

Extreme Weather Based Insurance Underwriting Decision Modeling and Solutions

Haoruo Li ^{1,*}, Weiru Luo ²

¹ School of Shipping Economics and Mangement, Dalian Maritime University, Dalian, China

² School of Economics and Mangement, China Agricultural University, Beijing, China

* Corresponding author: DUMhaoruo@163.com

Abstract. This paper addresses the problem of insurance companies' underwriting decisions under extreme weather conditions. Firstly, the data were preprocessed, including identifying and processing missing data and outlier analysis. Then an underwriting decision-making model based on the extreme weather hazard model and the insurance insolvency factor model is established, and the corresponding solutions are proposed. Finally, through model validation and empirical analysis, the underwriting strategies of insurance companies under different risk levels are derived, and corresponding suggestions are given.

Keywords: Extreme weather, Insurance underwriting, Risk prediction.

1. Introduction

The frequent occurrence of extreme weather events has brought challenges to insurance companies' underwriting decisions, and how to formulate reasonable underwriting strategies under the premise of minimizing risks has become an important issue in insurance business management. This paper aims to establish an insurance underwriting decision model for extreme weather events and propose corresponding solutions to help insurance companies better cope with the risks brought by extreme weather [1-2].

2. Modeling and Solving

2.1. Model 1(a) Model establishment and solution

This study examines insurance companies' decisions to accept policies at certain periods and conditions and how insurance users affect these decisions. We used data from the China Bureau of Statistics (<https://www.stats.gov.cn/sj/>) and the Guangdong Provincial Bureau of Statistics (<https://www.gddat.cn/newGlobalWeb/#/home>) to do this. Table 1, shows some of the results after data preprocessing:

Table 1. Insurance data for some regions in China

year	region	Area affected	Drought	Flood typhoon	Hailstorms	Low temperature snow storm	Affected population	death	Economic loss
2021	Hunan	436.2	157.9	272.4	5.9	0.0	652.4	8	82.4
2021	Guangdong	76.4	31.2	17.8	1.7	25.6	100.1	2	24.1
...	...	...	...	...	...	...	...	...	...

This study identifies five extreme weather damage types: drought, hail, flooding, temperature anomalies, and high winds. Python programming tools are used to visualize the acquired data more intuitively. Refer to Figures 1 to 4 for details.

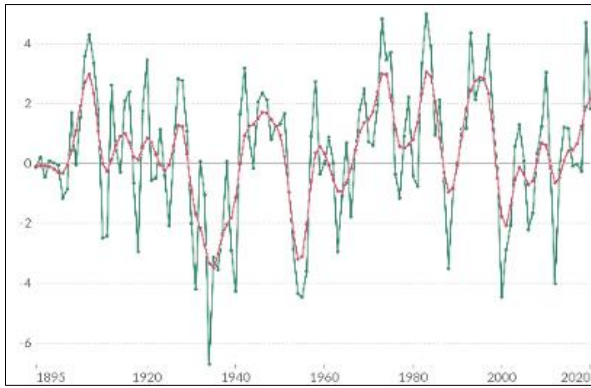


Figure 1. Drought severity index

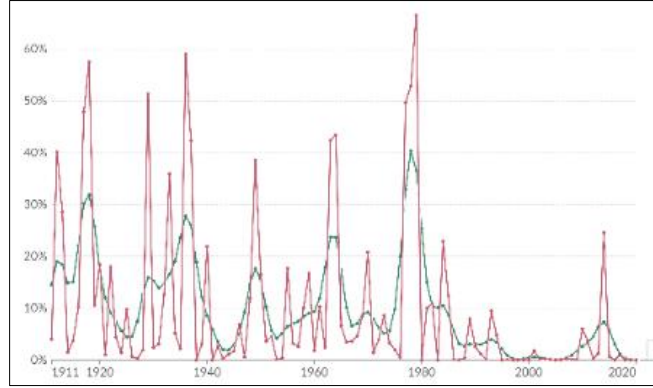


Figure 2. Abnormally low temperature

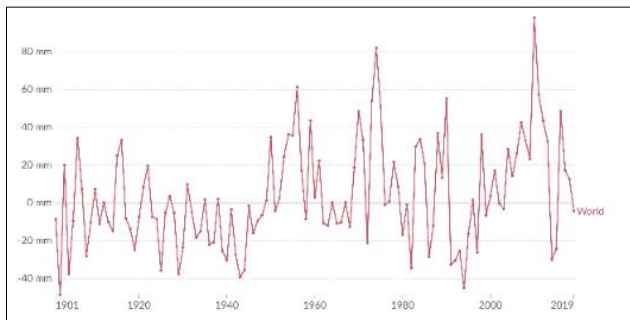


Figure 3. Abnormal precipitation

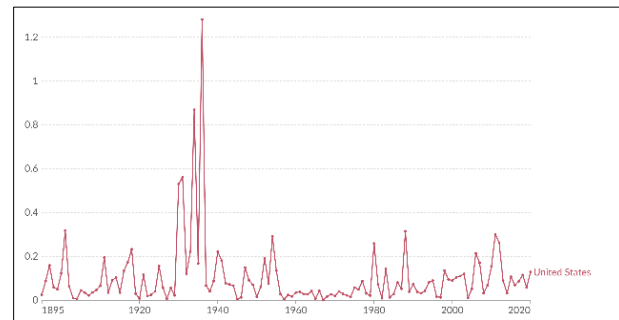


Figure 4. Abnormal hail

Step 1: Establish a disaster model

First of all, it is necessary to determine the main risk disaster standards brought by extreme weather, and the coefficients of different disaster types are denoted as, and the total extreme weather coefficients are denoted as H_j H .

(1) Drought disasters. According to the official drought early warning standards, the disaster coefficient is determined H_1 , and the level is divided into, and $i_1 = \{0.05, 0.25, 0.50, 0.80\}$ the random probability is introduced:

$$H_1 = \{rand(i_1, \sigma), 1\} \quad (1)$$

(2) Hail and snow disasters. Except for special areas, the hail disaster $i_2 = 0.7$ coefficient is determined to be the coefficient of the introduction H_2 of random rate:

$$H_2 = i_2 \cdot \min\{rand(\mu, \sigma), 1\} \quad (2)$$

(3) Flood disaster. Flood disasters are greatly affected by topography, assuming that the land water w content is ml, the precipitation is ml, l and the land water content is the next day $\omega' = \omega + L - 50$. Flood disasters affect agriculture in three stages, and the degree of disaster is closely related to flood height. The functional relationship is:

$$\begin{aligned} H_{31} &= 59.57(r + 0.04)^{0.44} - 10 \\ H_{32} &= 47.02(r + 0.25)^{0.83} - 10 \end{aligned} \quad (3)$$

(4) Cold wave and high temperature disasters. According to the official drought early warning standard, the disaster coefficient 4 is determined, and the level is divided into, and $i_4 = \{0.05, 0.25, 0.50, 0.80\}$ the random probability is introduced:

$$H_3 = \{rand(i_3, \sigma), 1\} \quad (4)$$

(5) Wind disasters. The disaster mainly caused crop loss and infrastructure damage, and it is assumed that the coefficient H_5 has an exponential relationship with wind speed v . $H_5 = 1 - \exp[-0.017(v - 10.8)]$ The total hazard factor is defined as:

$$H = \max\{H_1, H_2, H_3, H_4, H_5\} \quad (5)$$

Step 2: Establish an insurance bankruptcy factor model

Suppose the profitability process of the insurance company in year t is $U(t)$:

$$U(t) = x + ct - k \sum_{i=1}^{N(t)} Y_i \quad (6)$$

Among them, it represents the x initial capital (0 in this article), $c > 0$ represents the premium income $0 < k < 1$ ratio, represents the loss compensation ratio $N(t)$, represents the number of extreme weather occurrences in the $[0, t]$ period, $\lambda > 0$ obeys the Poisson process, and $\{Y_i, i \geq 1\}$ represents the economic loss caused by the i th disaster [3-4].

Recorded as the $T = \inf\{t : t \geq 0, U(t) < 0\}$ moment when the bankruptcy of an insurance company occurs, the probability can be expressed as:

$$\psi(x) = P\{T < \infty | U(0) = x\} \quad (7)$$

Step 3: Insurance Coverage Conditions

Suppose the starting point of the insurance premium (x_0, y_0) is set and the loss ratio is set as $\bar{V}$, so the starting rate is defined as $\bar{V} = [x_0 / y_0]$. Construct the functional relationship between the insurance and the insurance bankruptcy factor and premium

$$E = x - [I_1 \frac{x_0}{y_0} (x - x_0) + I_2 \frac{(x - x_0)^2}{2} + y_0 \cdot \bar{V} \cdot \psi(x)] \quad (8)$$

Among others $a = -\frac{I_2 \bar{V}}{2} x^2$; $b = -(\frac{I_1 \bar{V}}{x_0} - I_2 \bar{V} x_0 - 1)$; $c = I_1 \bar{V} y_0 - \frac{I_2 \bar{V} x_0}{2} - \bar{V} y_0$.

Step 4 Determination of threshold E

The functional relationship between insurance and premium is established using the extreme weather catastrophe coefficient and insurance bankruptcy factor model[5-6]. Only when E exceeds the threshold can the insurance company underwrite the income without bankruptcy. MFF is used to set the threshold, and the empirical mean transcendence function is defined for the above data (y_i) :

$$e_n(E) = \frac{\sum_{i=1}^n (y_i - E) I_{\{y_i > E\}}}{\sum_{i=1}^n I_{\{y_i > E\}}} \quad (9)$$

Step 5 Underwriting Model Solution

In this paper, we use Python to make statistics on the types of extreme weather disasters in various regions of China.

Finally, a summary forecast (regardless of the type of disaster) is made to obtain the statistical results shown in Figure 5 below.

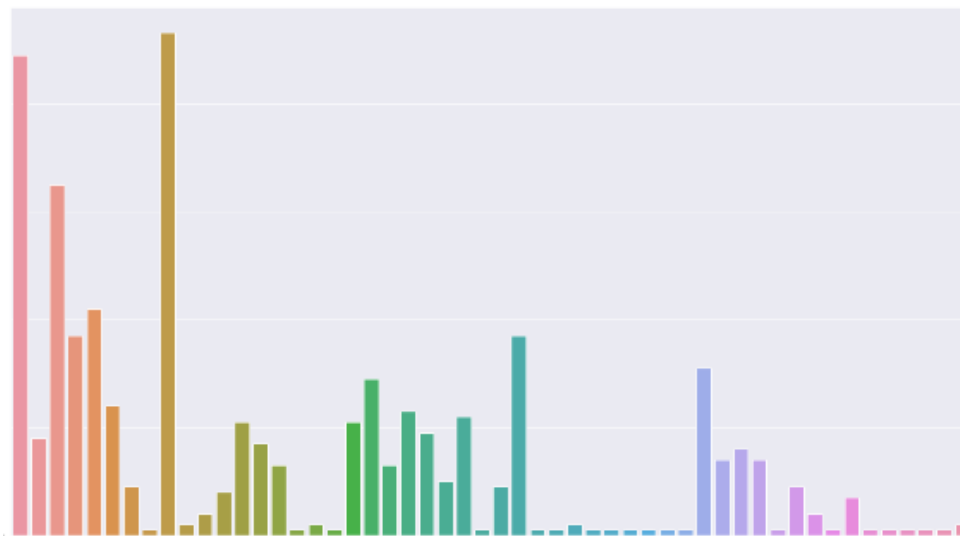


Figure 5. Statistics of extreme weather disasters (regardless of type)

Result analysis: Insurance companies tend to choose underwriting when the insurance coverage conditions are close to the upper limit $E(a)$ and exceed a certain threshold $e_n(E)$, in which case, the possibility of underwriting $\psi(x)$ is relatively low. Only under such criteria can the insurer accept and benefit from the coverage. Thus, insurance companies should underwrite based on the assumption that bankruptcy risk is low and natural disaster likelihood is low, as stated in the preceding analysis [7].

2.2. Model 1(b) Model establishment and solution

Insurers must evaluate their insolvency factor below a specified threshold before taking risk to ensure minimal risk and profitability. Insurers must foresee bankruptcy trends in the next years due to the nature of the insurance business, which covers from application to claim[8-10].

Step 1: Set the level of risk to be covered

Using the above model, we have calculated the insurance bankruptcy factors in various regions of the country in recent years. Based on these bankruptcy factors (i.e., risk levels), we can develop an underwriting strategy accordingly. For example:

When the risk level exceeds A: Choose not to cover or ask for a higher premium.

When the risk level is between A and B: Adopt a standard underwriting model strategy and adjust the underwriting fee appropriately.

When the risk level exceeds B: Implement incentives such as reduced premiums to attract underwriting.

Step 2: Build an underwriting risk prediction model

To calculate the risk level, region-specific indicators were used and grey forecasting models were used to make predictions.

Grey forecasting is a model suitable for small samples and short-term forecasting, which can provide more accurate prediction results based on limited time series data. Therefore, the grey prediction model was selected as an auxiliary tool for time series forecasting. Specifically, we used the grey GM (1,1) prediction method, where "G" represents gray, "M" represents the model, and "(1,1)" represents a first-order differential equation with one variable in the model and a first-order differential equation with $=GM(1, N)$ N 14 variables.

Suppose $X_1^{(0)} = \{x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(n)\}$ there are n observations in the sequence, and after accumulation, "a new sequence is formed $\chi^1 = \{\chi^1(1), \chi^1(2), \dots, \chi^1(n)\}$, and the corresponding model is:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = \mu \quad (10)$$

Among them, the development ash number is a , and the endogenous control ash number is the μ parameter vector to be evaluated

$$\hat{a} = \begin{pmatrix} a \\ \mu \end{pmatrix} \quad (11)$$

Use least squares to solve:

$$\hat{a} = (B^T B)^{-1} B^T Y_n \quad (12)$$

Solving the differential equation to derive the prediction model:

$$\hat{\chi}^{(1)}(k+1) = \left\{ \chi^{(0)}(1) - \mu / a \right\} e^{-ak} + \frac{\mu}{a} \quad (13)$$

Step 3: Verify and solve the model

The model must be tested for accuracy and applicability after construction. Common verification methods include relative error detection and posterior difference test. We may classify the prediction model into four accuracy classes using the posterior difference ratio C and small error probability P . Table 2 provides details.

Table 2. Indicator ranges

grade	C	P
outstanding	0.35	0.95
passing	0.50	0.80
grudging	0.65	0.70
Unqualified	0.80	0.60

After the posterior difference ratio C and small error probability P test of the gray prediction model, the C values for insurance bankruptcy factors in all regions of the country were 0.429 and 0.362, and the P values were 0.72 and 0.71, respectively, indicating that the model was qualified and could be predicted. Table 3 shows some projected findings.

Table 3. Results of the Underwriting Conditions Model

Total Damage	Total effect	Total Deaths	Mass movement	Flood	Storm	Earthquake	Drought	Extreme temperature
24	234629	128060	117	2	2	0	0	0
9	5905	16032	135	2	1	0	2	0
...	...	...	...	...	...	...	...	...

Result analysis: Combined with the above analysis and historical data, we set the threshold for the insurance bankruptcy factor (i.e., risk level) at $a=2.5$ and $b=0.8$. From the data, it can be seen that the risk level of extreme weather events such as storms and strong winds is relatively low, and the overall risk level is low, and insurance companies can consider actively obtaining underwriting opportunities. However, during the period of high incidence of disasters such as earthquakes and floods (about every 5-6 years), insurance companies should try to avoid underwriting or increase premiums or set stricter conditions when underwriting. For extreme weather risks in other years, insurers can adopt the standard insurance model described above and adjust the thresholds and underwriting costs as appropriate. The relevant visualizations are shown in Figure 6 below.

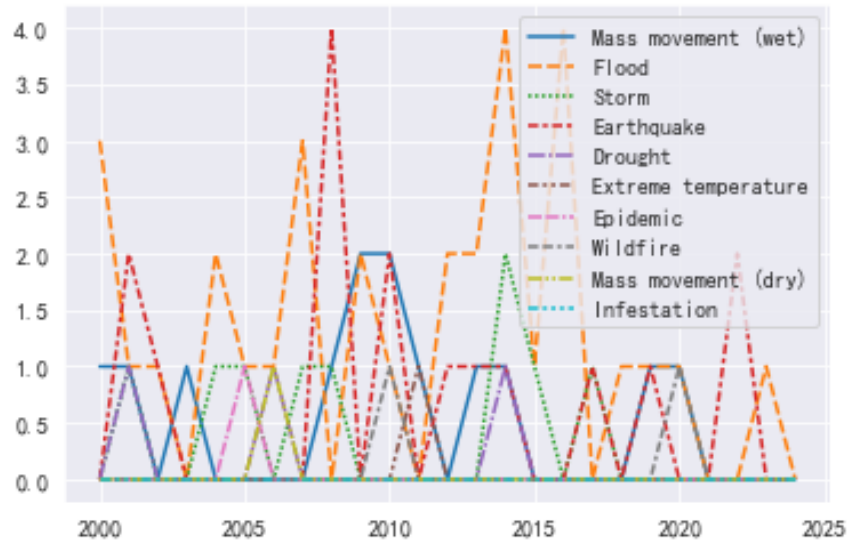


Figure 6. Risk bankruptcy factor level coefficient

2.3. Model 1(c) Model construction and solution

The main impact of owners (users) on underwriting choices is whether they change key indicators in the model. If flood hazard payouts significantly impact the extreme weather underwriting model and bankruptcy factor, the owner's adjustment to this statistic may affect the underwriting decision.

Step 1: Model creation and SHAP calculation

For the trained $f_x(S) = E[f(x)|x_S]$ model, the definition, here $f(x)$ is the output of the model, which is S consistent with the meaning in Shapley Value (a subset of the output feature set), so it is S in Shapley Value $f_x(S) \vee(S)$. Given a sample, how do you get the model's interpretation of the sample? $E[f(x)|x_1, x_2, x_3, x_4]$

Suppose the value of the features in the sample is $\{x_1 = a_1, x_2 = a_2, x_3 = a_3, x_4 = a_4\}$, there is: $S = \emptyset$ can be extrapolated, that is, the sample-independent expectation of the model prediction value $\psi_0 = f_x(\emptyset) = E[f(x)]$, and the average approximation of the model prediction value of the training sample. This value is not sample dependent.

$S = \{x_1\}$ It is possible to deduce $\psi_1 = f_x(\{x_1\}) - f_x(\emptyset) = E[f(x)|x_1] - E[f(x)]$ that the estimated difference between the expected expectation of the $\{x_1 = a_1\}$ immediate prediction and the expectation of the sample is independent of this. $E[(x)|x_1]$

For $S = \{x_1, x_2\}$ the $S = \{x_1, x_2, x_3\}$ same with the same.

In the same way, it $S = \{x_1, x_2, x_3, x_4\}$ is possible to launch:

$$\psi_4 = f_x(\{x_1, x_2, x_3, x_4\}) - f_x(\{x_1, x_2, x_3\}) = E[f(x|x_1, x_2, x_3, x_4)] - E[f(x|x_1, x_2, x_3)] \quad (14)$$

Note that this is just one of the sorts. Just like the table where we pushed the Shapley Value, in order to extrapolate to the general form, we need to apply the formula for calculating the Shapley Value to deduce the explanatory formula:

$$\psi_i = \frac{1}{p!} \sum_{S \subseteq \{x_1, \dots, x_p\} \setminus \{x_i\}} |S|!(p-|S|-1)!(f_x(S \cup \{x_i\}) - f_x(S)) \quad (15)$$

Where is the p number of total features S , a subtable of the shapley value, and f_x the mean

prediction of the alliance, which is the example in the above example, $E[f(x|x_1, x_2, x_3, x_4)] - E[f(x|x_1, x_2, x_3)]$ and the solution can obtain the SHAP value.

Step 2: Model solving

The local correctness, missingness, and consistency of SHAP are considered when solving the value in Python.

Results: The SHAP value graph study shows that measures like "Total Deaths", "Magnitude", and "No Homeless" have a higher impact on the insurance model. Property owners' efforts to reduce life and property risks, such as disaster preparedness training, strengthening residential structures, and reducing travel in extreme weather, lower the bankruptcy factor of the insurance model, lowering the risk threshold underwritten by the insurance.

3. Conclusion

By establishing an underwriting decision-making model based on the extreme weather hazard model and the insurance insolvency factor model, this paper provides an effective underwriting decision-making method for insurance companies. Meanwhile, the accuracy and feasibility of the model are verified through validation and empirical analysis of the model, which provides a basis for insurance companies to formulate reasonable underwriting strategies under different risk levels.

References

- [1] Subedi LRegmi M C, and Giri Y. Assessment of Community Based Health Insurance in Sunsari District. Kathmandu University medical journal (KUMJ) 16. 61 (2018): 53 - 59.
- [2] Wang Z, Zhang W, Wu T, et al. Time series models in prediction of severe fever with thrombocytopenia syndrome cases in Shandong province, China [J]. Infectious Disease Modelling, 2024, 9 (1): 224 - 233.
- [3] Naethe P, Sanctis D A, Burkart A, et al. Towards a standardized, ground-based network of hyperspectral measurements: Combining time series from autonomous field spectrometers with Sentinel-2 [J]. Remote Sensing of Environment, 2024, 303114013-.
- [4] Römogens N, Abbassi A, Jonscher C, et al. On using autoencoders with non-standardized time series data for damage localization [J]. Engineering Structures, 2024, 303117570-.
- [5] Bhavsar S, Pitchumani R, Maack J, et al. Stochastic economic dispatch of wind power under uncertainty using clustering-based extreme scenarios [J]. Electric Power Systems Research, 2024, 229110158-.
- [6] Guo H, Yang X, Tian Z, et al. DEM modeling of the dilute-to-dense transition of granular flow in silos [J]. Powder Technology, 2024, 436119472-.
- [7] David W, Balaji R, Edith Z. Long-Lead Forecasting of Runoff Season Flows in the Colorado River Basin Using a Random Forest Approach [J]. Journal of Water Resources Planning and Management, 2024, 150 (4):
- [8] Moradi E, Tavili A, Darabi H, et al. Assessing wildfire impact on Trigonella elliptica habitat using random forest modeling [J]. Journal of Environmental Management, 2024, 353120209.
- [9] Lu Xiaojiao. On the Current Situation and Response Strategies of Agricultural Insurance with Local Specialties [J]. Shanxi Agricultural Economics, 2022, (09): 184 - 186.
- [10] Zhang Peipei. Research on optimization of trademark insurance underwriting management in China [D]. Chongqing University of Technology, 2021.