



Estimation of the Extreme Distribution Model of Economic Losses Due to Outbreaks Using the POT Method with Newton Raphson Iteration

Riza Andrian Ibrahim^{1*}, Sukono², Riaman³

^{1,2,3} *Department of Mathematics, FMIPA, Universitas Padjadjaran
Jl. Raya Bandung-Sumedang Km 21, Jatinangor 45363, Jawa Barat, Indonesia*

**Corresponding author email: riza17005@mail.unpad.ac.id, sukono@unpad.ac.id, riaman@unpad.ac.id*

Abstract

Extreme distribution is the distribution of a random variable that focuses on determining the probability of small values in the tail area of the distribution. This distribution is widely used in various fields, one of which is reinsurance. An outbreak catastrophe is non-natural disaster that can pose an extreme risk of economic loss to a country that is exposed to it. To anticipate this risk, the government of a country can insure it to a reinsurance company which is then linked to bonds in the capital market so that new securities are issued, namely outbreak catastrophe bonds. In pricing, knowledge of the extreme distribution of economic losses due to outbreak catastrophe is indispensable. Therefore, this study aims to determine the extreme distribution model of economic losses due to outbreak catastrophe whose models will be determined by the approaches and methods of Extreme Value Theory and Peaks Over Threshold, respectively. The threshold value parameter of the model will be estimated by Kurtosis Method, while the other parameters will be estimated with Maximum Likelihood Estimation Method based on Newton-Raphson Iteration. The result of the research obtained is the resulting model of extreme value distribution of economic losses due to outbreak catastrophe that can be used by reinsurance companies as a tool in determining the value of risk in the outbreak catastrophe bonds.

Keywords: Extreme Distribution, outbreak catastrophe, Extreme Value Theory, Peaks Over Threshold, Maximum Likelihood Estimation, Newton-Raphson iteration

1. Introduction

Extreme events are events that rarely occur, but can cause very large losses. This incident can occur in various fields such as health, finance, agriculture and others. Examples are outbreak catastrophe, monetary crises, death of family head, crop failure, and so on. One tool that can be used to explain the risk of these events is the extreme distribution. The extreme distribution is the distribution of a random variable which is only limited to values that have a small probability of occurrence. In other words, the determination of risk using this distribution only focuses on the tail area of the distribution. Extreme distribution is widely used in various fields, one of which is the field of reinsurance. To prevent bankruptcy, reinsurance companies must determine the amount of claims so that the probability of it occurring is not too great. Thus, this extreme distribution will greatly help reinsurance companies in determining the amount of claims it bears because the occurrence of claims is included in extreme events with a small probability of occurrence.

Outbreak catastrophe that occurs in a country not only cause health problems, but also cause great economic losses. These losses can occur in various sectors, such as tourism, households, corporations, finance, and others (Qiu et al., 2018). To anticipate the risk, the government can insure it to reinsurance companies. Then, to expand their coverage capacity, they transfer the risk to the capital market (Cox and Pedersen, 2000). One of the securities that can be used as a means to accepting this risk is a bond. These securities were called outbreak catastrophe bonds (Liu et al., 2014). The most important part in the pricing of these bonds is the determination of the risk of losses. Therefore, knowledge of this matter is urgently needed, one of which can be obtained by extreme distribution (Jockovic, 2012). Zimbidis et al. (2007) determined the extreme distribution of catastrophe losses with the Extreme Value Theory

(EVT) Approach, whose interest rates are stochastically determined through the Cox-Ingersoll-Ross Model (CIR). Chao and Zou (2018) determined the extreme distribution of catastrophe losses with two trigger events each obtained by the EVT Approach which was then combined with the Copula function. Then, Residori (2019) determined the extreme distribution of catastrophe losses with the EVT Approach through the Block Maxima Method.

Based on the explanation above, this study will focus on determining the extreme distribution of losses due to outbreak catastrophe. The approach and method used to determine this distribution are Extreme Value Theory (EVT) and Peaks Over Threshold (POT), respectively. The quantile model of the distribution will also be estimated by the High Quantile Estimation Model. The EVT approach was chosen because of the extreme nature of outbreak catastrophe events. Then, the POT method was chosen because the classification of extreme data is determined based on data that exceeds a certain threshold value, regardless of the time of catastrophe, so it is very effective to use considering the small number of data.

2. Methods and Materials

2.1. Methods

The object of research used is data on economic losses due to outbreak catastrophe in various countries from 1976 to 2020. Data is collected from various sources such as the World Bank (<http://pubdocs.worldbank.org>), Asian Development Bank (<https://www.adb.org>), and so on. The tail of the data distribution must be fat. It is examined via its kurtosis. The tail of the data distribution is said to be fat if the kurtosis is greater than three, and vice versa. Then, the threshold value parameter will be estimated by Kurtosis Method, while the parameters of Generalized Pareto Distribution (GPD) which are assumed to match the data that has been resampled will be estimated by Maximum Likelihood Estimation Method based on Newton-Raphson Iteration. After that, the assumption of match of distribution will be proven by Kolmogorov-Smirnov Test. Finally, extreme distribution models and their quantile models will be built with Peaks Over Threshold and High Quantile Estimation, respectively.

2.2. Materials

2.2.1. Kurtosis Method

Threshold value is the boundary value between extreme data and non-extreme data. If $x_i, i = 1, 2, \dots, n$ are defined as the data sorted from smallest to largest, then iteration of kurtosis calculations can be done with the following equation:

$$K_k = \frac{\frac{1}{k} \sum_{i=1}^k (x_i - \bar{x}_k)^4}{s_k^4}; \quad k = n, n-1, \dots, 1, \quad (1)$$

where n represents the number of observed data, k represents the number of members of the sub data, K_k represents the kurtosis of the k -subdata, \bar{x}_k represents the average of the k -subdata, and s_k^2 represents the standard deviation of the k -subdata. If k -subdata has been obtained with $K_k \leq 3$ for the first time, then the iteration is stopped and the largest data from the k -subdata is selected as the threshold value. After that, a data resample is done by reducing each extreme data by the threshold value.

2.2.2. Generalized Pareto Distribution (GPD)

According to Jockovic (2012), if $Y \sim GPD(\sigma, \xi)$ where σ and ξ each represent the scale and shape parameters of the GPD, then the Y distribution can be expressed as follows:

$$F_{\sigma, \xi}(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-\frac{1}{\xi}}; \quad \xi \neq 0, \quad (2)$$

while the density function is as follows:

$$f_{\sigma, \xi}(y) = \frac{1}{\sigma} \left(1 + \xi \frac{y}{\sigma}\right)^{-\left(\frac{1}{\xi} + 1\right)}; \quad \xi \neq 0, \quad (3)$$

where $\sigma > 0, y \geq 0$ for $\xi > 0$, and $0 \leq y \leq -\frac{\sigma}{\xi}$ for $\xi < 0$.

2.2.3. Maximum Likelihood Estimation

Maximum Likelihood Estimation is a method of estimating the parameters of a probability distribution that maximizes the likelihood function. If Y_{i^*} , $i^* = 1, \dots, m$ are defined as random variables which are independent and identically distributed with the probability function $f_Y(y_{i^*}|\theta)$ where θ is an unknown parameter, then the likelihood function of Y_{i^*} can be expressed as follows:

$$L(\theta|\mathbf{y}) = \prod_{i^*=1}^m f_Y(y_{i^*}|\theta), \quad (4)$$

where $\mathbf{y} = [y_1, \dots, y_m]$. The log-likelihood function is the natural logarithmic form of equation (4). According to Purba et al. (2017), this function is monotonically related to equation (4) so that maximizing this function is equivalent to maximizing equation (4). The following is an expression of the log-likelihood function:

$$\ln[L(\theta|\mathbf{y})] = \ln \left[\prod_{i^*=1}^m f_Y(y_{i^*}|\theta) \right]. \quad (5)$$

If the first derivative of equation (5) against θ exists, then the value θ which maximizes equation (5) is the solution of the following equation:

$$\frac{d}{d\theta} \ln \left[\prod_{i^*=1}^m f_Y(y_{i^*}|\theta) \right] = 0. \quad (6)$$

2.2.4. Newton-Raphson Iteration

According to Bakari et al. (2016), Newton-Raphson iteration is a method of estimating the maximum solution of a function numerically. Let $f(\theta|\mathbf{y})$ be a function with unknown parameter θ and the $(t-1)$ -th and t -th estimates of θ are denoted as $\hat{\theta}_{t-1}$ and $\hat{\theta}_t$, $t = 1, 2, \dots$. If the first and second derivatives of $f(\theta|\mathbf{y})$ against θ in $\hat{\theta}_{t-1}$ are known, then the general equation is used for finding the optimum solution of $f(\theta|\mathbf{y})$ is as follows:

$$\hat{\theta}_t = \hat{\theta}_{t-1} - \frac{f'(\hat{\theta}_{t-1}|\mathbf{y})}{f''(\hat{\theta}_{t-1}|\mathbf{y})}; \quad f''(\hat{\theta}_{t-1}|\mathbf{y}) \neq 0, \quad (7)$$

where $f'(\hat{\theta}_t|\mathbf{y})$ represents the first derivative of $f(\theta|\mathbf{y})$ against θ in $\hat{\theta}_t$, and $f'(\hat{\theta}_{t-1}|\mathbf{y})$ and $f''(\hat{\theta}_{t-1}|\mathbf{y})$ represent the first and second derivatives of $f(\theta|\mathbf{y})$ against θ in $\hat{\theta}_{t-1}$, respectively. The error value for each iteration, denoted by ε_t , is determined by the following equation:

$$\varepsilon_t = |\hat{\theta}_t - \hat{\theta}_{t-1}|; \quad t = 1, 2, \dots \quad (8)$$

Let ε be error that is tolerated. If the iteration t -th with $\varepsilon_t < \varepsilon$ has been obtained for the first time, the iteration is stopped. The value of $\theta = \hat{\theta}_t$ is called the maximum solution if $f''(\hat{\theta}_t|\mathbf{y}) < 0$.

2.2.5. Kolmogorov-Smirnov Test

Kolmogorov-Smirnov test is a formal test used to check the fit between certain empirical and theoretical distributions. According to Vribik (2020), the test statistic in this test is the largest absolute difference between the two distributions whose value can be determined by the following equation:

$$D = \max\{D_{i^*}\}; \quad i^* = 1, 2, \dots, m, \quad (9)$$

where D_{i^*} represent the maximum value between $a_{i^*} = |F_e(y_{i^*}) - F_t(y_{i^*})|$ and $b_{i^*} = |F_e(y_{i^*}) - F_t(y_{i^*+1})|$. To the significance level of α , D will be compared with the critical value D_m^α which is obtained from the Kolmogorov-Smirnov table according to the number of data and the value of α selected. There is a two-tailed hypothesis

formulation of this test, namely $H_0 : F_e(y) = F_t(y)$ and $H_1 : F_e(y) \neq F_t(y)$. Reject H_0 if $D \geq D_m^\alpha$, meaning that at the significance level α , the empirical distribution does not fit the specified theoretical distribution, and vice versa.

2.2.6. Extreme Value Theory (EVT)

According to Gilli and Kellezi (2006), Extreme Value Theory (EVT) is an approach used to determine extreme distributions so that it focuses on determining probability in the tail area. According to Jindrova and Pacakova (2016), there are two methods in the EVT approach, namely Block Maxima (BM) and Peaks Over Threshold (POT). The BM method identifies extreme data based on the highest value of each period (weekly, monthly, and so on), while the POT method identifies extreme data based on values that exceed a certain threshold value regardless of the time of the event.

2.2.7. Peaks Over Threshold Method

Peaks Over Threshold is a modern method of estimating the extreme distribution where the selection of the extreme values are based on values that exceed a certain threshold value regardless of the time of the event. Suppose X as random variable which represents economic losses due to outbreak catastrophe with distribution F_X and u represents a large threshold value. If the random variable $Y = X - u$ with $X > u$ is defined which represents the excess of extreme economic losses due to the outbreak catastrophe from the threshold value, then the Y distribution can be determined by the following equation:

$$F_Y(y) = P\{Y \leq y | X > u\} = P\{X \leq y + u | X > u\} = P\{X \leq x | X > u\} = \frac{F_X(x) - F_X(u)}{1 - F_X(u)}; x > u, \quad (10)$$

where $0 \leq y \leq x_f - u$ with $x_f = \sup\{x \in \mathbb{R}^+ + \{0\} | F_X(x) < 1\}$. If equation (10) is transformed, then the extreme distribution of X is as follows:

$$F_X(x) = [1 - F_X(u)]F_Y(y) + F_X(u); \quad x > u. \quad (11)$$

Based on the Pickands-Balkema-de Haan theorem, if the threshold value of u is large, then $F_Y(y)$ can be approximated by $GPD(\sigma, \xi)$ or $F_Y(y) \approx F_{\sigma, \xi}(y) = F_{\sigma, \xi}(x - u)$. Then, Galambos et al. (1993) formulated that the probability X is greater than the threshold value is $\frac{m}{n}$, where m and n represent the number of observed data and the number of extreme data, respectively. Based on this, equation (11) can be rewritten as follows:

$$F_X(x) = 1 - \frac{m}{n} \left(1 + \xi \frac{x - u}{\sigma} \right)^{-\frac{1}{\xi}}; \quad \xi \neq 0, x > u. \quad (12)$$

2.2.8. High Quantile Estimation Model

According to Deng et al. (2020), High Quantile Estimation (HQE) model is a model used to determine the p -quantile of the extreme distribution, where $p \in (F_X(u), 1)$. If $F_X(x_p) = p$ is defined, then the HQE model of equation (12) can be determined by the following equation:

$$x_p = u + \frac{\sigma}{\xi} \left\{ \left[\frac{m}{n(1-p)} \right]^\xi - 1 \right\}; \quad \xi \neq 0, p \in (F_X(u), 1). \quad (13)$$

3. Results and Discussion

The data must have a fat distribution tail. This can be seen from the data kurtosis. Calculation of data kurtosis is done by equation (1). The result of the calculation of kurtosis obtained is 10.8989. Since the kurtosis is greater than three, the tail of the data distribution is confirmed to be fat-tailed so that the process can continue.

3.1. Threshold Value Estimation

The iteration of the kurtosis calculation is done with equation (1), the process is assisted by the Scilab software. Snippets of the iteration results are presented in Table 1.

Table 1. The Iteration of the Kurtosis Calculation

Iteration	k -subdata	K_k
1	103-subdata	10.898918
2	102-subdata	8.152811
3	101-subdata	8.836517
⋮	⋮	⋮
41	63-subdata	3.329750
42	62-subdata	3.164317
43	61-subdata	2.949846

Based on Table 1, it appears that in the 43rd iteration, $K_{61} = 2.949846 \leq 3$ for the first time. Therefore, the iteration is stopped and the largest datum of the 61-subdata, 664, is chosen as the threshold value so that there are 42 extreme data ($m = 42$), namely the 62th datum to the 103th datum. After that, resample the data. Snippets of resample data results are presented in Table 2.

Table 2. Resample Data Results

i^*	Y_{i^*}
1	36
2	80
3	316
⋮	⋮
40	15092
41	15392
42	24336

Based on Table 2, it appears that the smallest data is USD 36 million, while the largest data is USD 24,336 million.

3.2. GPD Parameter Estimation

Suppose that $X_i, i = 1, 2, \dots, 103$ are random variables that are independent and have identical distributions that represent the economic loss due to the i th outbreak catastrophe. If $Y_{i^*} = X_i - u, X_i > u, i^* = 1, \dots, 42$, defined as independent random variables and have a $GPD(\sigma, \xi)$ are the excess of extreme economic losses from the threshold value, then based on equations (4) and (5), the likelihood and the log-likelihood functions of Y_{i^*} can be expressed as follows:

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i^*=1}^{42} \frac{1}{\sigma} \left(1 + \xi \frac{y_{i^*}}{\sigma}\right)^{-\left(1+\frac{1}{\xi}\right)}, \tag{14}$$

and

$$\ln[L(\boldsymbol{\theta}|\mathbf{y})] = -42 \ln(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i^*=1}^{42} \ln \left[\left(1 + \xi \frac{y_{i^*}}{\sigma}\right) \right], \tag{15}$$

respectively, where $\boldsymbol{\theta} = [\sigma, \xi]^T$. Based on (6), the value of $\boldsymbol{\theta}$ that maximizes the equation (15) is the solution of the following equation:

$$g(\boldsymbol{\theta}) = [g_{11}, g_{21}]^T = \mathbf{0}, \tag{16}$$

where $g(\boldsymbol{\theta})$ represents the first partial derivative of $\ln[L(\boldsymbol{\theta}|\mathbf{y})]$ against $\boldsymbol{\theta}$ whose elements are as follows:

$$g_{11} = \frac{\partial \ln[L(\boldsymbol{\theta}|\mathbf{y})]}{\partial \sigma} = -\frac{42}{\sigma} + \frac{1 + \xi}{\sigma} \sum_{i^*=1}^{42} \frac{y_{i^*}}{\sigma + \xi y_{i^*}}, \tag{17}$$

$$g_{21} = \frac{\partial \ln[L(\boldsymbol{\theta}|\mathbf{y})]}{\partial \sigma} = \frac{1}{\xi^2} \sum_{i^*=1}^{42} \ln\left(1 + \xi \frac{y_{i^*}}{\sigma}\right) - \frac{1 + \xi}{\xi} \sum_{i^*=1}^{42} \frac{y_{i^*}}{\sigma + \xi y_{i^*}}. \tag{18}$$

It appears that equations (17) and (18) have a closed form so that $\boldsymbol{\theta}$ will be estimated through the Newton-Raphson Iteration. The second partial derivative of $\ln[L(\boldsymbol{\theta}|\mathbf{y})]$ against $\boldsymbol{\theta}$, denoted $\mathbf{H}(\boldsymbol{\theta})$, is determined first. $\mathbf{H}(\boldsymbol{\theta})$ is stated as follows:

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}, \tag{19}$$

where

$$h_{11} = \frac{\partial^2 \ln[L(\boldsymbol{\theta}|\mathbf{y})]}{\partial \sigma^2} = \frac{42}{\sigma^2} - \frac{1 + \xi}{\sigma^2} \sum_{i^*=1}^{42} \frac{y_{i^*}}{\sigma + \xi y_{i^*}} - \frac{1 + \xi}{\sigma} \sum_{i^*=1}^{42} \frac{y_{i^*}}{(\sigma + \xi y_{i^*})^2}, \tag{20}$$

$$h_{12} = h_{21} = \frac{\partial^2 \ln[L(\boldsymbol{\theta}|\mathbf{y})]}{\partial \sigma \xi} = \frac{1}{\sigma} \sum_{i^*=1}^{42} \frac{y_{i^*}(\sigma - y_{i^*})}{(\sigma + \xi y_{i^*})^2}, \tag{21}$$

$$h_{22} = \frac{\partial^2 \ln[L(\boldsymbol{\theta}|\mathbf{y})]}{\partial \xi^2} = -\frac{2}{\xi^3} \sum_{i=1}^{42} \ln\left[\left(1 + \xi \frac{y_i}{\sigma}\right)\right] + \frac{2}{\xi^2} \sum_{i=1}^{42} \frac{y_i}{\sigma + \xi y_i} + \frac{1 + \xi}{\xi} \sum_{i=1}^{42} \left(\frac{y_i}{\sigma + \xi y_i}\right)^2. \tag{22}$$

Based on equation (7), the Newton-Raphson iteration is done with the following equation:

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} - \mathbf{H}^{-1}(\hat{\boldsymbol{\theta}}_{t-1})g(\hat{\boldsymbol{\theta}}_{t-1}); \det[\mathbf{H}(\hat{\boldsymbol{\theta}}_{t-1})] \neq 0, t = 1, 2, \dots \tag{23}$$

Next, the tolerable error value is chosen, namely $\varepsilon = 0.000001$. For the first iteration, the initial estimate vector is determined first. According to Jockovic (2012), the initial estimated vector of GPD is determined by the following equation:

$$\hat{\boldsymbol{\theta}}_0 = \left[\frac{1}{2} \bar{y} \left[\left(\frac{\bar{y}}{s}\right)^2 + 1 \right], \frac{1}{2} \left[\left(\frac{\bar{y}}{s}\right)^2 - 1 \right] \right]^T, \tag{24}$$

where \bar{y} and s represent the mean and standard deviation of \mathbf{y} , respectively. Iteration is done with the help of Scilab software. Snippets of iteration results are presented in Table 3.

Table 3. Newton-Raphson Iteration

Iteration -t	$\hat{\sigma}_t$	$\hat{\xi}_t$	ε_t
1	5024.7906987	-0.185757149	1009.272534791
2	5635.4567170	-0.195067124	610.666018432
3	5455.6182602	-0.154398012	179.838461380
⋮	⋮	⋮	⋮
9	3204.6723139	0.359740363	1.822813535934
10	3204.6666983	0.359744618	0.005615644666
11	3204.6666982	0.359744618	0.000000055364

Based on Table 3, it appears that the iteration stops at the 11th iteration with an error of $\varepsilon = 0.000000055364$ and the parameter estimators obtained are $\sigma \approx 3204.6666982$ and $\xi \approx 0.359744618$. It appears that $\forall \mathbf{b} = [b_1, b_2]^T \in \mathbb{R}^2, \mathbf{b}^T \mathbf{H}(\hat{\boldsymbol{\theta}}_t) \mathbf{b} = -[(0.000002(b_1 + 3,067b_2))^2 + 12.401592b_2^2] \leq 0$, so $\hat{\boldsymbol{\theta}}_{11}$ is the maximum solution of $\ln[L(\boldsymbol{\theta}|\mathbf{y})]$.

3.3 Fit Test for Empirical Distribution and GPD

The following is the Probability-Probability Plot (P-P Plot) and Quantile-Quantile Plot (Q-Q Plot) of the empirical and GPD distributions which can be seen in Figure 1.

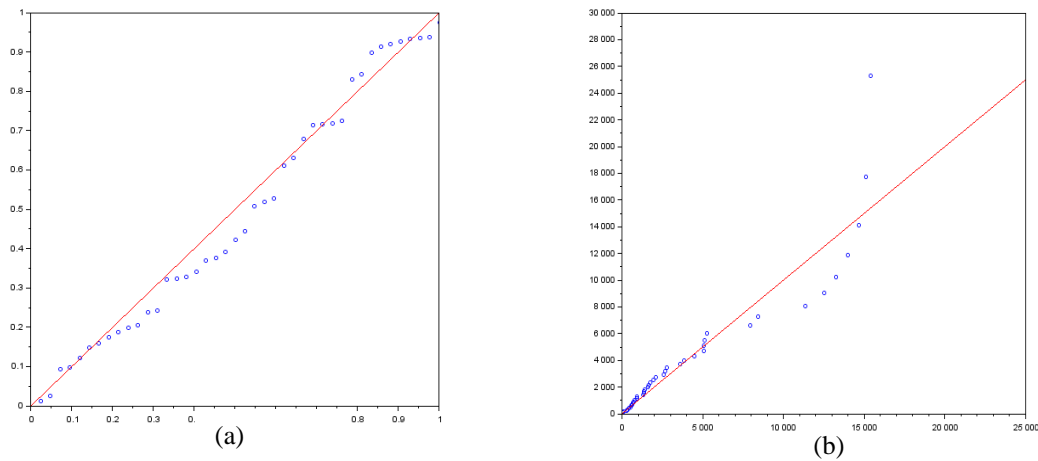


Figure 1. P-P Plot (a) and Q-Q Plot (b)

Based on Figure 1, it appears that the set of points on the P-P Plot and Q-Q Plot are scattered around the one-gradient line, so it can be seen that the empirical distribution fits the GPD. Formal testing is carried out by the Kolmogorov-Smirnov test. The level of significance chosen is $\alpha = 0.05$. Visualization of fitting empirical and GPD distributions using the Kolmogorov-Smirnov test was made with the help of Scilab software. The result of the visualization is presented in Figure 2.

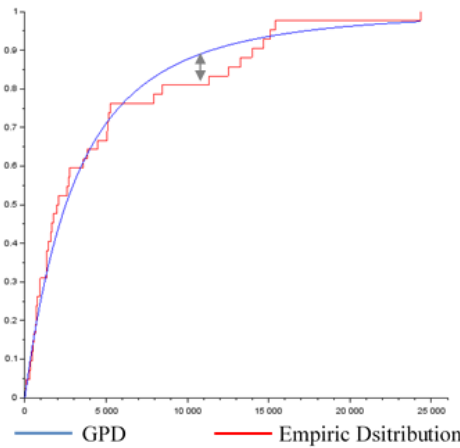


Figure 2. The Fitting of Empirical Distribution and GPD with the Kolmogorov-Smirnov Test

Based on Figure 2, it appears that the test statistics are at intervals (10,000, 15,000). Based on equation (9), the test statistic obtained is $D = 0.086490$. Based on the number of data and the selected significance level, the critical value obtained from the Kolmogorov-Smirnov table is $D_{42}^{0.05} = 0.208310$. It appears that $D < D_{42}^{0.05}$, so the decision taken is to accept H_0 , which means that the empirical distribution follows GPD.

3.4. Building Extreme Distribution from Economic Losses Due to Outbreak Catastrophe

Based on equation (12), the following is the extreme distribution of economic losses due to outbreak catastrophe:

$$F_X(x) = 1 - \frac{42}{103} \left(1 + 0.359744618 \frac{x - 664}{3204.6666982} \right)^{-\frac{1}{0.359744618}}, x > 664. \tag{25}$$

Based on equation (25), if the amount of the bond claim is USD 1,167.26 million, then the probability of the claim will be 0.35. Then, based on equation (13), here is the quantile model of equation (25):

$$x_p = 664 + \frac{3204.6666982}{0.359744618} \left\{ \left[\frac{42}{103(1-p)} \right]^{0.359744618} - 1 \right\}; p \in (0.59, 2). \tag{26}$$

Based on equation (26), if the probability of a claim is 0.35, then the of the bond claim will be USD 1,167.26 million.

3.5. Discussion

Reinsurance companies must determine the amount of the claim so that the amount is greater than USD 664 million so that the probability of it happening is not too big. The sponsor as the insured must know the relationship between the amount of the claim expected and the probability of its occurrence. Investors also have to think carefully about the risk of losing the principal and the amount of return they might get. For outbreak catastrophe bonds with a principal amount of 1, a coupon rate of 10%, an interest rate of 8%, and a term of one year, these relationships are visualized in a Cartesian Diagram. Visualization is done with Scilab software. The results of the visualization are presented in Figure 3.

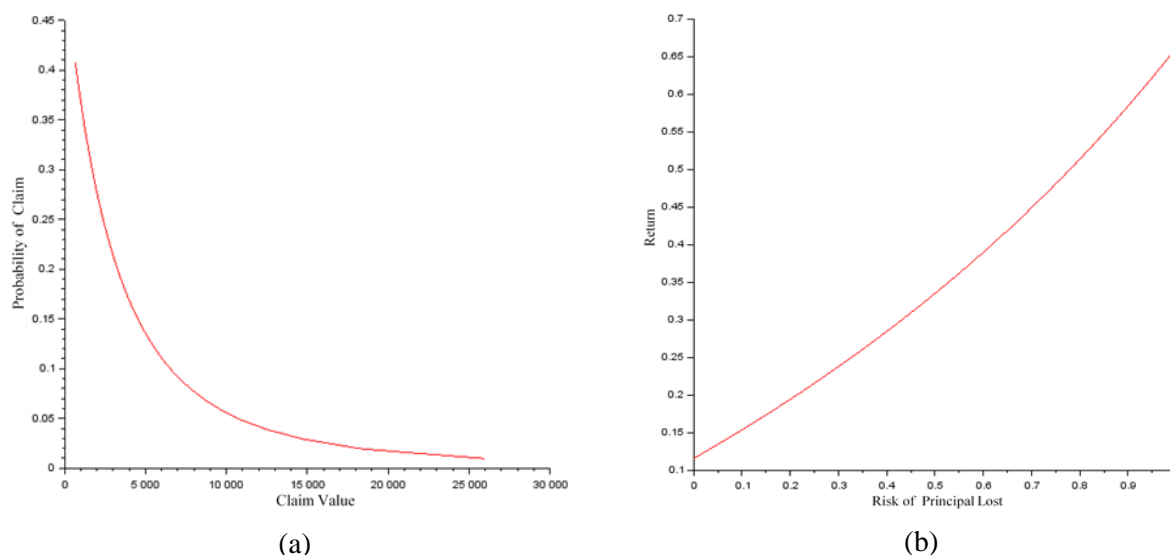


Figure 3. The relationship between the amount of the claim and the probability of the claim (a), and the relationship between the risk of loss of principal and the amount of return (b)

Based on Figure 3, it appears that the greater the claims expected by the sponsor, the smaller the probability of claims occurring. Therefore, if the sponsor does not want to lose the benefits of the bond, the determination of the amount of the claim must be calculated in such a way that the amount is not too large so that the probability of the claim is not too small. Then, it also appears that the higher the risk of losing principal, the higher the return that investors may get. The minimum return that investors may get is 11.55%, while the maximum return that investors may get is 66.15%. Therefore, if investors dare to take risks, then buying outbreak catastrophe bonds assuming the principal is lost is more profitable.

4. Conclusion

Based on the extreme distribution of economic losses due outbreak catastrophe, the threshold value of its losses is USD 664 million. Therefore, reinsurance companies must determine the amount of the claim so that it is greater than this value so that the probability of its occurrence is not too large. For sponsors, the greater the expected claim value, the less likely it is that claims will occur. Therefore, if the sponsor does not want to lose the benefits of the bond, the amount of the claim must be calculated in such a way that the amount is not too large so that the probability of its occurrence is not too small. Meanwhile, for investors, the higher the risk of losing principal, the higher the probability of a return. Therefore, if investors dare to take risks, bonds with the assumption of losing principal are more profitable.

References

- Bakari, H. R., Adegoke, T. M., & Yahya, A. M. (2016). Application of Newton-Raphson method to non-linear models. *International Journal of Mathematics and Statistics Studies*, 4(4), 21-31.
- Chao, W., & Zou, H. (2018). Multiple-event catastrophe bond pricing based on CIR-copula-POT model. *Discrete Dynamics in Nature and Society*, 2018, 1-9.
- Cox, S. H. & Pedersen, H. W. (2000). Catastrophe Risk Bonds. *North American Actuarial*, 4(4), 56-82.

- Deng, G., Liu, S., Li, L., & Deng, C. (2020). Research on the Pricing of Global Drought Catastrophe Bonds. *Hindawi*, 2020 (3898191), 1-7.
- Galambos, J., Lechner, J., & Simiu, E. (1994). *Extreme Value Theory an Application*. Kluwer Academic Publisher: Netherland.
- Gilli, M. & Kellezi, E. (2006). An Application of Extreme Value Theory for Measuring Risk. *Computational Economics*, 27(1), 1-23.
- Liu, J., Xiao, J., Yan, L., & Wen, F. (2014). Valuing Catastrophe Bonds Involving Credit Risks. *Hindawi*, 2018(563086), 1-6.
- Jindrova, P. & Pacakova, V. (2016). Modelling of Extreme Losses in Natural Disasters. *International Journal of Mathematical Models and Method in Applied Sciences*, 10, 171-178.
- Jockovic, J. (2012). Quantile Estimation for The Generalized Pareto Distribution with Application to Finance. *Yugoslav Journal of Operations Research*, 22, 297-311.
- Purba, S. A., Sutarman, and Darnius, S. (2017). Maximum Likelihood Based On Newton Raphson, Fisher Scoring, and Expectation Maximization Algorithm Application On Accident Data. *International Journal of Advanced Researh*, 6(1), 965-969.
- Qiu, W., Chu, C., Mao, A., and Wu, J. (2018). The impact on Health, Society, and Economy of SARS and H7N9 Outbreaks in China: A Case Comparison Study. *Hindawi*, 2018(2710185), 1-7.
- Residori, A. (2019). *Pricing Catastrophe Bonds Using Extreme Value Theory*. Thesis. Zurich: Postgraduate Program University of Zurich.
- Vribik, J. (2020). Small-Sample Corrections to Kolmogorov-Smirnov Test Statistic. *Pioneer Journal of Theoretical and Applied Statistics*, 15(1), 15-23.
- Zimbidis, A. A., Frangos, N. E., & Pantelous, A. A. (2007). Modeling Earthquake Risk via Extreme Value Theory and Pricing the Respective Catastrophe Bonds. *ASTIN Bulletin*, 37(1), 163-184.