Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 13, No. 2, 2024

Research on Extreme Weather Risk Assessment and Underwriting Decision Based on PCA-AHP Algorithm and ARIMA Modeling

Meijia Guo *, Shurui Zhao

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, China
* **Corresponding author**: Meijia Guo (Email: 3439639058@qq.com)

Abstract: In this study, the disaster frequency prediction model and underwriting risk assessment model based on ARIMA are established by combining principal component analysis-hierarchical analysis method in China and the United States. Considering the personal factors of owners, the weights of subjective factors are incorporated into the risk assessment index system. A pricing model is developed through nonlinear optimal programming to help insurance companies decide underwriting strategies under extreme weather conditions. This study provides insurance companies with an effective decision-making tool to develop appropriate insurance policies in high-risk areas and to achieve a balance between benefits and costs in order to enhance their ability to cope with the challenges posed by extreme weather events.

Keywords: ARIMA; PCA-AHP Algorithm; Nonlinear Programming.

1. Introduction

In the context of global climate change, the frequent occurrence of extreme weather events poses a great challenge to the insurance, real estate, and tourism industries [1]. To address this challenge, this study takes China and the United States as examples and adopts principal component analysis-hierarchical analysis combined with ARIMA model to establish a disaster frequency prediction model and an underwriting risk assessment model. In addition, the owner's personal factors are considered and included in the risk assessment index system. Finally, nonlinear optimal programming is introduced to construct a pricing model to help insurers develop underwriting strategies under extreme weather conditions. The study aims to provide insurers with an effective decision-making tool to achieve intelligent underwriting decisions in high-risk areas, optimize insurance policy allocation, ensure operational profitability, and effectively deal with potential future claim costs. The innovation of this methodology lies in the integrated use of multiple models to provide comprehensive and reliable support to the insurance industry in the face of extreme weather risks, thus contributing to the sustainable development of the industry and the enhancement of risk management capabilities.

2. ARIMA-based modeling of extreme weather forecasts

The ARIMA model utilizes autocorrelation and differencing of time series data to analyze and predict subsequent outcomes of the data through curve fitting and parameter estimation [2]. The specific steps are as follows [3].

Step 1: Acquire the time series data and analyze the time series plot.

Step 2: Assess the seasonality of the time series using the Augmented Dickey-Fuller (ADF) test. The ADF test is a widely used unit root testing method, particularly apt for sequences with higher order or lagged correlation.

For any AR (m) process：

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_m X_{t-m} + \varepsilon_t \quad (1)$$

If the autoregressive coefficient equation: $1-\emptyset_1 L-\emptyset_2 L^2-\cdots-\emptyset_m L^m=0$. If there is a unit root, then $\emptyset_1+\emptyset_2+\cdots+\emptyset_m=1$.

The AR (m) process is deformed:

$$\Delta X_t = \delta X_{t-1} - \beta_1 \Delta X_{t-1} - \cdots - \beta_m \Delta X_{t-m} + \varepsilon_t \quad (2)$$

In the formula: $\delta = \emptyset_1 + \emptyset_2 + \cdots + \emptyset_m - 1$.

If the sequence $\{X_t\}$ is stationary, the reverse is true. Therefore, the null hypothesis and the alternative hypothesis of the ADF test are:

$$H_0: \delta = 0 (X_t \text{nonstationary}), H_1: \delta < 0 (X_t \text{stationary}) \quad (3)$$

Construct ADF statistics:$\tau_\delta = \frac{\hat{\delta}}{S_{\hat{\delta}}}$; $S_{\hat{\delta}}$ is the standard error of the parameter estimate $\hat{\delta}$. $S_{\hat{\delta}}$ is the standard error of the parameter estimate? At a significance level of α,The null hypothesis is rejected, and the time series is considered stationary if the corresponding p-value is less than α; otherwise, if the p-value is greater than or equal to α, the null hypothesis cannot be rejected and it indicates that the time series has a unit root, meaning it is non-stationary.

Step 3: Determine the appropriate differencing order, denoted as d. Non-stationary time series that do not pass the ADF test should be iteratively difference by d-order until achieving data stability.

Step4: Involves examining the autocorrelation (ACF) plot and partial autocorrelation (PACF) plot of a stationary time series in order to determine the appropriate values for p and d, and subsequently fitting the ARIMA (p, d, q) model.

Step 5: Model diagnosis and prediction.

Combined with the steps of the ARIMA model solution, the time series of the number of disasters caused by extreme weather in the two countries is plotted (Figure 1), and the ADF test is carried out to determine that the number of disasters in the two countries is a non-stationary time series, as shown in Table. 1, in the absence of a first-order differencing of the data between the $p(p > \alpha)$ it is not possible to reject the original hypothesis, the existence of a unit root, and the time series is non-stationary, in the aftermath of a first-order differencing of the $p(p < \alpha)$, it is rejected the original hypothesis, the absence of a unit root, and the time series is stationary, and the order of the differencing is determined as a result of this, the order of the differencing of the $d(d = 1)$.

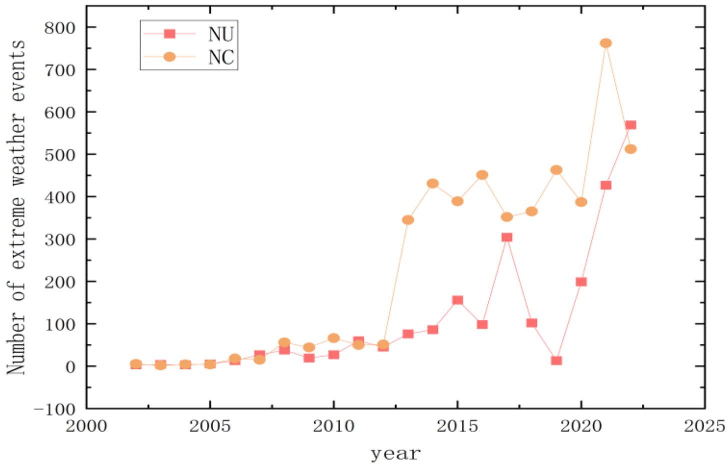

Figure 1. Sequence diagrams

Table.1 ADF test results

	United States	China
Before	0.8941	0.5695
After	0.001	0.001

Plotting the ACF and PACF plots of the smooth time series after first-order differencing for China and the U.S. (Figure 2, Figure 3) determines that the prediction models for the number of extreme weather events in the two countries take the values $q(q = 0)$,$p(p = 10)$. The ACF plots for the two countries show that the autocorrelation coefficients are within the confidence intervals, which is a white noise series.

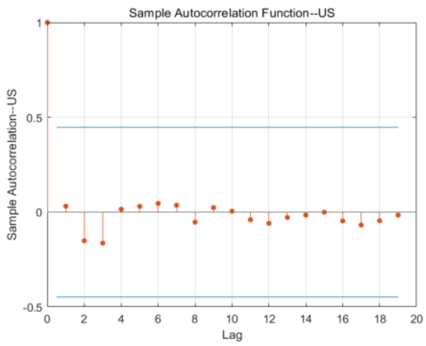

Figure 2. U.S. ACF test

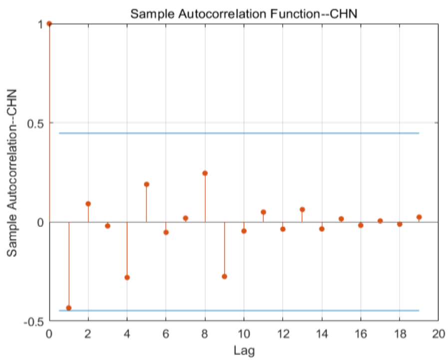

Figure 3. CHN ACF test

Therefore, ARIMA (0,1,10) model was chosen to fit and predict the occurrence of extreme weather, and the prediction results are shown below (Figure 4 and Figure 5).

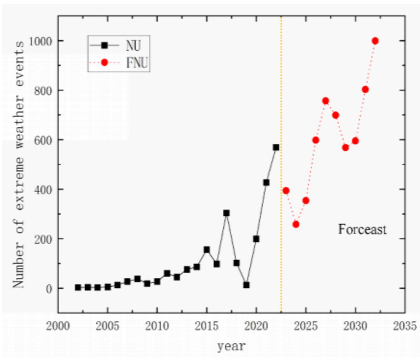

Figure 4. U.S. Prediction of the number of damages caused

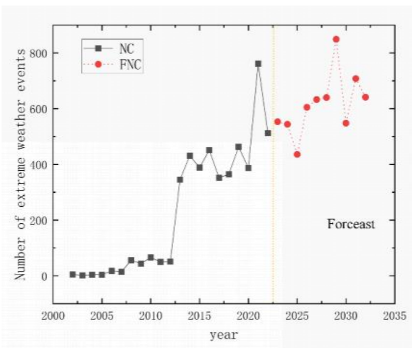

Figure 5. CHN Prediction of the number of damages caused

The forecast graphic shows that the number of extreme weather initiation occurrences in China and the U.S. will continue to fluctuate and grow in the future.

3. Establishment of a property insurance underwriting risk indicator system

Review the literature and information to establish the relevant indicators of the relevant risk evaluation system [4], and the indicators are selected as follows Figure 6.

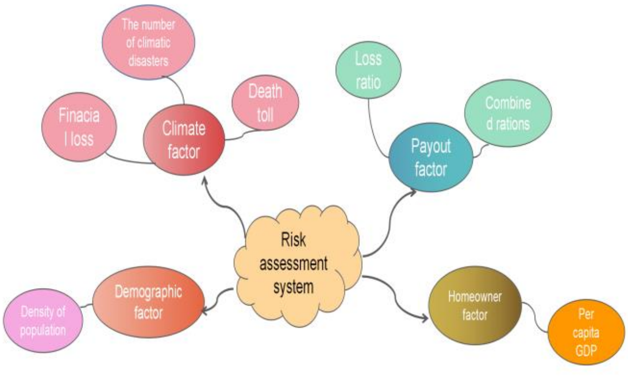

Figure 6. Risk assessment indicator system

In terms of claim factors, a higher insurance payout ratio and comprehensive cost ratio in previous data indicate a greater probability of claims in the case of frequent extreme weather events in the sample data area. This also increases the risk for insurance companies underwriting property insurance. Among climate factors, indicators such as the number of extreme weather deaths, occurrences, and economic losses are utilized to quantify the severity of extreme weather. A higher index value indicates myasthenia greater possibility of occurrence risk. Regarding demographic factors, higher population density coincides with an increased risk assessment. As for rude owner factors, it is believed that owners with higher subjective economic levels are less likely to receive compensation from insurance companies.

4. Property risk assessment model based on principal component-hierarchical analysis method

The determination of the weights of hierarchical analysis method is strongly influenced by the subjective judgment in the judgment matrix, therefore, the variance contribution rate of each indicator can be selected as the objective weight by principal component analysis and the subjective weight of hierarchical analysis method can be combined to establish the risk assessment model [5].

4.1. Hierarchical analysis to determine subjective weights

The steps of the hierarchical analysis method are as follows:

Step 1: Compare the degree of importance of the indicators under the same level and construct the judgment matrix according to the scale of 1-9. Construct the judgment matrix A, the a_{ij} is the importance of the indicator A_i relative to A_j, and the method of taking its value is shown in Table. 2.

Step 2: Eigenvalue method for weights w_{ej}.

$$w_{ej} = \frac{e_j}{\sum_{j=1}^{n} e_j}, j = 1,2,\cdots,n \tag{4}$$

The e_j is the eigenvector corresponding to the largest eigenvalue C_{max} in the judgment matrix A.

Step 3: Geometric averaging to find the weights w_{gj}.

$$w_{gj} = \frac{\left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}}}{\sum_{k=1}^{n}\left(\prod_{j=1}^{n} a_{kj}^{\frac{1}{n}}\right)}, j = 1,2,\cdots,n \tag{5}$$

Step 4: Arithmetic average method to find the weights w_{aj}.

$$w_{aj} = \frac{1}{n}\sum_{j=1}^{n}\frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}, j = 1,2,\cdots,n \tag{6}$$

Step 5: Three methods to find the average weights w_j.

$$w_j = \frac{1}{3}\left(w_{ej} + w_{gj} + w_{aj}\right) \quad j = 1,2,\cdots,n \tag{7}$$

Step 6: The data samples are de-scaled and normalized to obtain the correlation coefficient matrix S_{ij} ($i, j = 1, 2, 3 \cdots\cdots n$).

The relative evaluation value of the ith object is,

$$x_i = \bigcup_{i=1}^{n} V_{ij} W_J \tag{8}$$

Step 7: The consistency test is performed on the judgment matrix and the formula is calculated as follows:

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{9}$$

$$CR = \frac{CI}{RI} \tag{10}$$

The RI is the average random consistency index, and the consistency of the judgment matrix A can be considered acceptable if $CR < 0.1$.

Table.2 Meaning of scales at all levels

Scale value	Implication
1	The former is as important as the latter
3	The former is slightly more important than the latter
5	The former is clearly more important than the latter
7	The former is more strongly important than the latter
9	The former is more important than the latter.
2, 4, 6, 8	Intermediate value of the above proximity judgment
Inverse of number	The ratio of element-to-element importance is a_{ij}, and the ratio of element-to-element importance is inverse

4.2. Principal component analysis to determine subjective weights

Principal Component Analysis steps:

Step 1: Normalize the raw data x_i to create a data matrix S_{ij}. The standardized processing formula is given below:

$$x_i^* = \frac{x_i - E(x_i)}{\sqrt{Var(x_i)}}, i = 1,2,\cdots,n \tag{11}$$

Step 2: Calculate the correlation coefficient matrix R.

Step 3: Calculate eigenvalues m_j, eigenvectors e_j.

Step 4: Calculate the information contribution rate a_j and cumulative contribution rate of the eigenvalues a_{sj}.

Step 5: Select P （$P < 7$)principal components according to whether the cumulative contribution rate is greater than 80% or not.

4.3. Principal component-hierarchy analysis

Through the combination of both principal components and hierarchical analysis, the risk factor score can be calculated by subjective-objective combination assignment based on the realization of indicator dimensionality reduction.

The incorporation steps are specified below:

Step 1: Hierarchical analysis method, construct judgment matrix A, calculate and obtain weight vector W_j.

Step 2: Principal component analysis, standardize the original data matrix S_{ij}, calculate the correlation coefficient matrix R, calculate the eigenvalues m_j and eigenvectors e_j, calculate the principal component contribution rate a_j and cumulative contribution rate a_{sj}.

Step 3: The standardized data matrix S_{ij}, was weighted with hierarchical analysis weights W_j, and the principal component scores were calculated.

5. Underwriting risk assessment model solving

5.1. Principal Component Analysis Solving

Combined with the relevant steps of principal component analysis, the MATLAB software was used to carry out principal component analysis on the four secondary indicators and seven tertiary indicators in the risk index evaluation system, and the heat map solved and plotted to derive the correlation coefficient matrix is shown in Figure 7 and Figure 8.

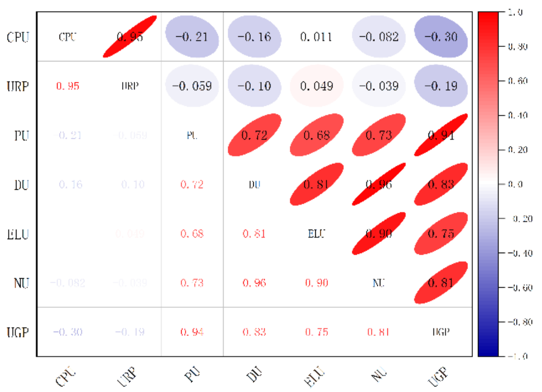

Figure 7. U.S. Correlation coefficient chart of each indicator

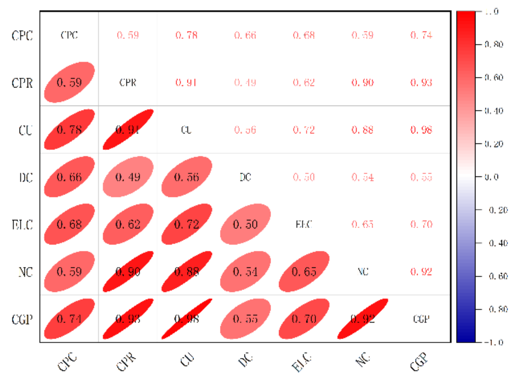

Figure 8. CHN Correlation coefficient chart of each indicator

5.2. Hierarchical analysis for solving

Through the comparison of the importance of the seven evaluation indicators to construct a judgment matrix and consistency test, the use of MATLAB was selected three methods to solve the judgment matrix of the weight of the seven indicators, the results are shown in the Table 3 below.

Table.3 Weights of each indicator

	D1	D2	D3	D4	D5	D6	D7
Eigenvalue (math.)	0.0596	0.0932	0.0271	0.4452	0.137	0.212	0.0259
Geometric mean method	0.0602	0.0968	0.0265	0.4252	0.1453	0.2187	0.0273
Arithmetic average method	0.0699	0.1051	0.0321	0.4091	0.1453	0.2104	0.0281
The average weight of the three methods	0.0632	0.0984	0.0286	0.4265	0.1425	0.2137	0.0271

The consistency ratio $CI(CI = 0.1297)$of the consistency indexes $CR(CR = 0.07321 < 0.1)$ indicates that the judgment matrix meets the requirements of consistency test, and the results of the weighting solution are reliable.

6. Optimal underwriting decision model based on nonlinear optimal programming

To model the decision to underwrite or not to underwrite property insurance from the perspective of an insurance company, the insurance company will decide whether or not to underwrite property insurance for homeowners in a given area based on a risk assessment of the area, and therefore build a nonlinear programming model using maximization of the insurance proceeds as the objective function to determine the probability of claiming for the key factor, denoted as P [6].

6.1. Fitting of claim probability *P*

It is generally accepted that when the insurance underwriting risk increases, the claim probability will also increase. When the insurance underwriting probability increases to a certain level, the claim probability is usually stable, and the claim probability ranges from 0 to 1. Therefore, the Logistic population retarded growth model can fit its change trend well. The logit model is established as follows.

$$logit(p) = ln\frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots\cdots \beta_m x_m \quad (12)$$

$$\hat{p} = \frac{e^x}{1+e^x} \quad (13)$$

6.2. Establishment of optimal underwriting decision model

(1) Objective function determination of operating profit S of insurance company

The operating profit of insurance company is taken as the objective function to make the decision. In order to ensure the time comparability of the monetary value during the insurance period, it is necessary to take the coefficient of the final value of the annuity into account to calculate the time value of the currency. The expression of the operating profit of an insurance company is established as follows:

$$S = \sum\left[a \cdot t \cdot (1-p) \cdot FVIFA_{i,n} - b \cdot t \cdot p \cdot FVIFA_{i,n}\right] \quad (14)$$

$$FVIFA_{i,n} = \sum_{t=1}^{n}(1+i)^{t-1} \quad (15)$$

a represents the premium, b represents the insured amount, t represents the insured years, $FVIFA_{i,n}$ represents the final value of the annuity coefficient, r represents the annual interest rate.

(2) Determination of constraints

In order to ensure the normal operation of an insurance company, it is generally required that the total insured amount within a certain period of time does not exceed the total premium of the company, and the insured period, premium and insured amount are all within the upper and lower limits of the property insurance of the insurance company.

$$\begin{cases} \sum b \le \sum a \\ lb_a \le a \le ub_a \\ lb_b \le b \le ub_b \end{cases} \quad (16)$$

(3) Optimal property insurance policy decision model based on nonlinear programming

$$\max(s) = \sum\left[a \cdot t \cdot (1-p) \cdot FVIFA_{i,n} - b \cdot t \cdot p \cdot FVIFA_{i,n}\right] \quad (17)$$

$$s.t.\begin{cases} p = p(I) \\ \sum b \le \sum a \\ lb_a \le a \le ub_a \\ lb_b \le b \le ub_b \end{cases} \quad (18)$$

In summary, the optimal decision-making model of insurance premium, insurance amount and insurance life are given.

6.3. The basic decision model of whether to cover or not

It is generally believed that if the company's operating profit is greater than 0, it can undertake property insurance in the place, that is $S = \sum\left[a \cdot t \cdot (1-p) \cdot FVIFA_{i,n} - b \cdot t \cdot p \cdot FVIFA_{i,n}\right] > 0$, Assume that there is an insurance plan with an annual insurance cost of $700, an insurance term of 5 years, an insurance amount of $1,000, and an annual bank interest rate of 12%, and the coefficient of the final annuity value of the policy is 6.323. Only when the claim probability of the policyholder in the region is less than 0.4118 can this transaction be profitable.

Therefore, combined in practical applications, the specific operations are as follows:

The risk assessment model evaluates the claim risk of residents in this area.

Bring the risk coefficient into the logit model to solve the claim probability of residents in the region.

When the premium, contract term and coverage range of the insurance package are known, genetic algorithm is used to solve the optimal underwriting decision model of nonlinear programming to determine whether to carry out underwriting.

7. Conclusions

In this study, we developed a disaster frequency prediction model and an underwriting risk assessment model by combining principal component analysis-hierarchical analysis and ARIMA model, which provides an effective decision-making tool for insurance companies. By incorporating homeowners' personal factors into the risk assessment index system and building a pricing model using nonlinear optimal programming, we were able to help insurers make informed underwriting decisions under extreme weather conditions. The results show that the weight assigned to the owner's subjective factors in the risk score is 0.0271, which plays a key role in pricing and underwriting decisions. Our model can help insurers choose the best underwriting options, including premium, period of insurance, and coverage amount, to ensure operational profitability and effectively manage potential future claim costs. The application of these methods provides reliable support to the insurance industry in dealing with extreme weather risks and helps improve the industry's risk management and promote sustainable development.

References

[1] Gasper R, Blohm A, Ruth M. Social and economic impacts of climate change on the urban environment[J]. Current opinion in environmental sustainability, 2011, 3(3): 150-157.

[2] Mohamed J. Time series modeling and forecasting of Somaliland consumer price index: a comparison of ARIMA and regression with ARIMA errors[J]. American Journal of Theoretical and Applied Statistics, 2020, 9(4): 143-53.

[3] Xue D, Hua Z. ARIMA based time series forecasting model[J]. Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), 2016, 9(2): 93-98.

[4] Khameneh A H, Taheri A, Ershadi M. Offering a framework for evaluating the performance of project risk management system[J]. Procedia-Social and Behavioral Sciences, 2016, 226: 82-90.

[5] Zhao B, Liu Q, Zhu J. Risk Assessment and Zonation of Roof Water Inrush Based on the Analytic Hierarchy Process, Principle Component Analysis, and Improved Game Theory (AHP–PCA–IGT) Method[J]. Sustainability, 2023, 15(14): 11375.

[6] Pantelous A A, Passalidou E. Optimal strategies for a non-linear premium-reserve model in a competitive insurance market[J]. Annals of Actuarial Science, 2017, 11(1): 1-19.