters on the Value of (Re)Insurance Companies[J]. Zeitschrift für die gesamte Versicherungswissenschaft, 2023, 112(4):337-368.

[4] Patrick L, Brockett, Linda L et al. A comparison of neural network statistical methods and variable choice for life insurers financial distress prediction [J]. Journal of Risk and Insurance, 2006, 73(3):397-419.

[5] Hou Xuhua, Peng Juan. Research on Financial Risk Early Warning of the Internet Insurance Company Based on Entropy Method and Efficacy Coefficient Method [J]. The theory and practice of finance and economics, 2019, 40(5):7.

[6] Mikosch T. Heavy-tailed modelling in insurance [J]. Stochastic Models, 1997, 13(4):799-815

[7] Samuel R, Peter J.R, Wouter J.W.B. Insights into the complementarity of natural disaster insurance purc hases and risk reduction behavior.[J].Risk analysis:an official publication of the Society for Risk Analys is, 2023, 44(1):141-154.

[8] Hou Xuhua, Xu Xiang. Research on the development evaluation and promotion path of green insurance for insurance companies under the goal of carbon peaking and carbon neutrality——Take Ping An Insurance Company of China for example[J]. Shanghai Insurance, 2023, (12):34-42.

[9] Sandow S, Zhou X. Data-efficient model building for financial applications: a semi-supervised learning approach[J]. The Journal of Risk Finance, 2007, 8(2):133—155.

[10] Huang Yifan, Meng Shengwang. The Design and Pricing of Earthquake Index Insurance in China [J]. statistical research, 2022, 39(04):108-121.

[11] Chao W .Valuing Multirisk Catastrophe Reinsurance Based on the Cox–Ingersoll–Ross (CIR) Model[J]. Discrete Dynamics in Nature and Society, 2021, 2021.