Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 10, No. 2, 2023

Research on the Estimation of The Scale of Catastrophe Insurance Funds in Various Provinces in China

-- A Case Study on Geological Disasters

Jinrui Liu[1], Taotao Zheng[2, *]

[1]Department of Applied Statistics, Zhejiang University of Science and Technology, Hangzhou,CO 310023,China
[2] Faculty of Mathematics, Zhejiang University of Science and Technology, Hangzhou,CO 310023,China
[*] Corresponding author: Taotao Zheng (Email: 1939521048@qq.com)

Abstract: The catastrophe insurance fund, as an important means for the steady and robust development of catastrophe risk management, has attracted significant attention from various sectors of society. However, natural disasters in different regions of China vary significantly in terms of frequency and severity, the centralized management of a uniform catastrophe insurance fund severely affects the efficiency and fairness of catastrophe risk management. This study selects the geological disaster loss data from various provinces of mainland China from 2004 to 2021 as the research sample. Based on the characteristics of relative losses caused by geological disasters, the geological disaster losses in various provinces of China are divided into four categories using cluster analysis, and the Generalized Pareto Distribution (GPD) is utilized to describe the loss distributions of each category; The scale of geological disaster insurance funds in each province is estimated through the optimal reinsurance strategy based on the Conditional Value-at-Risk (CVaR) - Expected Premium principle; Provinces with similar characteristics of relative losses from geological disasters have significantly different requirements for the scale of catastrophe insurance funds. Compared to traditional catastrophe models, this study introduces cluster analysis to consider the heterogeneity of geological disaster occurrences. In this paper, a new approach to enhance the management of catastrophe risks by fitting the distribution of geological disaster losses more reasonably is provided.

Keywords: Geological disasters, Cluster analysis, Generalized Pareto Distribution, Optimal reinsurance, Fund scale.

1. Introduction

China is one of the most severely affected countries by natural disasters globally, facing nearly all types of natural hazards, the frequent occurrence of catastrophic events has resulted in significant economic losses for the society. In recent years, China has established relevant laws and regulations to address the economic losses caused by catastrophic events. In 2006, the State Council of China issued the " Guo Shi Tiao ", which proposed that the government should provide financial support for the catastrophe insurance system. This initiative aimed to further enhance the multi-level agricultural catastrophe risk-sharing mechanism. In 2013, the Central Committee of the Communist Party of China issued the "Decision on Major Issues Concerning Comprehensively Deepening Reforms," aimed at improving the compensation mechanism of insurance economy and the system of catastrophe insurance. In 2014, the "New Ten Measures" were issued, explicitly proposing the establishment of systems such as insurance funds and catastrophe insurance reinsurance, as well as the innovation of a multi-level mechanism for dispersing catastrophe risks. In 2016, the "Outline of the 13th Five-Year Plan for the Development of China's Insurance Industry" proposed the research and establishment of earthquake catastrophe insurance funds, as well as the development of insurance systems covering multiple perils such as floods and typhoons. Up to now, China's catastrophe insurance has taken initial shape and demonstrated good operational effectiveness. On one hand, we have established a scientific operational model and formulated a market-based mechanism for loss sharing. On the other hand, we have innovatively developed numerous catastrophe insurance products. However, based on the current situation of catastrophe insurance payouts in China, there are evident issues in the management of catastrophic risks. Statistics indicate that the insurance payout rate for the 2008 Wenchuan earthquake was approximately 0.28%, while the insurance payout rate for the massive rainfall in Henan in 2021 was approximately 11%. Although the insurance claims payout rate has increased, there is still a significant gap when compared to the world average payout rate of 30% to 40%. Not to mention compared to the payout rates exceeding 60% in developed countries.

To better manage catastrophic risks, numerous scholar have conducted a series of meaningful research studies. Regarding the estimation of the scale of catastrophe insurance funds, the majority of scholars believe that the primary function of catastrophe insurance funds is to assume risks beyond the compensation limits set by insurance in various provinces. Therefore, the relevant literature primarily revolves around the underwriting capacity of insurance in various provinces and the maximum losses that catastrophic events may cause. Kremer provided a mathematical formulation for the general results of the potential maximum losses[1]. Ouyang Zisheng utilized Pareto distribution and lognormal distribution to fit the segmented geological disaster losses, furthermore, the researcher estimated the pure premium and the maximum potential losses for geological disasters[2]. Tian Ling et al. used Value-at-Risk (VaR) to estimate the scale of China's earthquake insurance funds from the perspective of optimal retention, the study indicated that the size of earthquake insurance funds is significantly influenced by risk tolerance and the additional premiums from reinsurance[3]. Geng Guizhen et al. selected data from the past fifty years to validate that the Peaks-over-Threshold (POT) - GPD model has a good fitting effect and accuracy for earthquake catastrophes[4]. Tian Ling et al. used CVaR as an optimization indicator for risk allocation in the reinsurance market to estimate the scale of earthquake catastrophe funds[5]. Yang Liang et al. approached the estimation of the scale of catastrophe insurance funds from the perspective of optimal reinsurance based on VaR - Expected Premium, they overcame the limitations of setting premium parameters manually and explored the impact of risk tolerance on reinsurance[6].

However, an increasing number of scholars argue for differentiated management of catastrophe insurance funds. Qiu Bo et al. believe that the differences in hazards and disaster bearing entities determine the need for "location-based fundraising" for catastrophe insurance funds, this approach can also mitigate moral hazards to a certain extent[7]. Tu Haiping, in the design of fund allocation, divided the country into different risk levels and proposed that fund allocation should adhere to the principle of "heavier burden on areas with severe losses and economically developed regions"[8]. Tian Ling et al. suggested that the payout ratio of earthquake catastrophe insurance (EQΠ) should be designed separately based on the specific characteristics of seismic disasters in each region[9]. However, currently, the fitting of catastrophe loss curves is based on nationwide data, which makes it difficult to reflect the quality differences in bearing catastrophe risks among provinces. This approach is unfair and inefficient in terms of catastrophe risk management. Therefore, this study incorporates cluster analysis into existing research to explore the similarities and differences in catastrophe losses among provinces, and subsequently quantitatively analyze the catastrophe insurance funds of each province. Seeking similarity in the development level among individuals and similarity in development trends are key aspects of cluster analysis. Li Yinguo et al. proposed using "absolute indicators","incremental indicators" and "volatility indicators" to construct a distance function for measuring similarity, they applied the Ward's method for cluster analysis[10]. Ma Junjie et al. incorporated the surface spatial volume indicator into the Li causality feature extraction, considering the distances between individuals in three-dimensional space[11]. Liu Yunxia proposed a dynamic time warping-based panel data clustering method. The method yields stable clustering results but comes with a higher computational complexity[12]. Wang Zedong et al. proposed a cluster analysis method for panel data based on trend distance, they applied symbolization to the data and measured the similarity of trend distances using the symbolized data[13].

This study aims to address the differentiated needs of regional risk management by employing symbolized time series processing and considering the influence of outliers to classify the relative losses of geological disasters in various provinces of China, and to construct different loss distribution functions. Building upon this, we introduce the principles of optimal reinsurance, taking into account the robustness of the CVaR reinsurance model and consistency of risk measures. We employ the CVaR-Expected Premium approach to calculate the geological disaster fund size for each province from a risk-minimization perspective. This approach aims to actively establish catastrophe risk management systems that align with the specific circumstances of each province.

2. Theoretical Model

2.1. Cluster Analysis

Hierarchical clustering is a commonly used clustering method that treats each sample or vector to be clustered as an individual group, then, an appropriate similarity measure is used to merge the closest two or several classes into a new class. This process is repeated iteratively until all samples or variables are merged into a single category. In this study, a comprehensive distance measure is used as a similarity statistic, and the clustering is performed using the squared Euclidean distance method.

Given two time series. $S_1=\{s_{11},s_{12}\ldots s_{1i}\ldots s_{1m}\},S_2=\{s_{21},s_{22}\ldots s_{2i}\ldots s_{2n}\}$,the corresponding symbolized sequences are as follows. $Q_1=\{q_{11},q_{12}\ldots q_{1i}\ldots q_{1m-2}\},Q_2=\{q_{21},q_{22}\ldots q_{2i}\ldots q_{2n-2}\}$,Since the elements in the symbolized sequences are represented by letters and cannot directly participate in calculations. Therefore, it is necessary to ensure that the interval between adjacent letters is the same in order to calculate the trend distance of the symbolized sequences.

Definition 1: Distance between two time series of real data

$$D_1=\sum_{i=1}^{n}\frac{\left(s_{1i}-s_{2i}\right)^2}{S_{max}^2}, \tag{1}$$

where $S_{\max}^2=\max\left\{\left|\max\left(S_1\right)-\min\left(S_2\right)\right|,\left|\min\left(S_1\right)-\max\left(S_2\right)\right|\right\}$.

Definition 2: Trend distance between two symbolized time series

$$D_2=\sum_{i=1}^{n-2}\frac{\left(q_{1i}-q_{2i}\right)^2}{Q_{max}^2}, \tag{2}$$

where $Q_{\max}^2=\max\{|\max(Q_1)-\min(Q_2)|,|\min(Q_1)-\max(Q_2)|\}$.

Definition 3: Comprehensive distance between two time series

$$D_3=\sqrt{D_1+\frac{n}{n+2}D_2}, \tag{3}$$

The details of proof are omitted here due to the limitation of space. More details can be found in Appendix .

The calculation formula for trend distance still incorporates the range values of the sequence, considering that outliers in the sequence can affect the trend distance. In this study, an improvement is made to the Q_{max}^2 in D_2 by excluding the influence of outliers, which is defined as the maximum difference between values in the sequence after removing the outliers. Due to the small sample size of the data, the outlier filtering method used is based on values that fall beyond two times the interquartile range (IQR) below the lower quartile or above the upper quartile in the box plot. To preserve the absolute nature of the original data distances, we only made improvements to the trend distance and did not modify the distances of the real sequences.

2.2. CVaR Estimation based on the Generalized Pareto Model

Compared to the fitting performance of the lognormal distribution and the standard Pareto distribution, the GPD is more suitable for fitting loss data with heavy-tailed characteristics, particularly for simulating loss data exceeding a certain threshold[14]. Considering the significant heavy-tailed characteristics of geological disaster loss data in China, we will employ the GPD model to simulate loss data exceeding a certain threshold.

2.2.1. Definition of the GPD Model

$$G_{\varepsilon,\beta}(x)=\begin{cases}1-\left(1+\varepsilon\dfrac{x}{\beta}\right)^{-\frac{1}{\varepsilon}} & \varepsilon\neq 0,\\ 1-\exp\left(-\dfrac{x}{\beta}\right) & \varepsilon=0,\end{cases} \tag{4}$$

where β is the scale parameter of the distribution, and ε is the shape parameter. When $\varepsilon\geq 0$, the GPD exhibits heavy-tailed behavior. Various methods can be employed to estimate the parameters of the GPD, this includes methods such as maximum likelihood estimation and probability-weighted moment method. In this study, the maximum likelihood estimation method is employed to obtain an estimate of ε,β. The advantages of maximum likelihood estimation can be referred to for further details[15].

2.2.2. Overall Distribution Function of the Loss Data

Usually, the Empirical Average Exceedance Function and the Hill plot are commonly used methods to select the threshold value. In this study, we employ the simplest and most effective method, the Empirical Average Exceedance Function, to determine the threshold value, and perform a GPD test on the loss data exceeding the threshold value. Let μ be a threshold value and $X-\mu$ be the excess loss, the excess loss distribution is defined as the distribution of the losses exceeding the threshold:for $0\leq y<x_0-\mu,\ F_\mu(y)=P(X-\mu\leq y\mid X>\mu)$: upon computation, we obtain $F_\mu(\mathrm{y})=\left[F(\mathrm{y}+\mu)-\mathrm{F}(\mu)\right]/\left[1-F(\mu)\right]$. For the excess loss distribution, the following limit theorems hold: For a wide range of distributions(almost including all commonly used) distributions of excess loss distribution function $F_\mu(\mathrm{y})$, there exists a function $\beta(\mu)$ such that:

$$\lim_{\mu\to x_0}\sup_{0\leq y<x_0-\mu}\left|F_\mu(\mathrm{y})-G_{\varepsilon,\beta(\mu)}(\mathrm{y})\right|=0. \tag{5}$$

Let N_μ be the number of sample data points exceeding the threshold μ, the empirical distribution function is used to replace $F(\mu)$, that is $F(\mu)=1-N_\mu/n$. Therefore, when $\varepsilon\neq 0$, the distribution function of the overall loss data can be represented as follows:

$$F(x)=1-\frac{N_\mu}{n}\left[1+\frac{\varepsilon}{\beta}(x-\mu)\right]^{-\frac{1}{\varepsilon}}. \tag{6}$$

2.2.3. Estimation of CVaR

VaR is the maximum potential loss that a financial asset or portfolio may experience within a specified future period, at a certain confidence level. However, the VaR method does not consider the tail risk beyond the VaR value and fails to satisfy sub-additivity and consistency with risk measurement. Therefore, to overcome the limitations of the VaR method, this study employs the CVaR method as a risk measure. CVaR refers to the average loss value conditional on the loss exceeding a given VaR value. In the case of a confidence level of p, we can obtain the expression for CVaR as follows:

$$VaR_p=\mu+\frac{\beta}{\varepsilon}\left(\frac{n}{N_\mu}(1-p)\right)^{-\varepsilon}-1, \tag{7}$$

$$CVaR_p(x)=\frac{1}{1-p}\int_{VaR_p}^{+\infty}xf(x)\,dx. \tag{8}$$

2.3. Optimal Reinsurance

Reinsurance refers to the practice where an insurance company, in order to diversify the risks and liabilities it undertakes, enters into a reinsurance contract with other insurance companies based on the original insurance contract, the insurance company pays a certain premium to the reinsurance company. In this study, the expected premium criterion is used to calculate the reinsurance premium.

2.3.1. Reinsurance Premium Criterion

$$r(h(X))=(1+\theta)E(h(X)), \tag{9}$$

where X represents the losses borne by the original insurance company, h(X) denotes the losses transferred to the reinsurer, and θ represents the premium parameter.

2.3.2. Distorted Risk Measures and Reinsurance Strategies

Distorted risk measures refer to the transformation of the original probability distribution using distortion functions, thereby altering the way risk is measured[17]. Due to its ability to better incorporate information about catastrophic risks, distorted risk measures have been widely used in the study of catastrophic risks. The general expression for distorted measures is as follows:

$$\rho_g(x)=\int_0^{+\infty}g\left[p(X>x)\right]dx=\int_0^{+\infty}g\left[S(x)\right]dx, \tag{10}$$

where $g(x)$ is the distortion function defined on the interval [0,1] and satisfies $g(0)=0$ and $g(1)=1$ as an increasing function; $S(x)$ is the survival function of the random variable.

2.3.3. Optimal Reinsurance Strategy without Premium Constraint

When engaging in reinsurance, the original insurance company needs to consider the relationship between the premiums paid and the risks transferred to the reinsurer, moreover, it is necessary to determine the premium parameter that minimizes the risk while satisfying the minimum premium requirement. Therefore, the following relationship needs to be satisfied to determine the premium parameter[6][18]:

$$\begin{cases}\min\left\{\mu_r\left[h^*(X)\mid\theta\right]\right\}=\int_0^{+\infty}\min\left\{r\left[S_X(t)\right],g\left[S_X(t)\right]\right\}dt,\\ h^*(X)=\underset{h\in F}{\arg\min}\left\{\rho_g\left[T_h(X)\right]\right\},\end{cases} \tag{11}$$

where $T_h(X)=X-h(X)+r\left[h(X)\right]$, as the total risk underwritten by the insurance company.

2.3.4. The conditional maximum loss for the insurance company

Based on existing research[18], it has been proven that under the criteria of expected premium and CVaR risk measure, stop-loss reinsurance is always optimal.where $d=S_X^{-1}\left(\frac{1}{1+\theta}\right)$,the potential maximum loss condition is given by:

$$\begin{cases} CVaR_\alpha\left[T_h(X)\right] = CVaR_\alpha(X) + \left[(1+\theta) - 1/\alpha\right]E\left[max\{(x-d),0\}\right], \alpha < 1/(1+\theta), 1-F_X(d) \le \alpha, \\ CVaR_\alpha\left[T_h(X)\right] = d + (1+\theta)E\left[max\{(x-d),0\}\right], \alpha < 1/(1+\theta), 1-F_X(d) > \alpha, \\ CVaR_\alpha\left[T_h(X)\right] = CVaR_\alpha(X), \alpha > 1/(1+\theta), \end{cases} \tag{12}$$

2.4. Estimation of Fund Size for Each Province

In this study, we utilize CVaR to measure the size of the catastrophe insurance fund. Taking the loss data of the i-th cluster of provinces from the results of the clustering analysis as an example, we estimate the fund size for these provinces. We employ the GPD to fit the loss distribution of the data exceeding the threshold value, the expression for calculating the catastrophic fund of province j in the i-th cluster of catastrophic provinces is derived. Let's assume that $\alpha < 1/(1+\theta), 1-F_X(d) > \alpha$,where α represents the significance level and x represents the relative loss data, namely, "direct economic losses from geological disasters / regional GDP", under the condition of CVaR as a risk measure, the optimal allocation function and the distortion risk measure function are given by:

$$\begin{cases} h(x) = \min\left\{\left[x - S_X^{-1}(1/(1+\theta))\right], 0\right\}, \\ g(x) = \min(x/\alpha, 1). \end{cases} \tag{13}$$

Based on the optimal reinsurance strategy, the premium parameter θ that minimizes the risk is calculated to satisfy:

$$\int_{S_X^{-1}\left(\frac{1}{1+\theta}\right)}^{+\infty} x f_i(x) dx = \frac{S_X^{-1}\left[1/(1+\theta)\right]}{1+\theta}. \tag{14}$$

The expression for calculating the catastrophe fund for province j in the i-th cluster of provinces, based on CVaR, is given by:

$$\begin{cases} CVaR_\alpha^i\left[T_h(X)\right] = S_X^{-1}\left[1/(1+\theta_i)\right] \\ \quad + (1+\theta_i)E\left\{max\left[\left(x - S_X^{-1}(1/(1+\theta_i))\right), 0\right]\right\}, \\ M_j^i = a_j * CVaR_\alpha^i\left[T_h(X)\right] * \frac{w_j}{\sum_{k=1}^{n} w_k}, \end{cases} \tag{15}$$

where a_j represents the GDP of province j in a specific year, and w_j represents the total relative loss of geological disasters for the data samples in province j; where $\sum_{k=1}^{n} w_k$ represents the sum of total relative losses of geological disasters for all provinces in the i-th cluster, and M_j^i represents the catastrophe insurance fund for province j in the i-th cluster in a specific year.

3. Empirical Analysis

3.1. Data Description and Preprocessing

The data used in this study are obtained from the annual "China Statistical Yearbook" and "China Environmental Statistical Yearbook" for multiple years. The data consists of annual direct economic losses from geological disasters and regional GDP for each province in mainland China. The time period for the data is from 2004 to 2021, and the cross-sectional data includes 31 provinces in mainland China. On the one hand, there are significant differences in economic development levels among different provinces, which can have an impact on the statistics of direct economic losses from geological disasters. On the other hand, the impact of direct economic losses from geological disasters on provinces is contingent upon their respective levels of economic development, to account for the influence of these two aspects, this study uses the relative indicator of "direct economic losses from geological disasters / regional GDP" as a substitute for the absolute indicator of "direct economic losses from geological disasters" to investigate and analyze the extent of geological disaster losses among provinces.

Table 1. Selected Indicators

	Indicators
Geological disasters	Direct economic losses from geological disasters (in thousands of yuan)
Level of economic development	Regional GDP (in billions of yuan)

During the data collection process, we observed that there were missing data on geological disaster occurrences for certain provinces in certain years. Therefore, we believe that the geological disaster situations in these provinces during these years did not meet statistical standards. The missing data for these provinces are replaced with zeros. Additionally, we found that the data on geological disaster occurrences in Shanghai were missing to a significant extent, therefore, this study does not include an analysis of geological disaster occurrences in Shanghai.

3.2. Cluster Analysis Results

In this study, we employed the hierarchical cluster analysis method, which takes into account both the true distance and the trend distance, to conduct the clustering. In terms of the selection of the number of clusters, we found that four clusters were appropriate for this analysis. The clustering results are presented in the following figure:

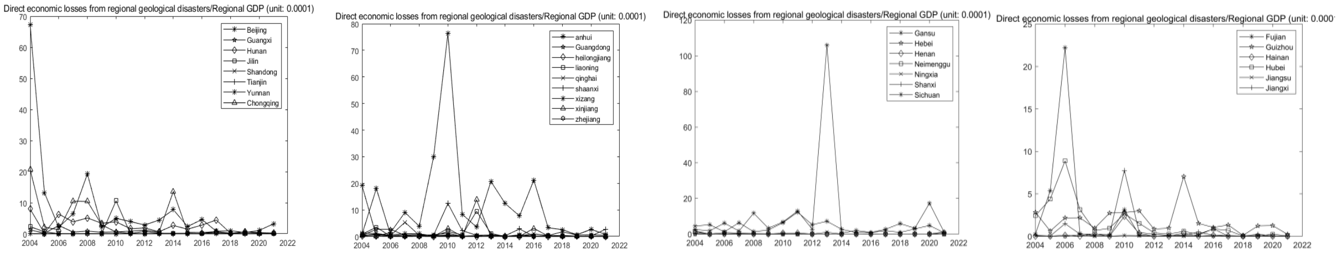

Figure 1: Classification Results (Top Left, Top Right, Bottom Left, Bottom Right correspond to Category 1, Category 2, Category 3, Category 4, respectively)

From the above classification results, it can be observed that each cluster largely preserves the similarity in terms of geological hazard relative loss levels and relative loss trend, furthermore, it can be visually observed that each cluster contains outlier information regarding the geological hazard relative loss of provinces. This can provide insights into estimating the likelihood of major geological hazards occurring in each cluster's provinces.

3.3. Descriptive statistical features

By analyzing the descriptive statistical measures of relative loss data for each category, the kurtosis and skewness values of the sample data are significantly greater than the kurtosis and skewness values of a normal distribution. The data intuitively reflects the characteristics of the data showing sharp peaks, thick tails, and right deviation.

Table 2. Basic Statistics of Relative Loss Data for Four Categories

	Mean	Standard	Kurtosis	Skewness
Category 1	1.9930	6.2492	75.565	7.862
Category 2	2.1464	9.7348	105.397	9.904
Category 3	2.0980	7.2910	68.764	7.428
Category 4	1.0242	2.5914	42.788	5.819

3.4. Selection of the threshold and testing of the GPD model

The Empirical Average Exceedance Plot is a simple and effective method for selecting the threshold value,the principle is that if the data follows a GPD, the portion exceeding the threshold value should exhibit a clear linear growth with a positive slope. Therefore, by observing the trend of the Empirical Average Exceedance Plot, an appropriate threshold can be selected to test whether the data follows a GPD.

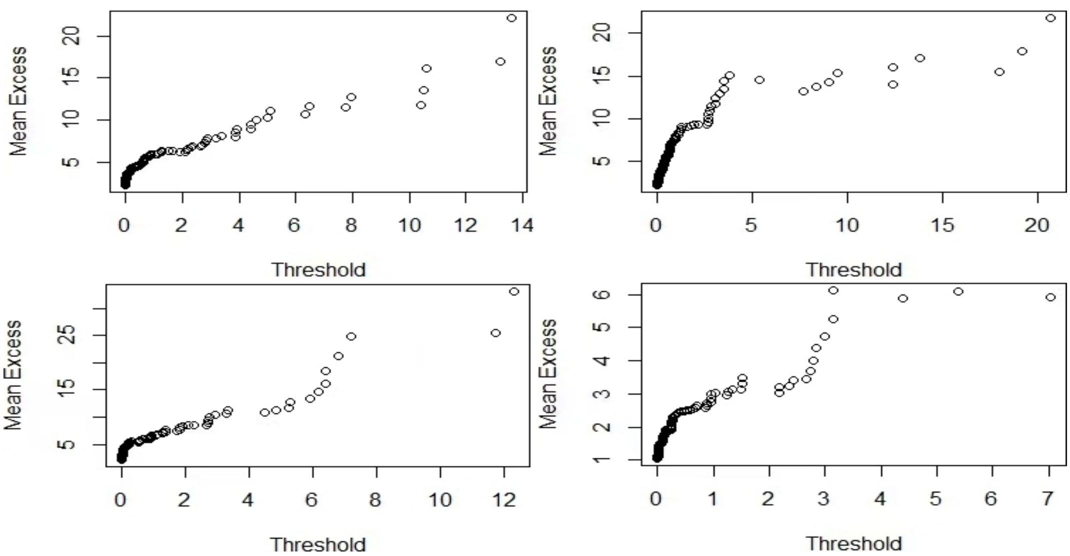

Figure 2. Empirical Mean Excess Function Plot for Four Categories of Relative Loss Data.

Based on the linear characteristics of the Empirical Average Exceedance Plot, and ensuring the linearity of the exceedance function and stability of the distribution parameter estimation, the selected threshold values are 4.8924,9.4800,5.9003 and 2.9807 respectively. When the exceedance exceeds the threshold value, The empirical exc- eedance function graph exhibits a clear linear trend, Ther- efore, the GPD can be used to fit the loss distribution of the exceedance data.

3.5. Calculation of Catastrophe Insurance Fund Size for Each Province

Under the metrics of CVaR and expected premium, a risk tolerance of 5% is chosen. Based on the principle of minimizing risk measures, the premium parameter β are calculated for each category of provinces as follows : 0.00027,0.000226,0.00051,0.00094. Based on this, the estimated sizes of the catastrophic insurance funds for each province are as follows:

Table 3. Geological Disaster Insurance Fund Size for Each Province. (Unit: thousands of yuan)

Region	$CVaR_{\alpha}\left[T_h(x)\right]$	Region	$CVaR_{\alpha}\left[T_h(x)\right]$	Region	$CVaR_{\alpha}\left[T_h(x)\right]$
Sichuan	378231.1	Guizhou	54053.88	Inner Mongolia	8034.363
Yunnan	288736.7	Liaoning	39338.89	Jiangsu	5862.289
Gansu	200399.1	Tibet	35284.25	Jiangsu	4190.128
Hunan	158302	Jiangxi	30287.06	Hainan	2036.991
Fujian	135201.6	Xinjiang	26640.57	Hebei	1524.43
Chongqing	134883.4	Zhejiang	20981.63	Ningxia	790.7257
Hubei	117929.8	Guangxi	17775.16	Shandong	746.4658
Anhui	79265.51	Henan	15823.1	Beijing	276.5931
Guangdong	78958.73	Jilin	13548.96	Heilongjiang	229.4369
Shaanxi	60462.2	Shanxi	10128.83	Tianjin	25.44016

The above results indicate that the provinces of Sichuan,Yunnan,Gansu, Hunan, Fujian,Chongqing and Hubei are characterized by higher geological disaster risks. The provinces of Ningxia, Shandong, Beijing, Heilongjiang, Tianjin and Hebei are characterized by lower geological disaster risks. The top five provinces in terms of direct economic losses from geological disasters, based on the National Statistical Yearbook from 2014 to 2019, are as follows:Sichuan:16 occurrences, Hunan:15 occurrences, Yu- nnan:12 occurrences, Gansu:8 occurrences, Hubei:6 occurre- nces, The provinces with the least occurrences are as follows: Ningxia:12 occurrences, Shandong:11 occurrences, Beijin- g:11occurrences, Hebei:9 occurrences, Tianjin:7 occurrences, The above analysis results are consistent with the actual situation.

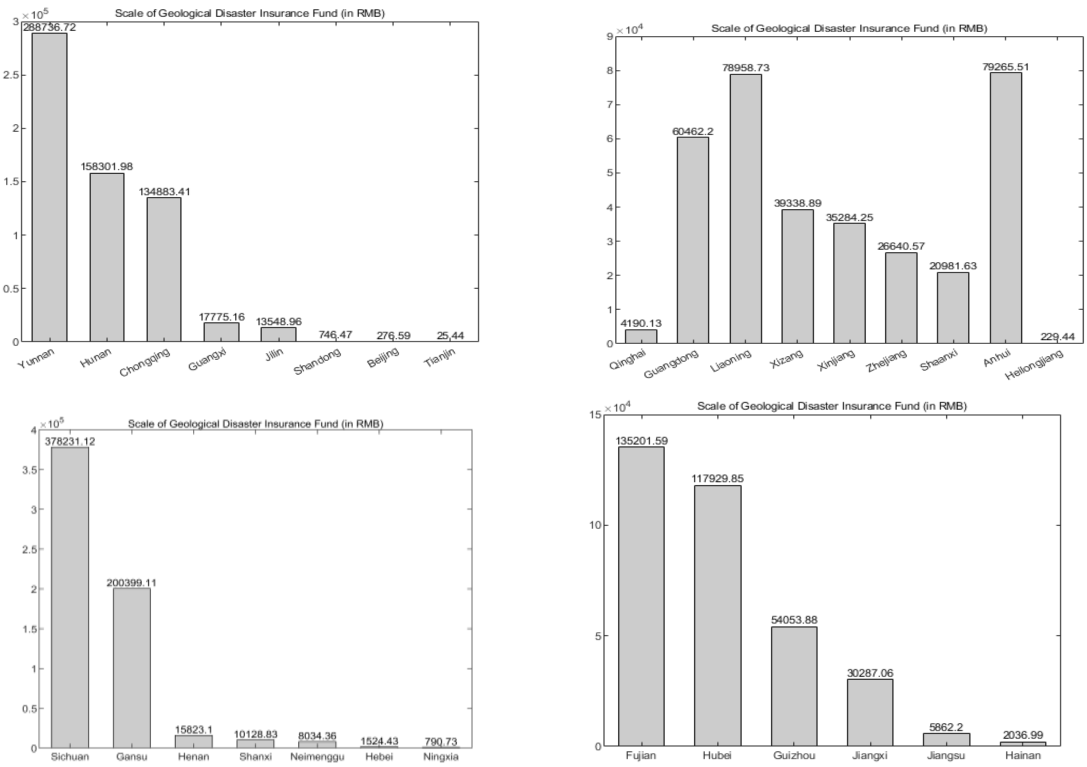

Figure 3. Geological Disaster Insurance Fund Size for Four Categories of Provinces

The geological disaster insurance fund sizes for each province are summarized according to the classification results, It can be observed that due to the varying levels of economic development among provinces, even among provinces of the same category with similar characteristics in terms of relative geological disaster losses, the absolute values of direct economic losses can still vary significantly. Therefore, if we consider jointly insuring the geological disasters of provinces within each category, not only can we promote the implementation of the geological disaster risk management model, but we can also achieve a certain level of risk diversification.

Different risk tolerance levels can lead to different CVaR values, which in turn affect the size of the geological disaster insurance fund. The following table provides examples of fund sizes for one province selected from each category of provinces, It can be observed that there are significant differences in fund sizes under different risk tolerance levels, as the risk tolerance level decreases, the size of the geological disaster fund increases continuously.

Table 4. Selected fund size results for four province (Unit : thousands of yuan)

Region	$CVaR_{0.9}$	$CVaR_{0.95}$	$CVaR_{0.99}$
Beijing	194.91	276.5931	899.46
Anhui	45308.27	79265.51	199007.19
Gansu	129979.56	200399.1	266542.04
Fujian	85183.19	135201.6	323750.92

4. Conclusion

Based on the results of cluster analysis and following the principle of optimal reinsurance, this article calculates the scale of geological disaster insurance funds for each province and draws the following conclusions.

First, based on the results of the cluster analysis mentioned above, it can be observed that different provinces exhibit certain similarities in terms of relative losses and trends in loss variation in geological disasters. This indicates that we can scientifically classify catastrophic risks in different regions, enabling precise management tailored to the risk conditions and economic development levels of each region. Consequently, effective control and management of risks can be achieved.

Secondly, by fitting the geological disaster loss distribution functions of each province using the Generalized Pareto Distribution, and then calculating the scale of the geological disaster insurance funds for each province based on the principle of optimal reinsurance, it helps to improve the efficiency and fairness of risk management.

Thirdly, provinces with similar relative losses in geological disasters have significant differences in their requirements for the scale of geological disaster insurance funds. If continental provinces are classified and combined for underwriting, it not only enables effective promotion of risk management models but also facilitates risk diversification.

The catastrophic insurance fund, as the core of catastrophic risk management, not only alleviates the financial burden on the government but also promotes the development of the catastrophic insurance market. This article's regional classification and research on the catastrophic insurance fund have significant practical significance for catastrophic risk management. However, China is still in the exploratory stage of catastrophic risk management, and further exploration is needed on how to diversify the management of catastrophic risks.

5. Appendix

5.1. Symbolization of Time Series

In this study, we adopted the symbolization method proposed by Deng Guangming (2019) for data symbolization. Division of Time Series into Nine Trend Patterns, The nine trend patterns are as follows: continuous decline, decline followed by stability, trough, stability followed by decline, continuous stability, stability followed by growth, peak, growth followed by stability, and continuous growth. These patterns represent different temporal behaviors and trends observed in the time series data. These nine trend patterns are sequentially labeled as A through I.

Given a time series $Z=\{Z_1,Z_2,...Z_i,...Z_n\}(\mathrm{i}=1,2...\mathrm{n})$, where n is the length of the time series. Define:

$$R(\mathrm{Z,i})=\frac{Z_i-Z_{i+1}}{X_{\max}}, \tag{16}$$

where $X_{\max}=\max(|Z_i-Z_{i+1}|)(\mathrm{i}=1,2...\mathrm{n})$, In the encoding process, the first and last values of the sequence are not included in the calculation. So, the length of the encoded sequence is n-2. The threshold value for trend interval division is denoted as ε. In practical applications, it is necessary to continuously adjust the threshold value in order to better showcase the changing trends of the time series.

Table 5. Symbolic representation of time series data.

	$R(Z,i)>\varepsilon$	$\lvert R(Z,i)\rvert\le\varepsilon$	$R(Z,i)<-\varepsilon$
$R(Z,i)>\varepsilon$	A	B	C
$\lvert R(Z,i)\rvert\le\varepsilon$	D	E	F
$R(Z,i)<-\varepsilon$	G	H	I

5.2. Trend distance

Given two time series $X_1=\{x_{11},x_{12}\dots x_{1i}\dots x_{1m}\}(i=1,2\dots m)$, $X_2=\{x_{21},x_{22}\dots x_{2i}\dots x_{2n}\}(j=1,2\dots n)$,Their trend distance is defined as follows:

$$Dist(0,0)=0;$$
$$Dist(i,0)=Dist(i-1,0)+\mu_d;$$
$$Dist(0,j)=Dist(0,j-1)+\mu_j;$$
$$Dist(i,j)=\min\{Dist(i-1,j)+\mu_d+Dist(i,j-1)+\mu_j,Dist(i\text{-}1,j-1)+\mu_r(i,j)\};$$
$$TD(X_1,X_2)=\mathrm{Dist}(M,N);$$

where the values of A and B are both 1. $\sigma_{\max}=\max\{|\max(X_1)-\min(X_2)|,|\min(X_1)-\max(X_2)|,|\min(X_1)-\max(X_2)|,|\min(X_1)-\max(X_2)|\},\mu_r(i,j)=|(x_{1i}-x_{2j})|/\sigma_{\max}$.

References

[1] E. Kremer. More on the Probable Maximum Loss.Bltter Der Dgvfm, vol. 21 (1994) , p.319- 326.

[2] Z.S. Ouyang. Fitting of Geological Hazard Loss Distribution and Risk Measurement. Statistical Research, vol. 11 (2011) , p.78-83.

[3] L. Tian, P.Yao. Study on the Scale of Catastrophe Insurance Fund in China - Taking Earthquake Risk as an Example. Insurance Research, vol. 4 (2013) , p.13-21.

[4] G. Geng, H.Y. Wang. Study on the Loss Distribution of Earthquake Catastrophes Based on the POT-GPD Model. Journal of Natural Disasters, vol. 3 (2016) , p.153-158.

[5] L. Tian, Y.L. Wu, X.C. Shen. Estimation of Earthquake Catastrophe Insurance Fund Size Based on CVaR. Economic Review, vol. 4 (2016) , p.141-150.

[6] L. Yang, S.W. Meng, T.T. Xu. Study on the Calculation of Catastrophe Insurance Fund Size in China. Statistical and Information Forum, vol. 3 (2022) , p.75-85.

[7] B. Qiu, T.T. Xu. Regional Differences in the Funding Standards of Agricultural Catastrophic Risk Funds in China: Based on Provincial-level Crop Loss Data. Journal of Ningbo University (Humanities and Social Sciences), vol. 3 (2015) , p.74-78.

[8] Y. Zhou, H.P. Tu. Design of Cross-Regional Typhoon Catastrophic Risk Insurance Fund. China Soft Science, vol. 6 (2017) , p.69-80.

[9] L. Tian, N. Sun, C. Yang. Pareto Optimal Indemnity Ratio in Earthquake Index Insurance: A Case Study of EQ II Product. Insurance Research, vol. 6 (2019) , p.39-50.

[10] Y.G. Li, Y. Dai, X.Q. He. Panel Data Clustering Method Based on Adaptive Weights. Systems Engineering Theory and Practice, vol. 2 (2013) , p.388-395.

[11] J.J. Ma, G.F. Ma, F.Y. Bai, J.X. You, R. Lu. Analysis of Provincial Patent Output in China Based on Fuzzy C-means Panel Data Clustering. Systems Engineering Theory and Practice, vol. 9 (2015) , p.2304-2314.

[12] Y.X. Liu. Research and Application of Panel Data Clustering Method Based on Dynamic Time Warping. Statistical Research, vol. 11(2016) , p.93-101.

[13] Z.D. Wang, G.M. Deng. Discussion on Panel Data Clustering Method Based on Trend Distance. Statistics and Decision Making, vol. 8 (2019) , p.35-38.

[14] A.J. Mcneil. Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory.Astin Bulletin, vol. 1 (1997) , p.117－137.

[15] S. Kang, J. Song. Parameter and Quantile Estimation for the Generalized Pareto Distribution in Peaks over Threshold Framework. Korean Statist. Soc. vol. 4 (2017) , p.487 -501.

[16] F.M. Longin. From Value at Risk to Stress Testing:The Extreme Value Approach. J. Banking & Finance, vol. 24 (2000) , p.1097-1130.

[17] S.S. Wang. A Class of Distortion Operators for Pricing Financial and Insurance Risks. Journal of Risk and Insurance, vol.1 (2000) , p.15-36.

[18] W. Cui, J.P. Yang, L Wu. Optimal Reinsurance Minimizing the Distortion Risk Measure under General Reinsurance Premium Principles.Insurance: Mathematics and Economics,2013, p.49-56.

[19] C.Z. Liu, W.Z. Shen, S.Huang. Strategic Considerations on China's Geological Disaster Prevention and Response, vol. 3 (2022) , p.1-4.

[20] X. Pei. International Experience and Lessons for the Development of Catastrophe Insurance.Chinese insurance, vol. 9 (2021) , p.55-59.

[21] L. Tian, J.Y. Peng, Z.W. Wang. Calculation of Catastrophe Insurance Fund Size in China under the Maximization of Underwriting Capacity. Insurance Research, vol.11 (2013), p.24-31.

[22] C. Wang. Factors Influencing Residents' Willingness to Purchase Geological Disaster Insurance--Taking residents of Lanzhou city community as an example. (Ph.D., Lanzhou University,China 2017), p.71.

[23] C.Q. Wang. Analysis of Difficulties in Catastrophe Insurance and Countermeasures. Fortune Life, vol. 10 (2022) , p.10-12.

[24] Q.B. Zhao, Y. Hou. Catastrophe Risk Management in Mature Non-life Insurance Markets and the Development Status in China. Insurance Theory and Practice, vol. 6 (2021) , p.87-98.

[25] Y.C. Chi, K.S. Tan. Optimal Reinsurance under VaR and CVaR Risk Measures.a Simplied Approach, vol.2 (2013) , p.487-509.