Research on insured risk of insurance companies based on TOPSIS comprehensive evaluation model and regression analysis

Huichun Bao[*, #], Zhaoyi Tong[#], Yinuo Zhou[#]

School of Architecture and Urban Rural Planning, Sichuan Agricultural University, Chengdu, China, 611830

* Corresponding Author Email: 18197066625@139.com

#These authors contributed equally.

Abstract. Extreme weather events due to climate change are one of the most significant risks facing the world today, and while insurers can help society share the resulting losses, for example through premium pricing, the industry is also facing increasing challenges. To address the issue of meeting the needs of insurers for long-term health in the face of current challenges, we have developed a risk assessment model and a profitability model for insurers, with a focus on the U.S. East Coast. In the risk assessment model we are based on entropy weighting method to find the weight of TOPSIS superiority and inferiority of the solution distance method to derive the score of each evaluation object, and ultimately obtain the low-risk, medium-risk, high-risk intervals are [0,0.014]; [0.014,0.055]; [0.055,0.386]. The profit function was built using multiple regression in the profitability model: $\hat{\Pi i} = 0.725\hat{\zeta i} - 7.13 \times 10^{-6}\hat{Y i}$, and to draw the final conclusion: when an insurer is in a medium-low risk area, or a high-risk area but with a profit value greater than 0, the insurer can write in that area and take on a certain amount of risk. This conclusion will help insurers make better insurance choices in light of the current risk challenges and will benefit the long-term health of the insurance industry.

Keywords: Extreme weather events, Property insurance industry, Sustainability.

1. Introduction

Climate change is posing serious challenges to the insurance industry [1,2], which is increasingly concerned about climate change worldwide [3].

In the past five years, the only more detailed inquiry is the exploration of agricultural insurance development under the impact of climate change [4, 5], and few studies in the past will explore the underwriting of insurance companies in the context of extreme climate change, Jiang Shengzhong et al. investigated the impact of regional heterogeneity and disequilibrium on the development of insurance companies in China, and Lin Bingqing analyzed the short-term of property and casualty insurance companies based on the non-expected output of the SBM-Malmquist model. health insurance operational efficiency.

Synthesizing the previous studies, it can be found that for the impact of extreme weather caused by climate change, the shortcomings of the previous studies are mainly in the following aspects: firstly, the research sample is not reasonably selected, and most of the literature studies a certain kind of insurance industry, such as agricultural insurance, real estate industry insurance, etc [6-8], Secondly, there is a lack of empirical research, with individual literatures' points of intention being too high and too large, and there are dilemmas in real-life applications; and thirdly, there is a lack of analysis of the results, with some of the literatures only pointing out the impacts of the insurance industry in the context of the challenge, and not making substantive recommendations to improve the situation accordingly. This study enriches the research in this area by developing a model for insurance companies to determine whether they should underwrite in an area where there is an increase in a particular extreme weather event, based on the relationship between underwriting volume, customer volume, and company operations, using extreme weather as a consideration [9, 10]. The innovation of this paper is that the establishment of a comprehensive evaluation model expands

a new reference basis for the insurance industry to insure risks, unlike previous studies, which set the payout rate of insurance companies, which is seldom paid attention to by the previous researchers, as a variable to establish a regression equation, thus filling the gap in this aspect of the research.

2. Model

2.1. The Insurance Model

The TOPSIS method is a commonly used comprehensive evaluation method, the basic principle of which is that if the evaluation object is closest to the optimal solution and at the same time farthest away from the worst solution, the object is considered to be the best. It can make full use of the information of the original data, and its results can accurately reflect the gap between the evaluation schemes, the method has no strict restrictions on the data distribution and sample content, and the data calculation is simple and easy to implement.

The TOPSIS model consists of three processes:

(1) The original data matrix is normalized by unifying the indicator types: all three indicators are very small and need to be normalized to convert all indicators into very large ones.

$$m'_{ij} = max - m_{ij} \tag{1}$$

(2) The normalized matrix is normalized to eliminate the effect of the scale of each indicator and to find the optimal and the worst solutions among the finite solutions. Assuming that there are n objects to be evaluated and m evaluation indicators, the normalization matrix is composed as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{2}$$

For each element in the normalized matrix O of this forward matrix is:

$$o_{ij} = m_{ij} / \sqrt{\sum_{i=1}^{n} m_{ij}^2} \tag{3}$$

(3) Calculate the distance between each evaluation object and the optimal program and the worst program respectively, and obtain the relative proximity of each evaluation object to the optimal program as the basis for evaluating the advantages and disadvantages.

Defining the maximum value:

$$O^+ = (O_1^+, O_2^+, \cdots, O_m^+)$$
$$= (max\{o_{11}, o_{21}, \cdots, o_{n1}\}, max\{o_{12}, o_{22}, \cdots, o_{n2}\}, \cdots, max\{o_{1m}, o_{2m}, \cdots, o_{nm}\})$$

Defining the minimum value:

$$O^- = (O_1^-, O_2^-, \cdots, O_m^-)$$
$$= (min\{o_{11}, o_{21}, \cdots, o_{n1}\}, min\{o_{12}, o_{22}, \cdots, o_{n2}\}, \cdots, min\{o_{1m}, o_{2m}, \cdots, o_{nm}\})$$

Define the distance of the ith (i=1,2, …, n) evaluation object from the maximum value:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} (O_j^+ - o_{ij})^2} \tag{4}$$

Define the distance of the ith (i=1,2, …,n) evaluation object from the minimum value:

$$D_i^- = \sqrt{\sum_{j=1}^{m} (O_j^- - o_{ij})^2} \tag{5}$$

In summary, we can calculate the score for the ith (i=1,2, ..., n) evaluation object without normalization:

$$G_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{6}$$

It is clear that $0 \leq G_i \leq 1$, and the larger G_i is the smaller D_i^+ is, i.e., the closer it is to the maximum value.

2.2. The Insurance Profitability Model

The profitability model consists of two parts: the multiple linear regression model and the time series model. Multiple linear regression model is a type of regression model in statistics that describes the linear relationship between multiple independent variables and a continuous dependent variable based on the least squares method of finding the best-fitting straight line or plane. Multiple linear regression modeling allows one to explore the extent to which the independent variables affect the dependent variable and to predict the value of the dependent variable. A time series model is a statistical model used to analyze time series data and can be used to describe the inherent structure and patterns in the data, which change over time.

The model consists of the following five steps:

A dataset containing the dependent and independent variables was collected and the results of the experiments on the number of adults as well as economic losses in the Philippines over the past 15 years were substituted into a multiple regression model.

Modeled on the basis that profit is equal to the difference between the funds received and the cost of casting:

$$\Pi_i = \theta_i \zeta_i - \beta_i Y_i + \mu_i \tag{7}$$

where, Π_i is profits of insurance companies, θ_i is the pricing of insurance purchases, ζ_i is the number of persons insured, β_i is the insurance company assumption rate, Y_i is the economic losses under the impact of the storm, and μ_i is perturbation terms that are unobservable and satisfy certain conditions.

Collect cleaned time series data and ensure that the data are in chronological order.

Forecasts are made using a fitted time series model to predict values at future points in time.

The established multiple linear regression model was utilized to process the predicted independent variable data to infer the value of the dependent variable. The profitability model is then used to make predictive inferences about whether insurance companies should underwrite in areas with increased extreme weather events.

3. Results

3.1. Results of the Insurance Model

In this subsection, we use the TOPSIS method to determine the risk assessment intervals using the size of the economic loss and the number of fatalities as indicators and draw conclusions as shown in the Table 1.

Table 1. Risk Prediction Level Evaluation Form

Risk Prediction Level	Low Risk	Medium Risk	High Risk
Di	[0, 0.014]	(0.014,0.055]	(0.055,0.386]

We then used the size of the economic loss, the number of deaths, and the composite score obtained in TOPSIS for each of the 40 testers to construct the distribution of the three clusters across the three metrics after clustering, the trends in the distributions, and the relationship between each cluster and

the two metrics. Parallel coordinate plots, area plots and scatter plot matrices were drawn to visualize and analyze the data based on the available data, as shown in Figures 1, 2, and 3:

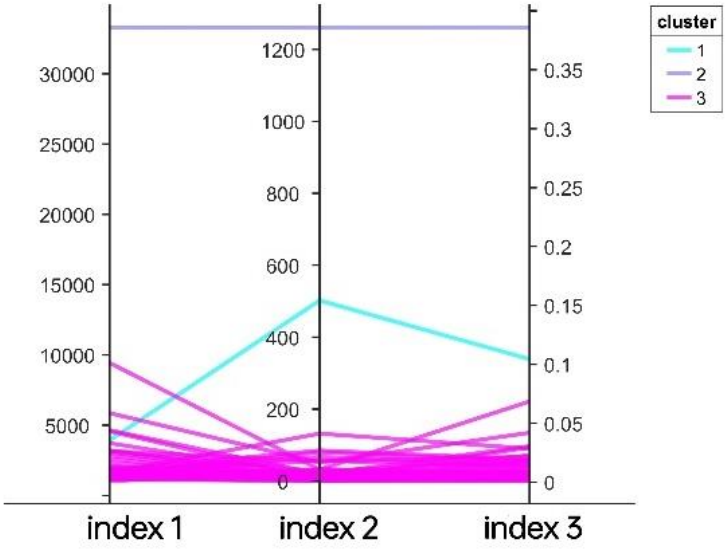


Figure 1. Parallel Coordinate Chart

As shown in Figure 1, where Index 1, Index 2, and Index 3 represent storm economic losses, deaths, and composite scores for the U.S. East Coast, respectively; Cluster 1, Cluster 2, and Cluster 3 represent the northern, southern, and central regions of the U.S. East Coast, respectively.

The range of values corresponding to indicators I, II and III in the third cluster is lower compared to the other two clusters; for the first cluster it has a lower value of indicator I, even lower than some of the values of indicator I in the first cluster, while the values of indicators II and III are significantly higher than the values corresponding to indicators II and III in the first cluster, but all of them are lower than the corresponding values of the indicators in the third cluster.

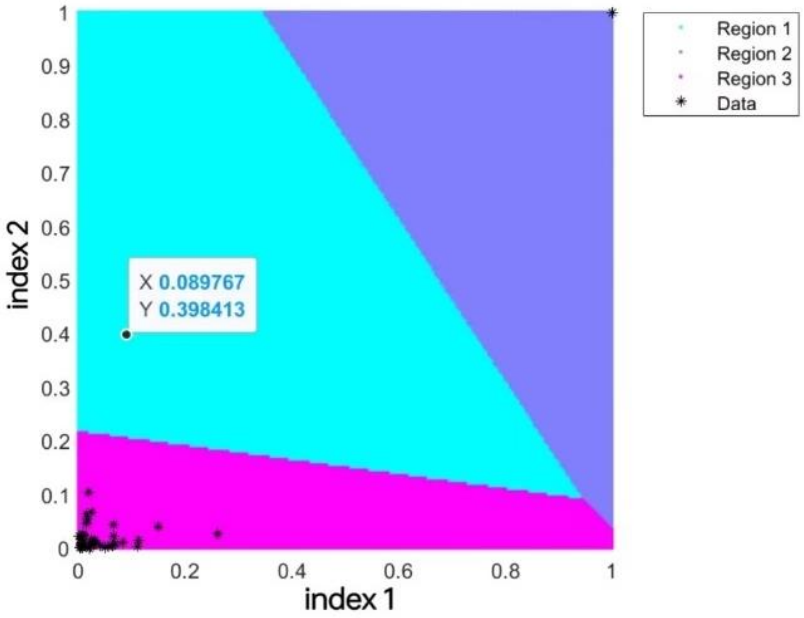


Figure 2. Area Chart

As shown in Figure 2, the light blue region has the largest area so the first cluster has the largest number and the purple region has the smallest third cluster also has the smallest number. The area map also reflects that when fixing the horizontal coordinates of the first cluster constant it is increasing with the vertical coordinates of the indicator two, i.e., the loss of property gradually decreases.

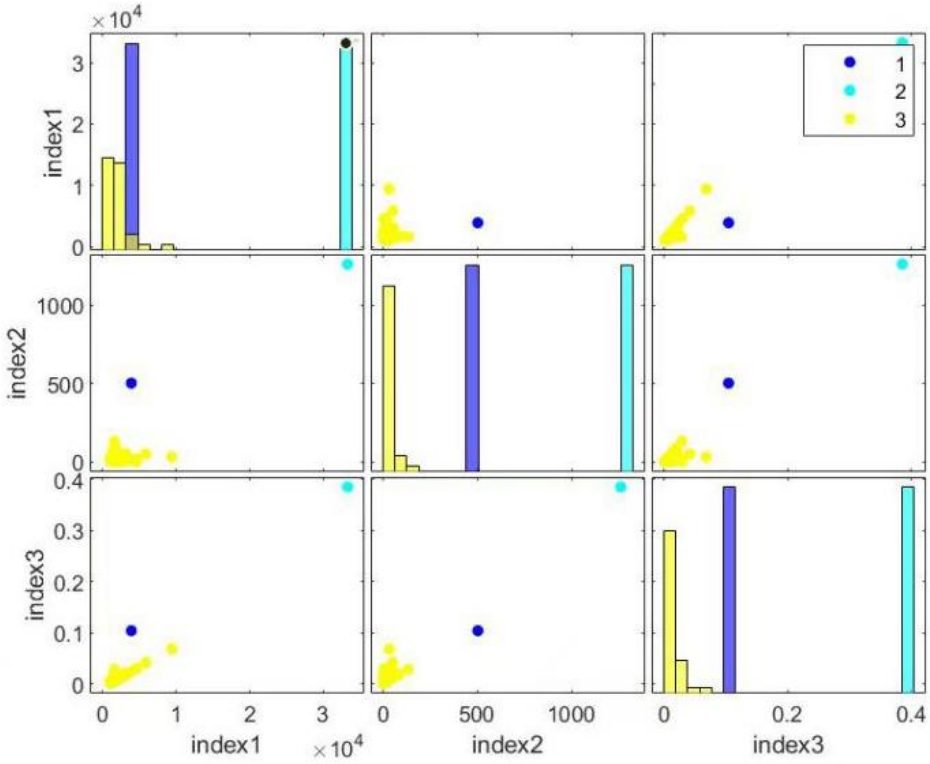


Figure 3. Scatterplot Matrix

As shown in Figure 3, the histograms at the diagonal positions of this matrix represent the distribution and size of the three clusters in terms of property damage, fatalities, and composite scores in their corresponding horizontal coordinates. The histograms in the second row and second column reflect the number of deaths in each of the three clusters, with the size of deaths in the third cluster concentrated in the 0 to 500 range; the remaining scatter plots show the interrelationships among the three clusters with respect to property damage, the population of deaths, and the composite score metrics; and the left-most scatter plots in the second row show the relationship between property damage and deaths in each of the three clusters.

3.2. Results of the the Insurance Profitability Model

In this subsection, we first used a multiple regression model to describe the relationship between the dependent variable and several independent variables. The model was first analyzed and the code was run to obtain the following data as shown in Table 2.

Table 2. Results of regression model runs

Variable	VIF	1/VIF
Adult population	1.09	0.920078
Storm damage	1.09	0.920078
Mean VIF	1.09	

We define the regression model VIF = max {VIF1, VIF2} = 1.09. Based on empirical judgment, VIF is less than 10, and it is considered that the regression model does not have serious multicollinearity and does not need further multicollinearity treatment.

Therefore, the result is:

$$\hat{\Pi i} = 0.725\hat{\zeta i} - 7.13 \times 10^{-6}\hat{Y i} \tag{8}$$

A time series plot reflecting the temporal relationship of the Philippine population from 2008 to 2022 and the original time series versus the fitted time and future projections were plotted based on the available data as shown in Figures 4 and 5:

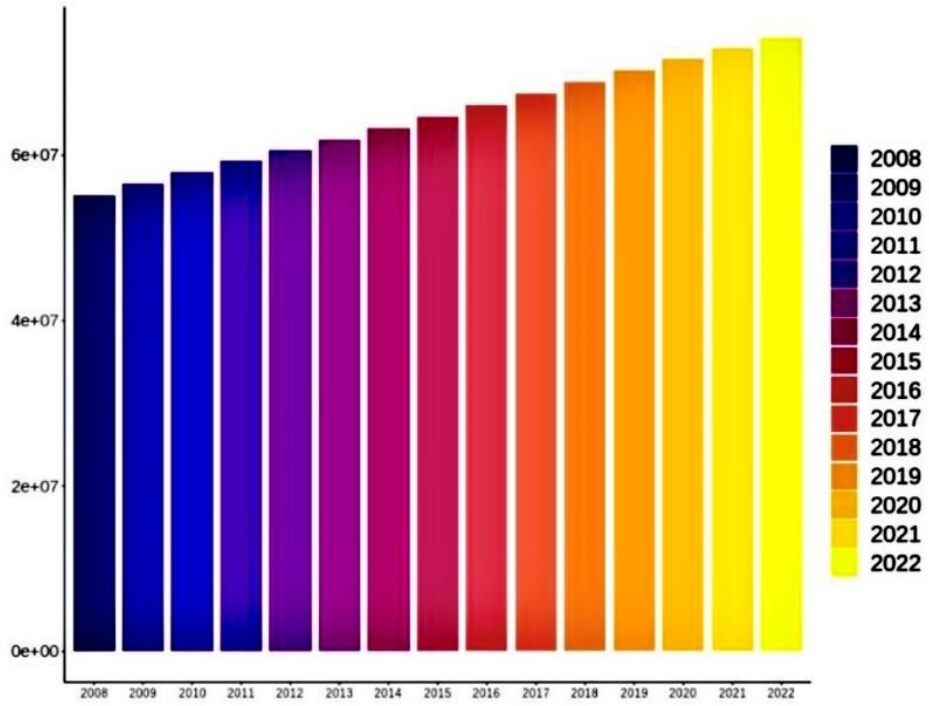


Figure 4. Philippine Population Time Series Chart

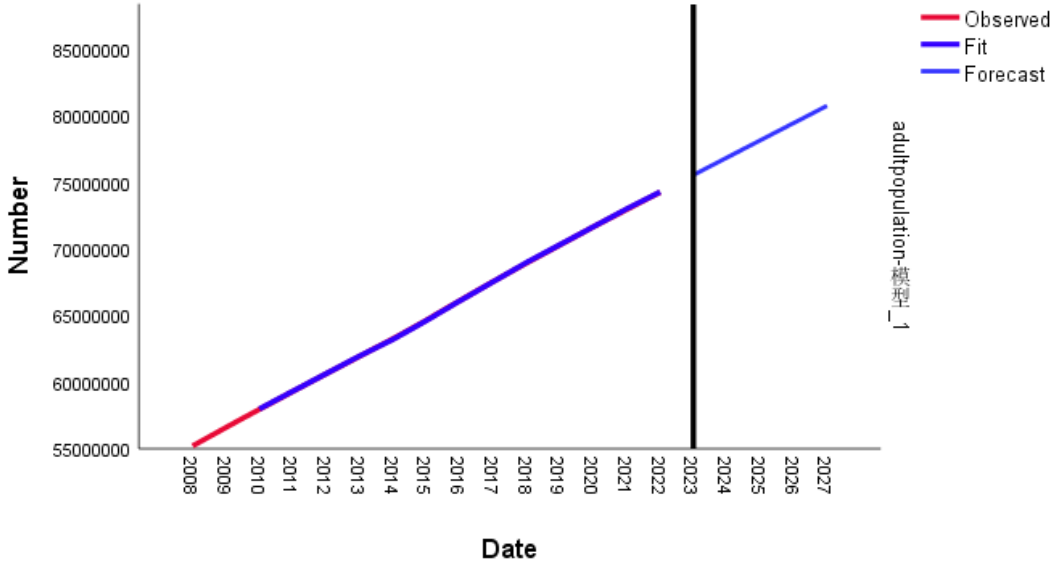


Figure 5. Comparison chart

As can be seen from the figure, the projected Philippine population for the latter five years shows a linear trend of simultaneous increase with yearly growth.

We then applied a time series model to forecast the future earnings of Philippine insurance companies:

Average population projections for the Philippines for the next five years, $\overline{\zeta_1}$ is 78103909, unit is person.

Projected economic losses from storms in the Philippines over the next five years, $\overline{Y_1}$ is 37288000, unit is dollar.

Banks in the Philippines forecast profits, $\hat{\Pi}_1$ is 56.625, unit is million dollars.

Note: We use a time-series model to forecast the next five years of data for both countries, using the average of these five years of data as the predicted value for a particular year in the future, avoiding extremes to some extent.

4. Conclusions and outlooks

This paper establishes a risk assessment model and a profit model for insurance companies in response to the challenges faced by insurance companies in extreme climates. The risk assessment model is based on the TOPSIS comprehensive evaluation method, which solves the problem of under what circumstances insurance companies must take risks and underwrite insurance. In practice, insurers can use the risk assessment model to determine the level of risk in different regions and can choose to underwrite in low to medium risk areas. The profit model is used to determine the underwriting profit of the insurance company, through the use of double coefficients of mutual constraints on the profit model, in order to achieve the optimal positioning of the insurance company to meet the conditions of the future payout capacity and long-term health, the two models not only reduces the risk of the insurance company's loss to improve underwriting profit at the same time for the long-term healthy development of the entire insurance industry to play an important significance.

There are also some shortcomings in the modeling such as the sensitivity of the TOPSIS method itself is not high therefore when sorting evaluation objects, it may be difficult to distinguish between their advantages and disadvantages for similar evaluation objects. In the multiple regression model, we use the time series method to get the average of the predicted value as a variable to solve the problem, but due to the limitation of valid data, we only use the predicted value of the next five years to calculate the mean value, and in the subsequent problem solving, we will find more samples to make predictions to ensure the effectiveness of the model.

References

[1] Daniels, S., Illinois' homeowners hikes rival those in disaster prone states: Insurers like Allstate have raised premiums here recently by hundreds annually. The industry says climate change is affecting Illinois, too[N]. 2023.

[2] Tyson, J., 5 tips for cutting insurance costs amid climate change[N]. Supply Chain Dive; Washington,2023.

[3] Feng A Q, Yue X L, Chao Q C, et al. Climate change risk and green insurance response under the goal of carbon peak and carbon neutral in China[J]. Environmental Protection, 2021, 49(8): 20-24.

[4] Yanmei, Y., Risks and challenges to the development of agricultural insurance in the face of climate change shocks[J]. Finance, 2022, 12: 510.

[5] COHEN A, FISHHENDLER I. An archetype for insurance thresholds for extreme natural events in the agricultural sector[J]. Climate Risk Management, Climate Risk Management, 2022.

[6] Fang Yunlong, L.H., Study on the impact of and response to climate-related risks on the financial sector[J]. 2023.

[7] ZHONG L, NIE J, YUE X, et al. Optimal design of agricultural insurance subsidies under the risk of extreme weather[J/OL]. International Journal of Production Economics, 2023: 108920.

[8] Braun A, Utz S, Xu J. Are insurance balance sheets carbon-neutral? Harnessing asset pricing for climate change policy[J]. The Geneva Papers on Risk and Insurance-Issues and Practice, 2019, 44: 549-568.

[9] Mingzhou, F.Z.J., Research and Application of Natural Disaster Risk Management System with Geoinformatics and Artificial Intelligence--Taking Ping An General Insurance's Self-built Eagle Eye System DRS2.0 as an Example[J]. Insurance Theory and Practice, 2024, (02):49-61.

[10] LINNEROOTH-BAYER J, MECHLER R, HOCHRAINER-STIGLER S. Insurance against Losses from Natural Disasters in Developing Countries [J]. Journal of Integrated Disaster Risk Management, 2011, 1(1): 59-81.