Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 8, No. 2, 2024

Research on Natural Disaster Risk Assessment and Insurance Decision-Making Using EWM-TOPSIS and ARIMA Models

Ji Zhou [1], Yi Bu [2]

[1] School of Information Science and Engineering, Shenyang University of Technology, Shenyang, Liaoning, China
[2] School of Economics, Shenyang University of Technology, Shenyang, Liaoning, China

Abstract: This study aims to evaluate natural disaster risks and provide decision support for insurance companies by integrating the EWM-TOPSIS (Entropy Weighted Method combined with Technique for Order Preference by Similarity to Ideal Solution) and ARIMA (AutoRegressive Integrated Moving Average) models. Initially, through literature review and online data collection, we identified multiple factors associated with various natural disasters, including PPM (Parts Per Million), degrees Celsius, temperature, rainfall, altitude, frequency of natural disasters, and actual water barometric pressure. The entropy weight method was utilized to calculate the weights of these factors, which were then combined with the TOPSIS comprehensive evaluation method to determine the extent of disasters in different regions. Additionally, a random forest model was constructed to predict the underwriting risk for insurance companies, considering factors such as population density, income levels, and the strength of agricultural, industrial, and service sectors. Furthermore, the ARIMA model was employed for time series analysis of risk factors to forecast future risk trends, thereby identifying the optimal timing for insurance companies to assume risk. Finally, from the perspective of real estate developers, the study evaluated the potential benefits of constructing houses in different regions by combining the entropy weight method with the TOPSIS model. The findings indicate that Washington State has a high disaster severity index when considering the potential benefits of future real estate development, suggesting that real estate development in this area could yield higher returns. This study provides a scientific basis for risk management in the insurance industry and site selection decisions for real estate developers.

Keywords: EWM-TOPSIS; ARIMA Model; Natural Disaster Risk Assessment; Insurance Decision-making; Real Estate Development.

1. Introduction

Natural disasters pose a great threat to human society. Not only do they cause human casualties, they can also cause economic losses and environmental damage. With global climate change and population growth, natural disasters are increasing in frequency and intensity, making natural disaster risk assessment and insurance decisions particularly important [1-2]. In the insurance industry, accurate risk assessment is the key to formulating insurance policies and pricing [3]. However, traditional risk assessment methods often rely on historical data and expert experience, and may not fully account for the complexity and uncertainty of multiple factors. Therefore, it is of great significance for the insurance industry to develop a new method that integrates multiple factors and can dynamically assess the risk of natural disasters [4].

This paper presents a research method of natural disaster risk assessment and insurance decision making based on EWM-TOPSIS and ARIMA models [5-6]. EWM-TOPSIS method uses entropy weight method to determine the weight of each factor, combined with good and bad distance method for comprehensive evaluation, while ARIMA model uses time series analysis to predict the future trend of risk factors. The combination of these two methods aims to provide a more accurate and dynamic risk assessment tool to aid insurance decisions.

2. Random Forest Prediction based on Ewm-Topsis

We learned what factors are associated with various natural disasters based on the literature and collected data online on PPM, Degress C, temperature, rainfall, altitude, frequency of natural disasters, and actual water barometric pressure for these different areas. Literature shows that: these factors can measure the extent of disasters in these areas. We utilize the entropy weighting method to calculate the weights of the various factors, combined with the TOPSIS comprehensive evaluation method to calculate the degree of disaster in these areas, and utilize the combination of the two models to make the solution more accurate.

2.1. Weight Calculation based on Entropy Weight Method

The basic idea of Entropy Weight Method to calculate the weight of indicators is based on the perfect evaluation index system, and then according to the size of the information entropy Ej of the evaluation indicators to determine the degree of dissimilarity of the indicators. The weights of the evaluation indicators are finally obtained value. The calculation step is as follows.

According to the Entropy Weight Method, the weights of PPM, Degress C, temperature, rainfall, altitude, disaster frequency, and actual water pressure can be calculated, as shown in the table 1 and Fig. 1.

Table 1. Weights of each factor calculated based on the Entropy Weight Method

Factor	Indicator weights
PPM	0.022
Degrees C	0.1258
Temperature	0.1378
Rainfall	0.1221
Altitude	0.1253
Disaster Frequency	0.1399
Actual water pressure	0.1542

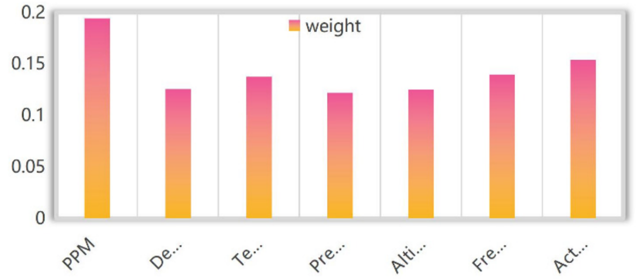

Fig 1. Weights of each factor calculated based on the Entropy Weight Method

The weights show that a place's exposure to disasters is most relevant to the PPM, and these weights are used to correct the raw data

2.2. TOPSIS Method for Evaluating the Extent of Disasters

The Entropy Weight TOPSIS method is to multiply the weights of each factor derived from the Entropy Weight Method with the standardized index matrix to get the canonical weighted evaluation matrix, and then calculate the positive and negative ideal solutions, the Euclidean distance, and the degree of closeness, and finally sort according to the value of the degree of closeness.

Specifically, the entropy weight method is used to determine the weight of each factor, and then these weights are multiplied with standardized evaluation indicators to obtain a normalized weighted evaluation matrix. Then, the positive and negative ideal solutions are calculated based on the matrix, and the Euclidean distance between the evaluation object and these ideal solutions is calculated. After that, the approximation degree between the evaluation object and the ideal solution is calculated. The greater the approximation degree, the better the performance of the evaluation object. Finally, the evaluation objects are sorted according to the value of proximity, so as to evaluate the disaster severity in different regions.

In addition to different orders of magnitude, the nature of the indicator (positive or negative, whether it is a benefit or cost or interval indicator) may also differ, there are three main types of indicators that can be processed by the TOPSIS method. Among them, the larger the benefit index, the better, the smaller the cost index, the better, and the better the interval index falls within a certain range.

The magnitude of the weights shows that the degree of exposure of a place to hazards is most relevant to the PPM, and the evaluation methodology yields a score for these areas, called the Hazard Severity Index (HSI). The table 2 illustrates the disaster severity index for selected areas.

Table 2. Catastrophe Degree Index by Region

Area	Disaster Severity Index
Washington	0.3587
Alaska	0.3864
Hawaii	0.4512
Utah	0.4008
New Mexico	0.3346

We have selected a few of the more prominent regions in the United States, and the calculated disaster severity index is shown in Fig. 2.

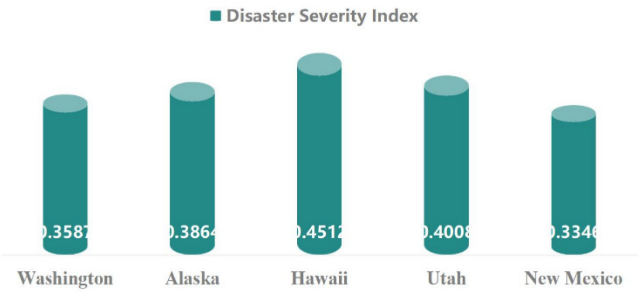

Fig 2. Hazard Degree Index by Region of the United States

2.3. The Random Forest Model Predicts Whether it is Underwritten

Fig 3. Heat maps of several types of underwriting risk factors

For insurance companies, the hazard index is a key factor in whether or not to underwrite a policy. In addition to this, a review of the literature shows that the choice of insurance companies to underwrite is also related to the population density of the area, its income bracket, and the strength of its agro-industrial and service sectors. Based on these data, we decided to build a random forest model to predict when insurance companies will choose to underwrite and when they will choose not to underwrite.

Income classes can be categorized into five categories by reviewing the literature; low income, low-moderate income, middle income, upper-moderate income, and high income. The correlation between several factors is plotted into a heat map, as shown in Fig. 3.

The Random Forest algorithm requires a number of decision tree models to be combined in parallel to determine the final result based on the majority voting principle, which is defined as follows: Using the majority voting principle or averaging to get the final result. Fig. 4 depicts the importance of random forest features.

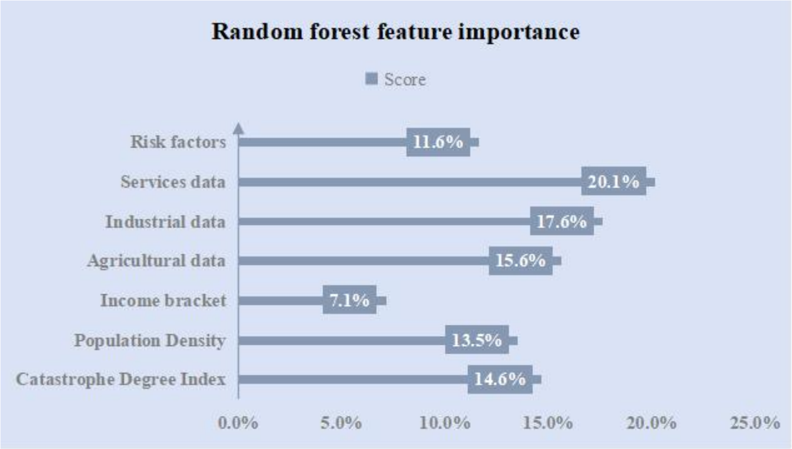

Fig 4. Random forest feature importance

The ROC curve is often used to evaluate the prediction performance of a random forest model, and Fig. 5 shows the ROC curve that uses the random forest model to predict whether an insurance company will underwrite it.

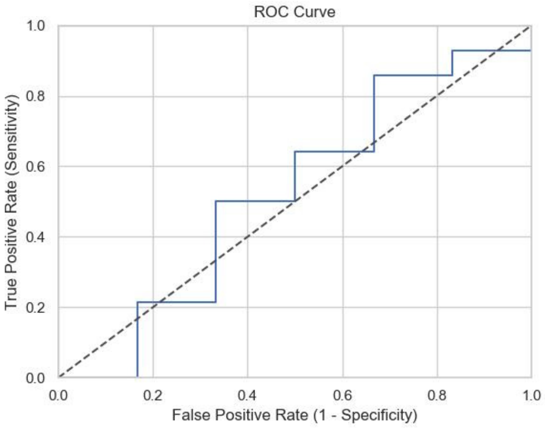

Fig 5. ROC curves predicted by the random forest model

The risk factor has a very important weight in the tree in terms of whether or not it is covered, and the size of the risk factor will basically determine whether or not the insurance company will cover it. Fig. 6 shows the important characteristics of the first tree in a random forest.

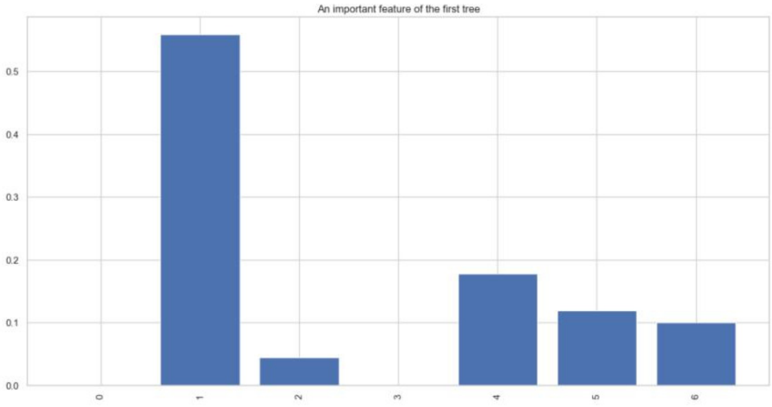

Fig 6. Important characteristics of the first tree in a random forest

3. Arima Prediction Model

3.1. Discussion of Risk Factors

We need to explore when insurance companies need to take risks. After reviewing the relevant literature, we found that we can predict the direction of the risk factor to determine when the insurance company has the greatest benefit from risk-taking, so we collected data on the risk factor at different times, and used it to make predictions on the risk factor in the following period of time.

The more severe the disaster the greater the risk of underwriting. Based on the disaster severity index we defined above, it can be seen that the disaster severity index is inversely proportional to the risk factor. The larger the population, the greater the risk theoretically; the lower the income bracket, the greater the risk; The stronger the agricultural, industrial and service sectors, the less risky.

3.2. Establishment of the ARIMA Model

A series of time series data can be obtained through the calculation of the risk factor formula, and the original time series diagram of the data is shown in Fig. 7.

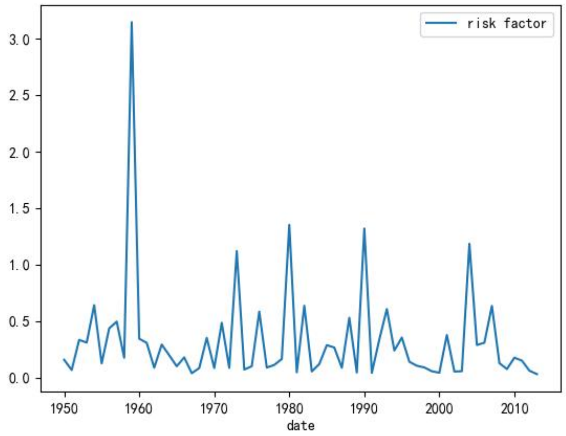

Fig 7. Temporal series diagram of risk factors

For this set of risk factor time series data need to be ADF test, the unit root statistic corresponds to a p-value significantly greater than 0.05, which is judged to be a non-stationary series, so it is necessary to carry out a first-order difference operation on this column of risk factor data, and then re-plot the time series. The sequence diagram of the first-order differential of the raw data is shown in Fig. 8.

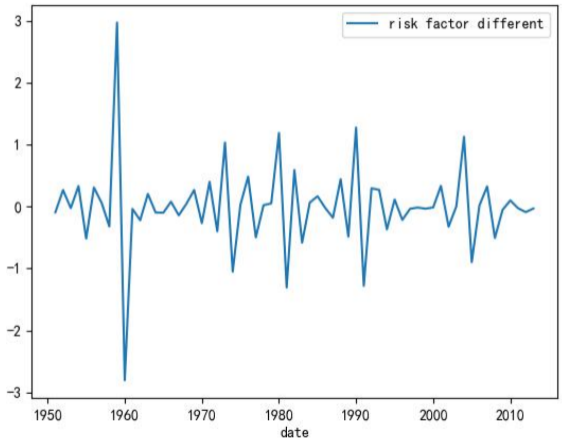

Fig 8. Timing diagram of the first-order difference series of risk factors

Sequence chart after first-order difference is stationary near the mean. Volatility auto correlation chart has strong short-term correlation. Unit root test p value is less than 0.05. Therefore, the sequence after first-order difference is stationary. Auto correlation chart and partial auto correlation chart are shown in Fig. 9.

The ARIMA (P, D, Q) model is based on ARMA fitting after D-order difference. A series of steps to build an ARIMA model can be summarized as follows. First, check the stationarity of the time series data of risk factors, confirm the non-stationarity of the series, and carry out the first-order difference operation to achieve the stationarity of the data. Then, the appropriate ARIMA model parameters are identified by drawing autocorrelates and partial autocorrelates. Then, based on the identified parameters, ARIMA (p, d, q) model is established, and the risk factors are fitted with the model. After the parameter estimation is completed, the significance test of the model is carried out, including the significance test of the whole model and the significance test of the parameters, to ensure the validity of the model. Then, the model is optimized to improve its prediction accuracy. Finally, the validated and optimized ARIMA model is used to predict the future risk factor values, and the future trend of the series is judged based on the predicted results.

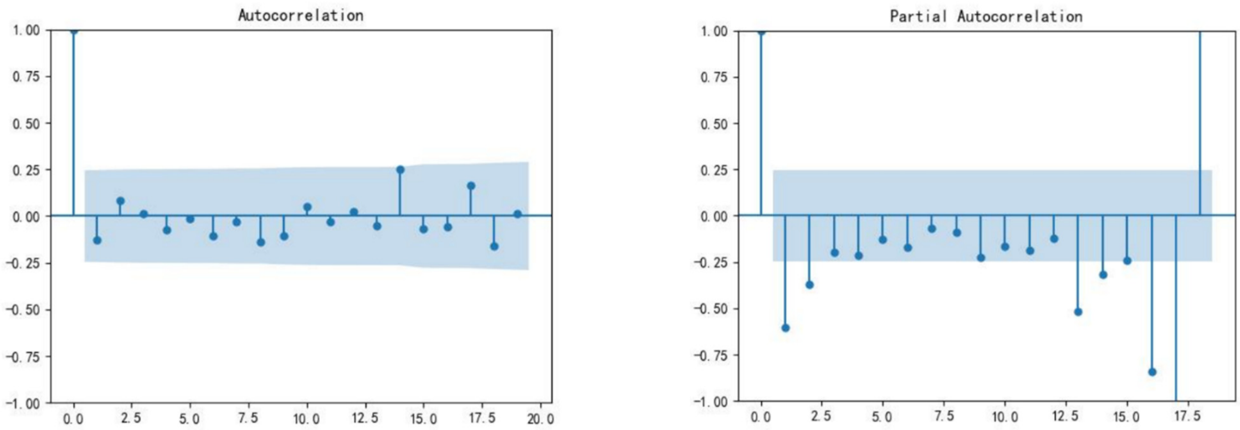

Fig 9. First-order difference/difference-biased auto correlation plot of risk factors

Here we select the risk factor data to be predicted backward for 10 periods to observe the trend of the data, and the diagram predicted by the ARIMA model is shown in Fig. 10.

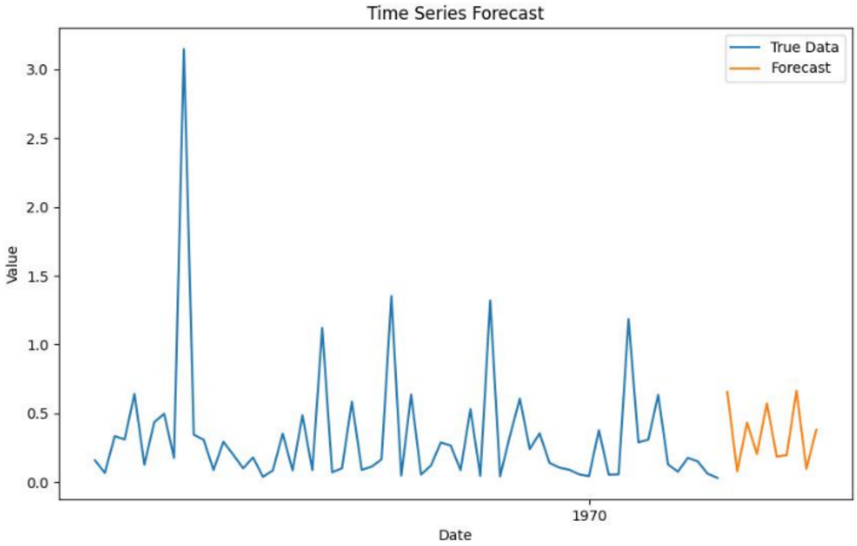

Fig 10. ARIMA predicted risk factor data graph

It can be seen from the diagram that the yellow part of the curve is the predicted value after the risk factor. Here, we predicted the data of the 10 periods backward, and we can see that the risk factor is low in the 2, 4, 6 and 8 periods. At this time, the insurance can choose whether the risk will benefit. The table 3 shows the forecast after 10 periods

Table 3. Risk factor data after 10 periods of prediction

Year	Risk Factor
2014	0.6544
2015	0.0768
2016	0.4317
2017	0.2030
2018	0.5710
2019	0.1845
2020	0.1946
2021	0.6630
2022	0.0953
2023	0.3792

3.3. Apply the Model from the Perspective of a Real Estate Developer

Earlier we selected several important factors according to the perspective of the insurance company, and now we need to modify the random forest model just now from the perspective of real estate developers, on the basis of the original factors, by consulting the literature, it is more beneficial to determine where the real estate developer builds a house, according to the entropy weight method just now combined with the TOPSIS model, comprehensive evaluation score, here we have selected several famous continents in the United States, in these continents, For real estate developers, in addition to the factors that insurance companies need to consider, there are more important factors to consider, such as: using the predicted data to score and determine where to build a house in the future is most beneficial to real estate developers.

By bringing the data into the insurance model of the first question, it is possible to get the most benefits of the future real estate to build a property in Washington State. The table 4 shows the real estate benefit scores for a select number of states in the United States.

Table 4. Real Estate Benefit Score Table

Area	Disaster Severity Index
Washington	0.45
Alaska	0.35
Hawaii	0.32
Utah	0.40
New Mexico	0.37

4. Conclusion

This study has successfully applied EWM-TOPSIS and ARIMA models in the field of natural disaster risk assessment and insurance decision making, providing a new decision support tool for insurance industry and real estate developers. By combining a number of factors associated with natural disasters, we are able to quantify risk exposure in different regions and make recommendations for risk management accordingly. A limitation of this study is that the predictive power of the model is limited by the quality and range of data used. Future studies can explore more risk factors and improve the generalization ability of the model. In addition, the parameter selection and optimization of the model is also an important direction of future research.

Overall, this study provides a new perspective for understanding and managing natural disaster risk, and provides practical tools and recommendations for decision makers in the insurance and real estate industries.

References

[1] Cui Yin, Zhang Shuxia. Research on the countermeasures for the high-quality development of animal husbandry insurance under the risk of major disasters: A case study of Hebei Province[J]. Agricultural Outlook,2023,19(10):25-30.

[2] Tsai C H, Chen C W. The establishment of a rapid natural disaster risk assessment model for the tourism industry[J]. Tourism management, 2011, 32(1): 158-171.

[3] Owoyemi A, Nnaemeka E, Benson T O, et al. Machine learning recommendation system for health insurance decision making in Nigeria[J]. arXiv preprint arXiv:2305.10708, 2023.

[4] Cremen G, Galasso C, McCloskey J. Modelling and quantifying tomorrow's risks from natural hazards[J]. Science of The Total Environment, 2022, 817: 152552.

[5] Zhang Y, Ng S T. Identification and quantification of node criticality through EWM–TOPSIS: a study of Hong Kong's MTR system[J]. Urban Rail Transit, 2021, 7(3): 226-239.

[6] Shumway R H, Stoffer D S, Shumway R H, et al. ARIMA models[J]. Time series analysis and its applications: with R examples, 2017: 75-163.