Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 9
GBMS 2024

Innovative Insurance and Profit Modeling: Integrating ARIMA and LSTM for Risk Assessment and Profit Maximization Strategies

Meng Yuan *, Zihao Zou, Shuhao Chang

School of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, China, 221116

* Corresponding Author Email: 03220605@cumt.edu.cn

Abstract. This research presents an innovative approach to insurance and profit modeling by integrating ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) techniques for risk assessment and profit maximization in the insurance industry. The study initially establishes sub-models focusing on income modes and loss modes, utilizing historical data and risk factors to calculate potential profits and losses. Through the development of a profit model based on nonlinear programming, the study aims to optimize costs associated with natural disaster claims while maximizing profitability. Leveraging machine learning techniques, particularly LSTM, the research refines time series models to predict and mitigate risks more accurately. The findings emphasize the importance of geographical variations and timely underwriting decisions for insurance companies to ensure business stability and sustainable growth amidst evolving risk landscapes.

Keywords: Innovation, Sustainability, Resilience.

1. Introduction

The insurance industry plays a critical role in enabling individuals and businesses to manage risks and uncertainties in an ever-changing world. With the increasing complexity and frequency of natural disasters, pandemics, and other unforeseen events, there is a growing need for innovative approaches to assess risks effectively and maximize profitability [1-2]. Traditional models often fall short in capturing the dynamic nature of risks and their financial implications. This research explores a novel methodology that combines ARIMA and LSTM techniques to enhance risk assessment and profit optimization in insurance. By delving into income and loss modes while incorporating historical data and risk factors, this study aims to pave the way for more informed decision-making processes in the insurance sector. Embracing advanced machine learning algorithms, particularly LSTM, holds the promise of revolutionizing how insurers predict, manage, and respond to risks in a rapidly evolving landscape. Thus, this research seeks to contribute to the resilience and competitiveness of insurance companies in the face of escalating uncertainties and challenges [3-5].

2. Insurance and Profit Model based on ARIMA and LSTM

2.1. The Establishment of Insurance and Profit Model

2.1.1. Sub-model I: Income Mode

Based on the initial data, we have decided to use the population of the area as potential policyholders. The actual number of customers has been calculated using historical insurance rates in the area. To simplify the model, we have consulted the insurance company's official website for information and set the insurance premium accordingly.

$$I = \sum_{i=1}^{N} I_i \tag{1}$$

The insured amount for a single type is

$$I_i = np_i m_i \tag{2}$$

Among which, represents the total population in the local area, corresponding to the insurance rate of this type of insurance in the local area, and the premium that each person needs to pay for this type of insurance [6].

2.1.2. Sub-model II: Loss Mode

To quantify the amount of loss, we choose to introduce risk factors as indicators. After categorizing different local disasters, we search for their relationship with risk loss.

$$S(t) = f(t)g(t) \tag{3}$$

After differentiating the expression, we obtain the derived expression for the risk factor

$$f(t) = \frac{d_{s(t)}}{d_{g(t)}} \tag{4}$$

Subsequently, we examine the relationships between various risk factors to identify the influencing factors and their quantitative representation. Through literature review and data fitting, we find a strong correlation with CPI (Consumer Price Index). Consequently, we establish and validate the risk factor impact model using the Monte Carlo simulation method.

$$f(t) = K\frac{CPI_t}{CPI_e - ln(k + CPI_s)} \tag{5}$$

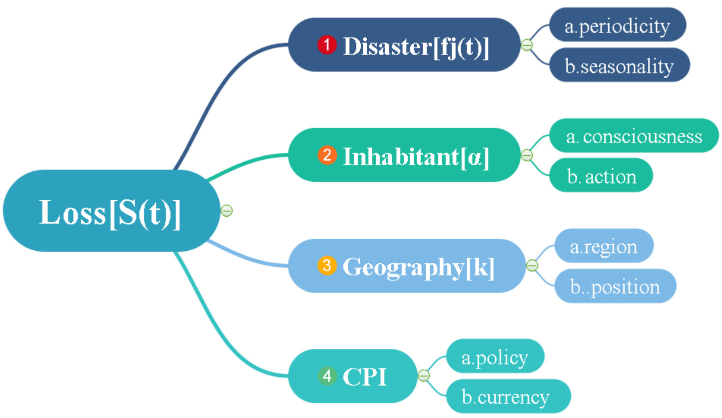

Figure 1. The composition of $S(t)$.

2.1.3. Core Mode

To tackle the issue of declining profitability in the insurance company, we first establish a profit model based on nonlinear programming. The model aims to minimize the costs associated with natural disaster claims and maximize profits in the insurance industry [7].

$$\omega = I - S(t) \tag{6}$$

$$S(t) = \sum_{i=1}^{N} f_i\,(t) IN_j + c \tag{7}$$

$$IN_i = \ g(t) \tag{8}$$

In addition to the objective conditions of natural disasters, the subjective factor of insurance customers' purchasing behavior is also included in our model. We propose the homeowner behavior factor $\alpha(0 < \alpha < 1)$, which is influenced by multiple factors such as the homeowner's personal awareness of safety precautions, the quality of local infrastructure. A higher homeowner behavior factor indicates a lower probability of encountering a disaster, resulting in smaller insurance payouts for the insurance company. Taking this indicator into account, we further refine and optimize the initial model [8-10].

$$\omega = I - (1 - \alpha) S(t) \tag{9}$$

This is the profit formula we have established for the insurance company.

$$\omega = \sum_{i=1}^{N} np_i m_i - (1 - \alpha) \sum_{j=1}^{N} K_j \frac{CPI_t}{CPI_e - ln(k + CPI_s)} g(t) - c \tag{10}$$

Taking into account real-world constraints, insurance companies are also subject to various constraints, such as the minimum operating cost of the company.

$$\text{s.t.} \begin{cases} f_i(t) IN_j + c > np_i m_i, & i,j = 1,2,\cdots,n \\ CPI_e - ln(k + CPI_s) > 0, & \\ \sum_{i=1}^{N} I_i > M, & i = 1,2,\cdots,n \\ S(t) \geq 0, I \geq 0, & \\ K > 0, 0 < \alpha < 1, & . \end{cases} \tag{11}$$

2.2. The Solution of Insurance and Profit Model

We collected data on natural disasters such as floods, storms and other major disasters in several countries from different websites, we obtained the Number or severity of disasters from 1960 to 2022.Additionally, we collected the corresponding loss information caused by disasters during the same period.

Table 1. Data source collation.

Database Names	Database Websites Data	Type
e-Stat	https://www.e-stat.go.jp	Population
EM-DAT	https://public.emdat.be/	Geography
CEIC	https://www.ceicdata.com.cn/	economy
WB	https://www.worldbank.org/	currency

Based on processed monthly storm statistics from Japan between the 2000s and 2022s, we have established a seasonal time series model for the assessment of risks. This model focuses on the intensity (impact) of disasters, represented as the variable IN.

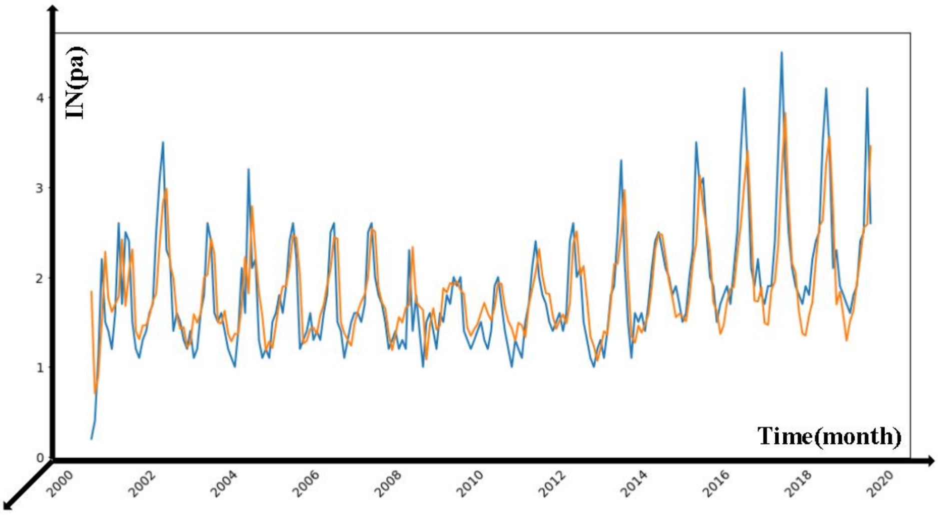

Figure 2. Seasonal ARIMA model of storm in Japan

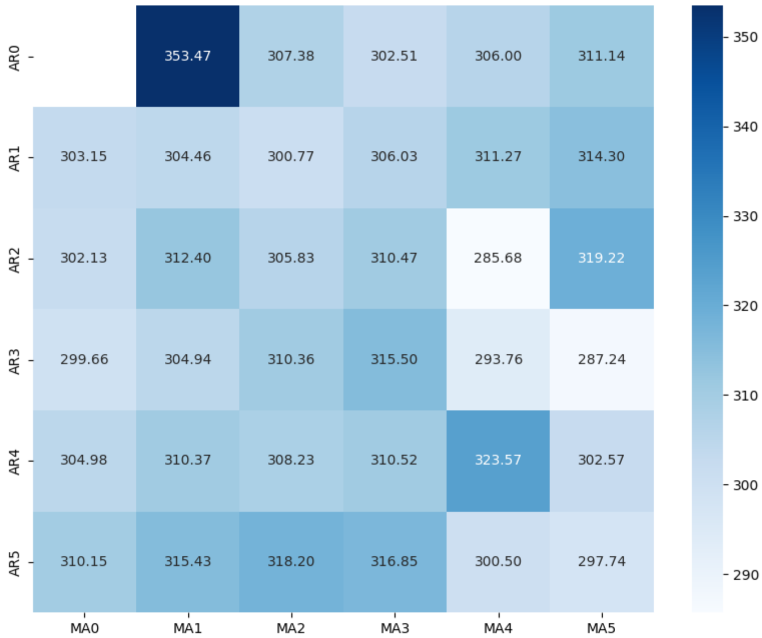

Figure 3. BIC criterion Test.

According to the figure, we set up a 5x5 grid to calculate the values of p and q for the time series. By allowing the system to automatically search for the optimal values, we obtained a minimum value of 285.68. Therefore, we can determine that $p = 2$ and $q = 4$.

At the 0th order of difference, the significance p-value is 0.856, which indicates that there is no significant evidence at the given level to reject the null hypothesis. Therefore, the sequence is considered a non-stationary time series.

At the 1st order of difference, the significance p-value is very small, indicating significant evidence at the given level to reject the null hypothesis. Therefore, the sequence is considered a stationary time series, and $d = 1$.

Finally, the computed goodness-of-fit measure R^2 is 0.64, indicating that the model performs relatively well. The model meets the basic requirements but can still be further optimized using machine learning techniques.

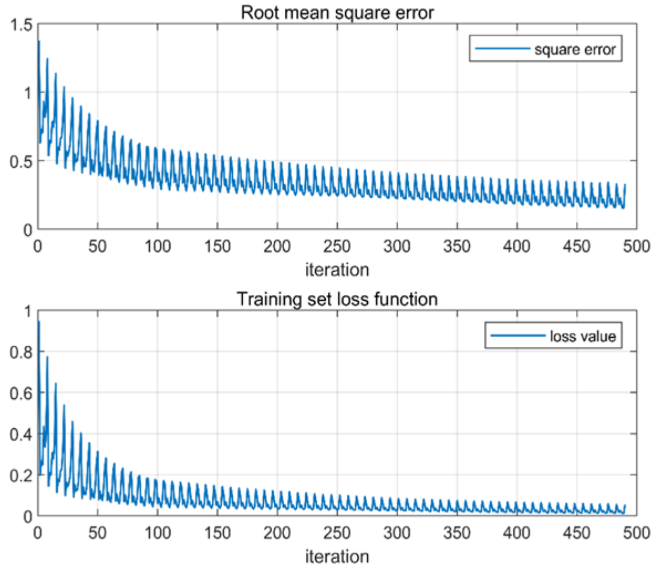

Figure 4. LSTM model, root mean square error and loss function.

To minimize the potential errors of the ARIMA time series model, we compared it with the LSTM neural network. To improve machine training, we applied CWT noise reduction to the data The curve shows that the data obtained after noise removal is more accurate, which improves the performance and robustness of the model. This not only reduces the calculation amount but also the calculation error, making it easier to establish an accurate time series model.

After data processing, we fed the data into the neural network for machine learning. We set the ratio of training, validation, and testing sets to 8:1:1, with a rolling step of 2. Additionally, we configured the neural network with 32 neurons and trained it for 70 epochs. To further minimize errors, we employed the Particle Swarm Optimization (PSO) algorithm for intelligent optimization.

The position of particle number is.

$$X_{id} = (x_{i1}, x_{i2}, \ldots, x_{iD}) \tag{12}$$

The velocity of particle number is.

$$V_{id} = (v_{i1}, v_{i2}, \ldots, v_{iD}) \tag{13}$$

The best solution found by particle number is

$$P_{id,pbest} = (p_{i1}, p_{i2}, \ldots, p_{iD}) \tag{14}$$

The best position found by the swarm search (global best solution) is.

$$P_{d,gbest} = (p_{1,gbest}, p_{2,gbest}, \ldots, p_{D,gbest}) \tag{15}$$

Velocity update formula

$$v_{id}^{k+1} = \omega v_{id}^{k} + c_1 r_1 (p_{id,\mathrm{pbest}}^{k} - x_{id}^{k}) + c_2 r_2 (p_{d,\mathrm{gbest}}^{k} - x_{id}^{k}) \tag{16}$$

By implementing the above formula in our programming, we can optimize the neural network. After several iterations of particle swarm optimization and observing the performance on the test set, we can obtain a more accurate time series prediction model. This model can be used to determine the changing relationship between the impact of extreme weather disasters and time

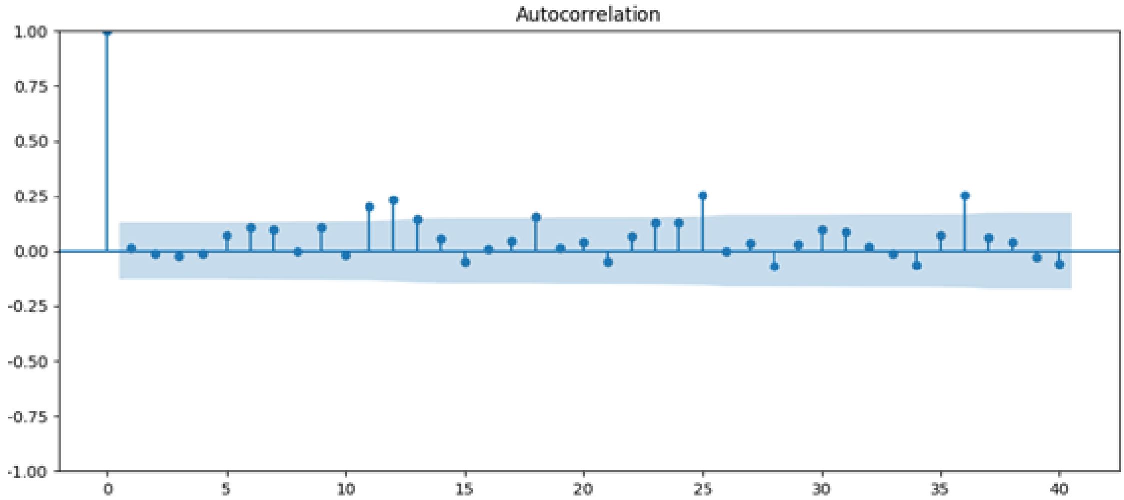

Figure 5. ACF-Test.

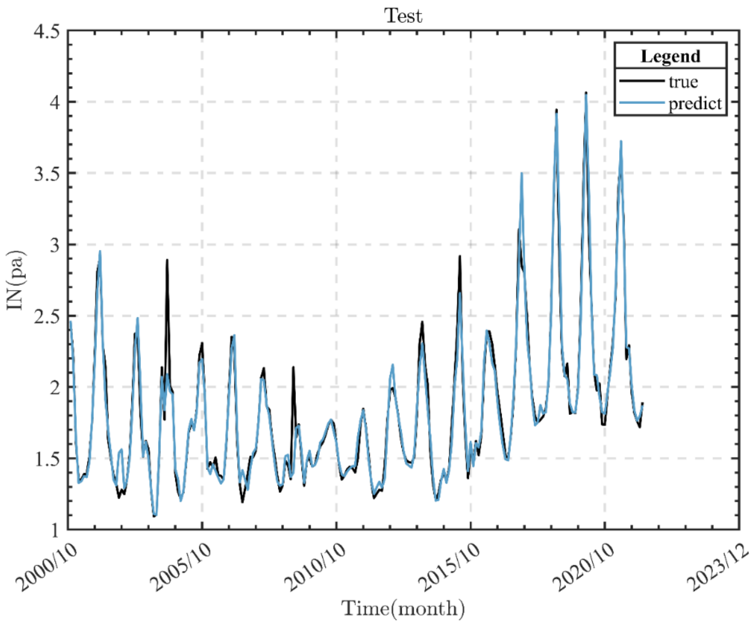

Figure 6. LSTM Test Set.

Given that the major contributing factors to extreme weather disasters in Japan are floods and storms, with secondary impacts being other losses, the total loss can be expressed as follows.

$$S(t) = f_{flood}(t)IN_{flood} + f_{storm}(t)IN_{storm} + \sum_{i=1}^{k} f_i\,(t)IN_j + c \tag{17}$$

First, we will analyze the relationship between flood disasters in Japan and their occurrence time each year. Then, we will establish a corresponding relationship between the loss situation and the impact of floods in the same temporal and spatial context. This will result in a correlation between the amount of flood losses and the frequency of their occurrence over time.

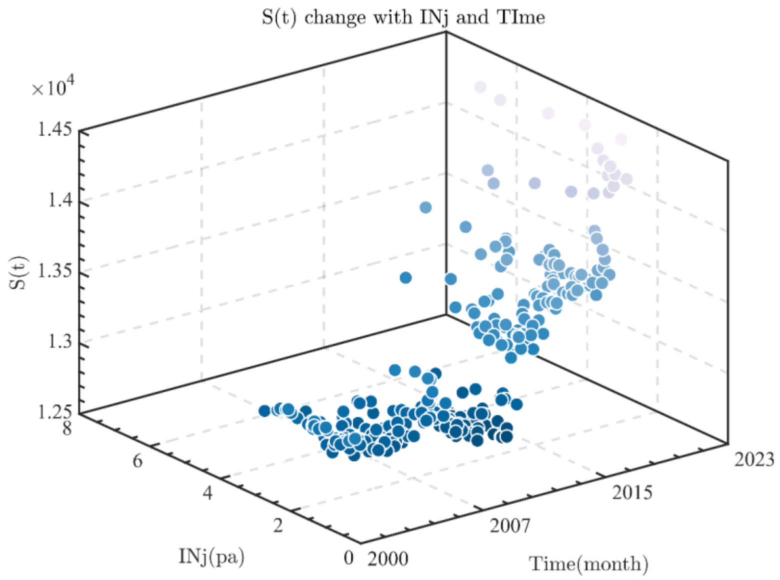

Figure 7. Scatter plot of $S(t)$ change.

The flood risk factor is calculated by differentiating the equation above with respect to time. This factor represents the mathematical relationship between the occurrence of flood disasters and the amount of loss incurred.

We will repeat the above steps to obtain the storm risk factor $f_{flood}(t)$.

Next, we will explore the general patterns of the risk factors $f_{storm}(t)$. We will compare the obtained flood risk factor with the storm risk factor and find a mathematical relationship with the local CPI.

The CPI is a macroeconomic indicator that measures changes in the price level of a basket of goods and services purchased by ordinary consumers. It reflects trends and the extent of price changes in items and services consumed by households in their daily lives.

Next, we will explore the relationship between the risk factors and CPI.

By using the commonly used logarithmic fitting in economics, we can achieve our desired design and determine the constant K, which represents the quantitative mathematical relationship between losses and local CPI over time in Japan. Through calculation and fitting, we obtain $K = 13341.05791$ for the Japanese region.

Referring to literature and combining it with information from official departments, we understand that CPI is subject to interventions and does not have unrestricted growth. After fitting verification, we finally determine the CPI logarithmic model as the quantitative expression of the risk factors.

$$f(t) = K \frac{CPI_l}{CPI_e - ln(k + CPI_s)} \tag{18}$$

CPI_e is the latest monthly CPI value, CPI_s is the initial CPI value since statistics, and CPI_l is the CPI value for the current month.

For computational purposes, let's first take the CPI-related constant k as 1. By calculation, we can obtain the geographical constant $K = 13341.05971$ for the Japanese region. This constant is closely related to the amount of loss incurred.

To evaluate specific locations in the problem, the constant K for extreme weather insurance will vary with different regions since the derived risk factors must be trained based on corresponding national data. The model's universality is enhanced. A study will now be conducted on the Gulf of Mexico to establish the correlation between disasters and geographical factors using the above steps.

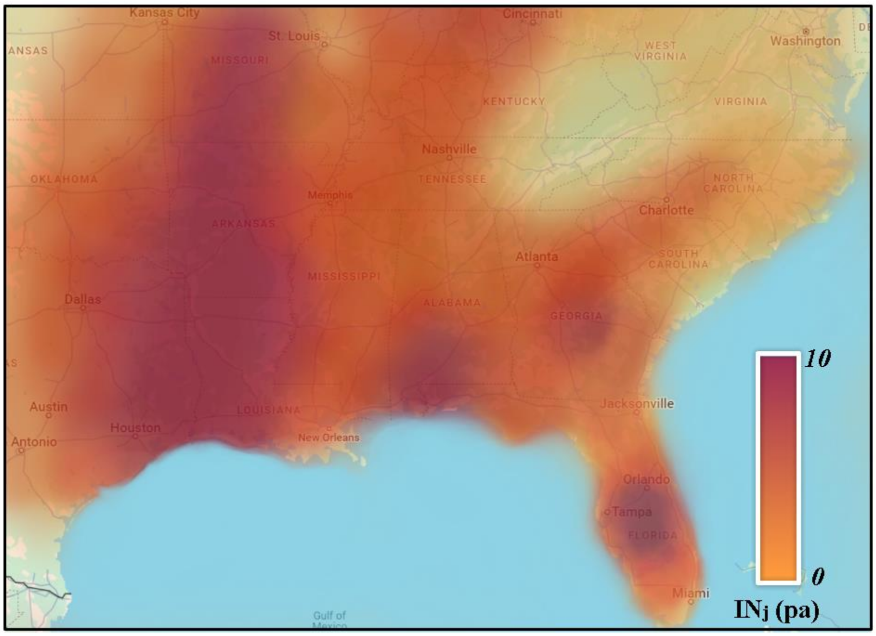

Figure 8. INj geographic heat map.

In addition, we searched for the disaster occurrence and loss data of China's Taiwan, Cuba, and the Bahamas from 1960 to 2022 to demonstrate our model.

We calculated the local extreme weather insurance constant "K" by analyzing the frequency of disasters and losses in the first 50 years. We trained the model using a subset of available data and then predicted the loss situation.

We conducted a differential test between the obtained predicted values and the actual values.

After the differential test, we believe that our model is consistent with the disaster loss situations in China's Taiwan, Cuba, and the Bahamas, this verifies that our model can be generalized

Based on the above model, when the profit function is negative, indicating that the risk factor exceeds a certain threshold, it is not recommended for insurance companies to underwrite policies in such circumstances. When the profit function reaches its maximum value, insurance companies should choose to underwrite the risk at that time.

In conclusion, insurance companies should consider geographical variations in the areas they underwrite and choose the underwriting time wisely. Insurance companies should monitor these changes closely to maintain business stability and ensure sustainable development.

3. Conclusions

In conclusion, the integration of ARIMA and LSTM techniques presents a promising avenue for enhancing risk assessment accuracy and profit optimization within the insurance industry. Through the utilization of historical data, income and loss models, and key risk factors, this combined approach offers a more comprehensive and dynamic framework for evaluating risks and making strategic decisions. By leveraging advanced machine learning algorithms like LSTM, insurers can better anticipate and mitigate potential threats, ultimately fostering resilience and sustainable growth in a volatile economic environment. As the insurance landscape continues to evolve with increasing complexities and uncertainties, embracing innovative methodologies such as the one proposed in this study is essential for staying ahead of emerging risks and maintaining a competitive edge. Moving forward, further research and practical implementations of these techniques hold the potential to transform how insurers operate, adapt, and thrive in an ever-changing risk landscape.

References

[1] Liu Rong. Research on optimization of cyber security insurance coverage in China in the era of digital economy [D]. Guangxi University, 2023.

[2] He Zhongming. Research on optimization of anti-fraud strategies for agricultural insurance at RB Insurance Yibin Branch [D]. Chongqing Technology and Business University, 2023.

[3] Zhu Ming, Yang Ruhua. The impact of reinsurance on agricultural insurance underwriting risks—Discussion on the adjustment effect of competitive advantage [J]. Rural Economy, 2023, (03): 69-77.

[4] He Shuang. Research on risk management of project insurance for large government investment projects [D]. Chongqing University, 2022.

[5] Li Qian. Research on factors influencing the claims ratio of car insurance at XX Property Insurance Company using the random forest algorithm [D]. Hunan University, 2022.

[6] Ni Jie. Research and case analysis on the extension of deposit insurance pricing model based on Merton option pricing framework[D]. Zhejiang Gongshang University, 2022.

[7] Chen Linwei. Study on the impact of export credit insurance on Wuxi's export trade and enhancement strategies [D]. Southeast University, 2021.

[8] Wang Yao. The impact of non-credit factors on the default risk of individual credit loans [D]. University of International Business and Economics, 2021.

[9] Zhang Peipei. Research on optimization of underwriting management of trademark insurance in China [D]. Chongqing University of Technology, 2021. DOI: 10.27753/d.cnki.gcqgx.2021.000043.

[10] Mao Qinjing. Policy orientation of Chinese export credit insurance and its impact on export trade [D]. Southwestern University of Finance and Economics, 2020.