IDENTIFYING THE PROBABILITY DISTRIBUTION MODELS OF ECONOMIC LOSSES DUE TO NATURAL DISASTERS

Ashish Jha, Vikas Kumar Sharma, Abhimanyu Singh Yadav

•

Department of Statistics, Banaras Hindu University, Varanasi, India
vikasstats@rediffmail.com

Abstract

Natural catastrophes have a tremendous influence on the environment and our economy, which has raised significant concerns and spurred scientific research. Several studies have been done to model the economic losses brought on by natural disasters. In this article, we primarily concentrate on examining the distributions of economic losses resulting from big catastrophes including wildfires, earthquakes, droughts, volcanic eruptions, and harsh weather. We recommend utilizing five well-known statistical distributions, including the Weibull, Log-logistics, Gamma, Generalized Pareto, and Lognormal distributions since we observe the skewed forms of the empirical distributions. We employ the maximum likelihood technique for each distribution for the available data sets in order to estimate the distributions. The parameter estimations are numerically computed using the PSO method. We select the distribution that best fits the economic losses using the Akaike Information Criterion and Kolmogorov-Smirnov statistics. We discovered that the Log-logistic distribution is the distribution that fits the total economic losses caused by all-natural disasters the best.

Keywords: Natural catastrophes, Economic losses, Probability distribution models, Maximum likelihood estimation, PSO Method, R-software, Goodness-of-fit tests

I. Introduction

Nature has been giving us gifts since the beginning of time. But we have also had to deal with the terrible things about it. Every year, a large number of different natural disasters, including floods, wildfires, earthquakes, extreme heat, cold, and volcanic activity, claim the lives of on average 60,000 people. Direct and indirect effects are distinguished by a recognised typology of disaster effects [1]. The destruction of fixed assets, raw materials, natural resources, high-yielding crops, and the loss of priceless lives are examples of direct repercussions. Indirect effects, which are frequently referred to as economic losses, are those that have an impact on economic activity over time, particularly in the goods and services sectors [2].

According to EM-DAT, catastrophes caused 0.1% of fatalities in the previous two decades. High-impact incidents accounted for 0.1% to 0.4% of all fatalities. Flood and drought were the deadliest natural calamities, but they no longer kill many. Earthquakes are the deadliest nowadays. Along with life, calamities also destroy resources. These risks affect economic activity, causing volatility and losses for the global economy; see [1,2,3,4].

Natural disasters have increased dramatically over the last three decades, posing a significant threat to the world's economies, particularly those of developing countries. The impact of economic losses on developing countries is far greater than that on developed countries. Between 1970 and 2002, 6436 natural disasters occurred, with developing countries bearing the brunt of the damage. It

demonstrates that developing countries are unable to combat these deadly disasters due to a lack of resources [2,3]. The relationship between natural disasters and economic losses is widely evident all over the world, and a number of studies have prompted the need for disaster mitigation strategies to reduce human and economic suffering. For such studies, we refer the readers to follow [5,6].

The preceding discussion highlights the necessity of evaluating the distribution of economic losses caused by natural disasters. Estimating economic losses due to disasters was a huge difficulty in the early days, and it was dependent on a hypothetical or singular historical occurrence rather than mathematical or statistical modelling. Few studies have been done to calculate the economic damages incurred by various natural disasters. [7] estimated economic losses for the whole spectrum of extreme weather, such as draught and flood, by combining stochastic hydro-meteorological crop-loss models with a regionalized computable general equilibrium model. [8] estimated the economic losses caused by natural disasters using the input-output model and associated modelling frameworks such as the social accounting matrix and the computable general equilibrium. Furthermore, [9] introduced a novel modelling framework known as the regional input-output model to explore the effects of natural catastrophes.

Coronese et al. [10] estimated the damage and mortality caused by natural catastrophes using a quantile regression model. They discovered an increasing trend in extreme natural catastrophe damages, which is consistent with a climate-change signal. Natural catastrophe casualties have reduced, despite an increase in economic damages. They also noticed an alarming increase in casualties associated with severe temperatures. [4] proposed using extreme value theory for modelling economic losses in a monograph, and they employed extreme value and extended pareto distributions for fitting heavy tailed distributions of economic losses. [11] discovered that generalised extreme value models and generalised Pareto distributions match well to the extreme losses of natural disasters and are helpful tools for calculating the tails of loss severity distributions. [12] used a generalised Pareto distribution to describe economic damages resulting from non-natural disasters.

The majority of academicians generally support the use of generalised extreme value distributions to describe economic damages brought on by natural disasters. Only one or two natural disasters are modelled using probability distribution models. Numerous probability models that are available in the literature could match such datasets more accurately than the generalised Pareto distribution. In this article, we use different probabilistic models to fit the economic losses caused by six significant calamities: drought, earthquake, extreme weather, extreme temperature, wildfire, and volcanic activity. For each of these data sets, we look at five three-parameter statistical distributions (size, shape, and placement). The new research completely contradicts the studies under consideration and discovers a very suitable distribution for natural disasters. For the purpose of calculating numerical maximum likelihood estimates of the unidentified model parameters, we use the particle swarm optimization approach (PSO). We employ goodness-of-fit tests like Kolmogorov-Smirnov and Akaike Information Criterion to choose the probability distribution model that best fits the distribution of natural catastrophes.

This paper is organized into different sections. In the first section, we review the literature on the economic loos due to natural disasters. Which is then followed by a discussion on data and variables viz. drought, earthquake, extreme temperature, extreme weather, volcanic activity, and wildfire. Which is then knowledge about the three parametric distribution (viz. Weibull, Log-logistics, Gamma, Gen. Pareto, and Log-normal) and estimation techniques. Analysis of results follows which is finally concluded by the discussion of the results with respect to the objectives of the study.

II. Data description

The Pro Vention Consortium of the World Bank Catastrophe Management Facility launched a coordinated effort to review the quality, accuracy, and completeness of three global disaster data sets after realising the need for higher quality data to enhance disaster preparedness and mitigation. These were EM-DAT managed by the Centre for Research on the Epidemiology of Disasters (CRED), Sigma

maintained by Swiss Reinsurance Company (Zürich), and Nat Cat maintained by Munich Reinsura nce Company (Munich).

Over 22,000 mega catastrophes have occurred throughout the world since 1900, and EM-DAT provides crucial core data on their incidence and consequences. The database is created using data from a variety of sources, including UN agencies, non-governmental organisations, insurance firms, research institutions, and press outlets. Our catastrophe information is taken from the EM-DAT International Disaster Database, which CRED and the US Office for Foreign Disaster Assistance both administer (OFDA). A disaster is one that meets at least one of the following criteria, according to the database: 10 or more fatalities, 2000 or more people impacted by hunger and drought, 100 or more by other calamities, a government disaster declaration, or an appeal for outside help.

We examine economic losses caused by draughts, earthquakes, volcanic activity, harsh weather, high temperature, and wildfires. Minimum value, maximum value, mean, variance, standard deviation, coefficient of variation, skewness, and kurtosis of economic losses due to natural disasters were computed. Table 1 shows an overview of the descriptive statistics for each economic variable.

Table 1: *The descriptive statistics of economic losses due to the natural disasters*

Variable	Mean	Variance	Coefficient Variation	Skewness	Kurtosis
Draught	340.33	264018.7	150.98	2.66	8.0
Earthquake	981.81	8821100.0	302.51	5.73	38.6
Extreme Tem.	196.46	185152.3	219.03	3.92	16.3
Extreme Weather	1363.40	7919309.0	206.40	3.60	16.4
wildfire	200.53	141181.8	187.37	4.23	21.9
Volcanic Activity	14.05	666.4	183.74	2.48	5.6

According to the overview, between 1900 and 2018, the minimum damages attributable to natural disasters ranged from $0.02 billion to $0.10 billion. Volcanic activity provided the lowest minimum losses, but drought and wildfire caused the highest minimum losses. Natural disasters can cause maximum losses ranging from $100 to $23030 billion. Volcanic activity has the lowest maximum losses, whereas earthquakes have the highest possible losses. The mean or average of all disaster-related losses from 1900 to 2018 ranges from $14.05 to $1363.40 billion, with extreme weather events having the highest mean value and volcanic activity having the lowest mean value. The range of the standard deviation for all six variables is $25.8 to $297 billion. Volcanic activity has the lowest standard deviation while earthquakes have the highest. The fact that some effects created by extreme value to the huge value in the raw data makes it evident that the standard deviation for each variable is always larger than the mean. The degree of asymmetries in a distribution around the mean is determined using the coefficient of skewness. All distributions are positively skewed as the skewness value is more than zero and lies in between 2.48 and 5.73. The skewness value for an earthquake is the highest at 5.73, clearly showing that it is very obviously skewed and that its asymmetric tail is extending to the right, while the skewness value for volcanic activity is the lowest at 2.48, showing that it is less obviously skewed and that its symmetric tail is also extending to the right.

The relative peak or flatness of a distribution can be assessed using the value of kurtosis. Kurtosis values range from 5.6 to 38.6, and they are always greater than 3. All distributions of economic losses have a greater peak than the typical normal distribution. Earthquake's height value is 38.6, which suggests the likelihood of a leptokurtic distribution, in which the data set tends to have a prominent peak close to the mean and a heavy tail. Volcanic activity obtains the lowest value and tends to have a flat peak close to the mean in the data set. The same can be deduced from Figure 1. It is possible for us to state here that one ought to employ the probability distribution models for the purpose of fitting such data sets which are positively skewed and have a frequency curve with high peaks.

III. Methodology

In this section, we will explain the statistical approaches that would be utilised to achieve the fitting of economic variables, which was covered in the previous section. Inclusion in this category includes that of probability distributions, parameter estimation, and the goodness-of-fit criterion.
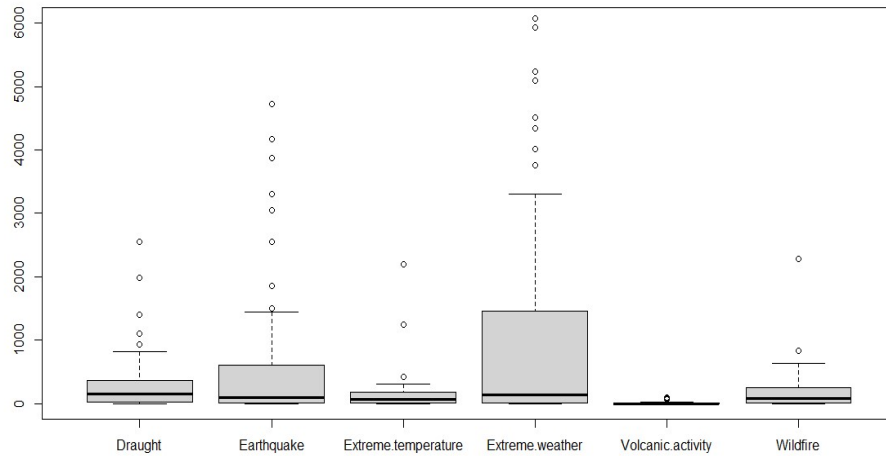

Figure 1: *Boxplots of economic losses due to six main natural disasters*

I. Probability distributions

The human mind is capable of incredible feats, and statistical modelling is one of them. It involves abstracting the results of observation in order to determine the similarities and differences between occurrences. When it comes to protecting ourselves against the effects of natural disasters, statistical models are a typical tool. The process of evaluating risks, making forecasts, and issuing warnings all depend heavily on modelling.

It is possible to draw the conclusion from the previous section that a statistical distribution with a right-skewed spread is the most accurate when it comes to modelling economic losses due to natural disasters. The current article makes use of five positively skewed distributions, namely the Weibull distribution, the Log-logistics distribution, the gamma distribution, the generalised Pareto distribution, and the Lognormal distribution, in order to fit the economic losses caused by natural disasters such as drought, earthquake, extreme temperature, extreme weather, volcanic activity, and wildfire. The probability density function (PDF), and cumulative distribution function (CDF) are given in Table 2.

The three-parameter Weibull distribution is commonly utilised in reliability and life data analysis [13]. Weibull distributions with β=1 have a constant failure rate, indicating usable life or random failures. Weibull distributions with β > 1 have a wear-out failure rate. Next is the Log-logistics three-parameter distribution, often known as the Fisk distribution in economics [14]. Characterizing the lifetime distributions log logistics distributions have property of a constant discrete Log-odds rate (LOR) with respect to t and ln t [15]. A random variable with a logistic logarithm has a log-logistic distribution. It resembles Lognormal but has heavier tails. Its cumulative distribution function is closed, unlike the lognormal. This distribution can exhibit a monotonically decreasing failure rate function for some parameter values. It is a survival analysis model for occurrences whose rate rises then falls. Some applications of the log-logistic distribution are discussed in economics to model wealth or income distribution [16] and in hydrology to estimate stream flow and precipitation [17].

Table 2: *The PDF & CDF of the distributions*

Model	PDF and its Support	CDF
Weibull	$f(x) = \frac{\beta}{\alpha}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-\gamma}{\alpha}\right)^{\beta}}, x \geq \gamma, \beta > 0, \alpha > 0.$	$F(x) = 1 - e^{\left(\frac{x-\gamma}{\alpha}\right)^{\beta}}$
Log-logistics	$f(x) = \frac{\left(1+\frac{\beta}{\alpha}(x-\gamma)\right)^{-\left(\frac{1}{\beta}+1\right)}}{\sigma\left(1+\left(1+\frac{\beta}{\alpha}(x-\gamma)\right)^{-\left(\frac{1}{\beta}\right)}\right)^2}, x \geq \gamma, \beta > 0, \alpha > 0.$	$F(x) = \left(1 + \left(1 + \frac{\beta}{\alpha}(x-\gamma)\right)^{-\left(\frac{1}{\beta}\right)}\right)^{-1}$
Gamma	$f(x) = \frac{1}{\Gamma\alpha\beta^{\alpha}}(x-\gamma)^{\alpha-1} e^{-\left(\frac{x-\gamma}{\beta}\right)}, \gamma < x < \infty, \beta > 0, \alpha > 0.$	$F(x) = \frac{\Gamma\left(\frac{x-\gamma}{\beta^{\alpha}}\right)}{\Gamma\alpha}$
Gen. Pareto	$f(x) = \frac{1}{\alpha}\left(1 + \beta\left(\frac{x-\gamma}{\alpha}\right)\right)^{-1-\left(\frac{1}{\beta}\right)}, x \geq \gamma, \beta > 0.$	$F(x) = 1 - \left(1 + \frac{\beta(x-\gamma)}{\sigma}\right)^{-\left(\frac{1}{\beta}\right)}$ for $\beta \neq 0$
Log-normal	$f(x) = \frac{1}{(x-\gamma)\alpha\sqrt{2\pi}} exp\left\{\frac{[ln(x-\gamma)-\beta]^2}{2\alpha^2}\right\}, x > \gamma \geq 0, -\infty < \beta < \infty, \alpha > 0.$	$F(x) = \Phi\left(\frac{ln(x-\gamma)-\beta}{\alpha}\right)$

It is positively skewed and the amount of skew depending inversely on the shape parameter. In gamma distribution median does not have a closed-form equation. Some applications of the gamma distribution are discussed in climatology to estimate the different behaviour of the natural climatic events [18] and in hydrological analysis [19]. Environmental studies use the Generalized Pareto distribution to model heavy-tailed data sets [4]. The distribution is called the "peaks over thresholds" model because it models flood control threshold exceedances. Generalized Pareto distribution models are used for extreme event [20]. The log-normal distribution is a function distributing a dependent variable in a normal or Gaussian fashion on a logarithmic scale of the independent variable (i.e., if the random variable X is log-normally distributed then Y=ln(X) has a normal distribution). A distribution that is log-normal in one of its moments will be log-normal in any of its moments with the same geometric standard deviation, describing the spread of the dependent variable [21]. The median size of any moment is connected to the median size of any other moment by an analytical relationship derived by [22]. One of the most common applications where log-normal distributions are used in finance is in the analysis of stock prices [23].

II. Maximum Likelihood Estimation

The most common method for obtaining estimators is by far the maximum likelihood approach. According to the MLE concept, the probability distribution that is "most likely" to accommodate for the observed data is the one that is wanted. As a result, one must look for the parameter vector value that maximises the likelihood function $L(\theta|x)$. The notion of maximum likelihood, which selects as the estimator that value of the parameter that maximises the PDF $f_{\theta}(x)$, effectively presupposes that the sample is representative of the population.

For each sample point x, let $\hat{\theta}(x)$ be a parameter value at which $L(\theta|x)$ attains its maximum as a function of θ, with x held fixed. Then $\hat{\theta}(x)$ is called the MLE of the parameter θ, (θ may be vector valued). Obtain n independent observations, $x_1, x_2, ..., x_n$ the estimates of parameters $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$ can be obtained by solving the differentiation of the logarithmic likelihood function as;

$$\frac{\partial log L(\hat{\theta}; x_1, x_2, ..., x_n)}{\partial \theta_j} = 0, j = 1, 2, ..., k. \tag{1}$$

Here, we discuss the complete producer of finding the MLEs of the Weibull distribution parameters. Consider the pdf and cdf of the Weibull distribution from Table 2. Assuming that the observations are independently distributed, the likelihood function is defined by,

$$L(\alpha,\beta,\gamma|data) = \prod_{i=1}^{n} f(x_i,\alpha,\beta,\gamma) \tag{2}$$

Our aim of estimation is to determine the three unknown parameter α,β,γ by the maximizing the likelihood (2) or equivalently log-likelihood function (3). The log-likelihood function is shown below.

$$logL(\hat\theta; x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n}\left[\ln(\beta) + (\beta-1)\ln(x-\gamma) - \beta\ln(\alpha) - \left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right] \tag{3}$$

Using the conventional approach, we take the partial derivatives of the log-likelihood function (3) in terms of α,β,γ and set them equals to zero. We obtain the following equations,

$$\begin{cases} \dfrac{\partial \ln(L)}{\partial\alpha} = \sum_{i=1}^{n}\left[\left(\dfrac{1}{\beta}\right) + \ln(x-\gamma) - \ln(\alpha) - \left(\dfrac{x-\gamma}{\alpha}\right)^{\beta}\ln\left(\dfrac{x-\gamma}{\alpha}\right)\right] = 0, \\[2ex] \dfrac{\partial \ln(L)}{\partial\beta} = \sum_{i=1}^{n}\left[-\left(\dfrac{\beta}{\alpha}\right) + \left(\dfrac{\beta}{\alpha}\right)\left(\dfrac{x-\gamma}{\alpha}\right)^{\beta}\right] = 0, \\[2ex] \dfrac{\partial \ln(L)}{\partial\gamma} = \sum_{i=1}^{n}\left[-\dfrac{(\beta-1)}{(x-\gamma)} + \left(\dfrac{\beta}{\alpha}\right)\left(\dfrac{x-\gamma}{\alpha}\right)^{\beta-1}\right] = 0. \end{cases}$$

It is commonly understood that obtaining estimates of unknown parameters by solving equations given above numerically is challenging. The particle swarm optimization (PSO) approach is used to find unknown parameters estimates and is inspired by the notion of heuristic algorithms. Using this process, we can find the MLE for all of the distributions under consideration.

III. Particle Swarm Optimization Method

The biologically inspired approach known as particle swarm optimization, which was initially described by [24], is based on the flocking behaviour of birds. PSO is a population-based, self-adaptive search optimization method also referred to as an optimizer. All of the particles in the swarm move faster toward the best individual and overall position while continuously evaluating the value of their present location according to the same controlling principle. Each particle has a memory that aids it in remembering its most recent optimal location. Particle positions are classified as either personal best (pbest) or global best (gbest). Each particle has a unique pbest that is based on the journey it has taken. The particle compares the fitness value of its present position to that of pbest at each step along its route. The pbest is changed to the present location if the latter has a greater fitness value. Each particle also had a method of knowing where the swarm's greatest concentration of flowers had been located. The gbest, is the name given to this site of the best fitness ever found. There is a single gbest to which every particle is drawn throughout the whole swarm.

In a n-dimensional search space, the position and velocity of individual (particle or solution) i are represented as the vectors $Xi = (xi1, xi2, \ldots, xin)$ denote a particle's position (coordinate) and $Vi = (vi1, vi2, \ldots, vin,)$ denote the particle's flight velocity over a solution space in the PSO algorithm. Each individual x in the swarm is scored using a scoring function that obtains a score (fitness value) representing how good it solves the problem. Let $pbest_i$ and gbest = $(x_1^{gbest}, \ldots, x^{gbest}_n)$ be the position of individual i and its neighbors' best position so far, respectively. Each particle records its own personal best position (pbest), and knows the best positions found by all particles in the swarm (gbest). Then, all particles that fly over the n-dimensional solution space are subject to updated rules for new positions, until the global optimal position is found. The modified velocity and position of each individual can be calculated using the current velocity and the distance from $pbest_i$ to gbest as follows:

$$V_i^{k+1} = \omega V_i^k + C_1 R_{ran\,1}\left(pbest_i^k - X_i^k\right) + C_2 R_{ran\,2}\left(gbest_i^k - X_i^k\right) \tag{4}$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \qquad (5)$$

where V_i^k velocity of individual i at iteration k, ω weigh parameter (inertia weight), c_1, c_2 acceleration coefficients, R_{rand_1} and R_{rand_2} random numbers uniformly distributed between 0 and 1, X_i^k position of individual i at iteration k, $pbest_i^k$ best position of individual i until iteration k, $gbest_i^k$ best position of the group until iteration k.

The fundamental structure and pseudo-code of PSO algorithm

---

for each particle

      generate an initial particle

  end

  do

  for each particle

      calculate fitness value

      if the fitness value is better than the best fitness value (pBest) in history

        set current value as the new pBest

      end

  end

  choose the particle with the best fitness values of all the particles as the gBest

  or each particle

      calculate particles velocity according eq (5)

      update particle position according eq (6)

  end

while maximum iteration criterion is not attained.

Marinho et al. [25] introduce the Adequacy Model computational library version 2.0.0 for the R statistical environment with two major contributions: a general optimization technique based on the PSO method (with a minor modification of the original algorithm) and a set of statistical measures for assessment of the adequacy of the fitted model. The goodness.fit() function provides some useful statistics to assess the quality of fit of probabilistic models. The function can also compute other measures such as AIC and KS test statistic. The general form for the function is given below:

Goodness.fit (pdf, cdf, starts=NULL, data, method="PSO", lim_inf, lim_sup, min(x), e, s, N, domain=c (0, inf)) where,

- pdf: probability density function (pdf);
- cdf: cumulative distribution function;
- starts: initial parameters to maximize the likelihood function;
- data: data vector;
- method: method used for minimization of the -log-likelihood function.
- method = "PSO", then all arguments of the PSO() function could be passed to the goodness.fit() function.
- lim_inf and lim_sup: define the inferior and superior boundaries of the search space, respectively;
- e: current error. The algorithm stops if the variance in the last iterations is less than or equal to e;
- S: number of considered particles
- domain: domain of the pdf. By default the domain of the pdf is the open interval (0, 1).

IV. Model selection criterion

The choice of the best probability distribution is a crucial step. The best distribution model for economic variables is determined using the goodness-of-fit (GoF) test and Akaike information criterion (AIC). The model with the lowest AIC value is the best fitting model. We discover the more accurate estimate for selecting the optimal model using the PSO approach. In addition, we carry out the same task as a probability plot using an empirical CDF plot. The empirical cumulative probability that is closest to the S-curve empirical one is chosen as the best fitting. The GoF test determines if a statistical model fits a collection of observations supplied in advance. Accordingly, the GoF measures are primarily used to summarise the discrepancy between observed values and predicted values under the specified statistical model. The minimal error produced, as assessed by the methods below, will be used to find the distribution that is best fitted:

The AIC, developed by [26], ranks models according to how well they fit the data and how little error they generate in their estimates. To move away from a solely inferential and limited approach to model selection, AIC has become part of a growing movement. It is defined as follows.

$$AIC = -2logL(\hat{\theta}_k) + 2k \qquad (7)$$

Among all investigated distributions, the model with the lowest AIC value is regarded to be the best fitting model. Kolmogorov- Smirnov test compares empirical and theoretical distributions. Let us consider $F_0(x)$ is the population CDF and $S_N(x)$ the observed cumulative step function of a sample (i.e., $S_N(x) = k/N$, where k is the number of observation less or equal to x), then KS test statistic is defined as

$$T = \max_x |F_0(x) - S_N(x)|. \qquad (8)$$

For implications, we reject the hypothesis at the level of significance, α, if T exceeds the $1 - \alpha$ quantile as given by the table of quantile for the KS test statistic.

IV. Results and Discussion

The economic losses caused by six natural catastrophes (drought, earthquake, extreme weather, extreme temperature, wildfire, and volcanic activity) are examined in this part and fitted to the five probability distributions discussed in section 2. First, the investigation focuses on determining the best-fitting model using the AIC value. Among all the models evaluated, the model with the lowest AIC was deemed the best. However, the PSO technique in R was used to estimate the parameters of the five theoretical probability distributions using maximum likelihood estimation. Tables [3–8] provide the MLEs, KS statistic (along with p-value), and AIC value for each fitted model for each economic variable. The fitting results show that some PDF characteristics are more suited for some places while being less appropriate for others.

Table 3: *MLEs, KS statistics, p-value and AIC for all five distributions for Drought data*

Distribution	MLE			P - Value	Statistic	AIC
Log-Logistic	0.7277	94.8424	0.1000	0.5207	0.1131	653.0388
Weibull(3P)	0.3469	87.8642	0.0978	0.0005	0.2852	659.8648
Gamma	0.3742	0.0005	0.0629	0.0173	0.2161	654.3794
Gen.Pareto	1.0106	99.7400	0.0660	0.2976	0.1360	662.5274
Lognormal	2.6033	4.5910	0.0068	0.4443	0.1202	662.7647

Figure 3 depicts the PDF plot of all heavy-tailed variables. Our findings are closed in terms of log-logistic and Weibull distributions. The empirical investigations show that Weibull considerably fits the

greatest value whereas Generalised Pareto underestimates it. Meanwhile, of all competing models in the research, Generalised Pareto provides the weakest match. The log-logistic model is the second best. The KS test statistic are used to select the distribution at 95% confidences interval from the Tables [3-8] we compare the results for all the distribution.

To fit the distributions for economic losses owing to drought, Table 3 shows that the p-value of Weibull and Gamma is less than 0.05 so we reject the null hypothesis. Log-logistic distribution has the lowest AIC value (653.0338), highest likelihood estimates (0.7277, 94.8424, 0.1), and smallest p-value (0.5207). The gamma distribution has the second-lowest AIC value (654.3794) among all distributions. Table 4 shows that the p-value of Weibull, Gamma and General Pareto is less than 0.05 that implies we reject the hypothesis. Log-logistic has the lowest AIC for earthquake economic losses (1117.5730). Weibull has a higher AIC value than Lognormal (1122.6180).

Table 4: *MLEs, KS statistics, p-value and AIC for all five distributions for earthquake data*

Distribution	MLE			P - Value	Statistic	AIC
Log-Logistic	0.5823	68.3169	0.0650	0.3430	0.1029	1117.5730
Weibull(3P)	0.4213	89.6176	0.0616	0.0046	0.1912	1145.4830
Gamma	0.4980	0.0015	0.0603	0.0225	0.1644	1240.6280
Gen. Pareto	2.4820	25.8344	0.0565	0.0332	0.1571	1136.3800
Lognormal	2.6557	4.0074	0.0439	0.1001	0.1343	1122.6180

Table 5 shows that the p-value of Gamma is less than 0.05 so we reject the hypothesis. Log-logistics has the lowest AIC value (384.8178) across all distributions. Generalized Pareto has the second lowest AIC, whereas gamma has the highest. According to Table 6, the p-value of Weibull and Gamma is less than 0.05 that implies we reject the null hypothesis. The gamma distribution has the highest AIC value among all distributions for the distribution of economic losses brought on by extreme weather, while log-logistic has the lowest value.

Table 5: *MLEs, KS statistics, p-value and AIC for all five distributions for Temperature data*

Distribution	MLE			P - Value	Statistic	AIC
Log-Logistic	0.7077	72.8837	0.0600	0.2893	0.1737	384.8178
Weibull(3P)	0.5810	89.9062	0.0303	0.4292	0.1546	388.0758
Gamma	0.6875	0.0025	0.0083	0.0068	0.2979	397.0683
Gen.Pareto	0.8057	62.7886	0.0392	0.9816	0.0824	386.1265
Lognormal	1.7809	3.8543	0.0110	0.6215	0.1332	390.8100

Table 6: *MLEs, KS statistics, p-value and AIC for all five distributions for Extreme Weather data*

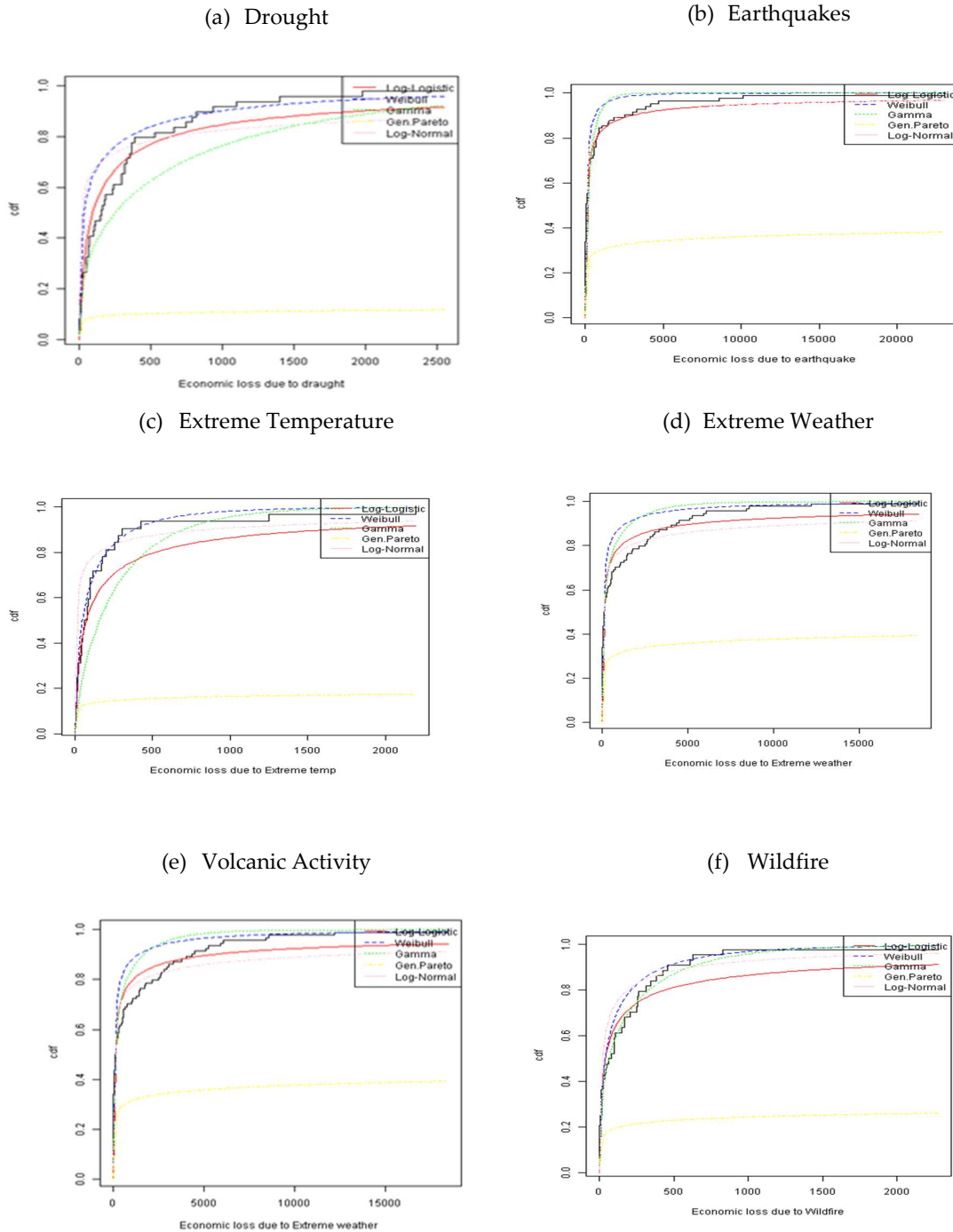| Distribution	MLE			P - Value	Statistic	AIC
Log-Logistic	0.5061	70.7698	0.0500	0.2772	0.1025	1343.7310
Weibull	0.2981	83.7310	0.0389	0.0003	0.2180	1372.3730
Gamma	0.2195	0.0004	0.0172	0.0068	0.1738	1384.3180
Gen. Pareto	3.0613	23.8658	0.0496	0.0984	0.1266	1366.3450
Lognormal	3.4142	4.7270	0.0423	0.5663	0.0811	1349.6740

(a) Drought

(b) Earthquakes

(c) Extreme Temperature

(d) Extreme Weather

(e) Volcanic Activity

(f) Wildfire

Figure 2: *Fitted CDF plots for economic losses.*

(a) Drought

(b) Earthquakes

(c) Extreme Temperature

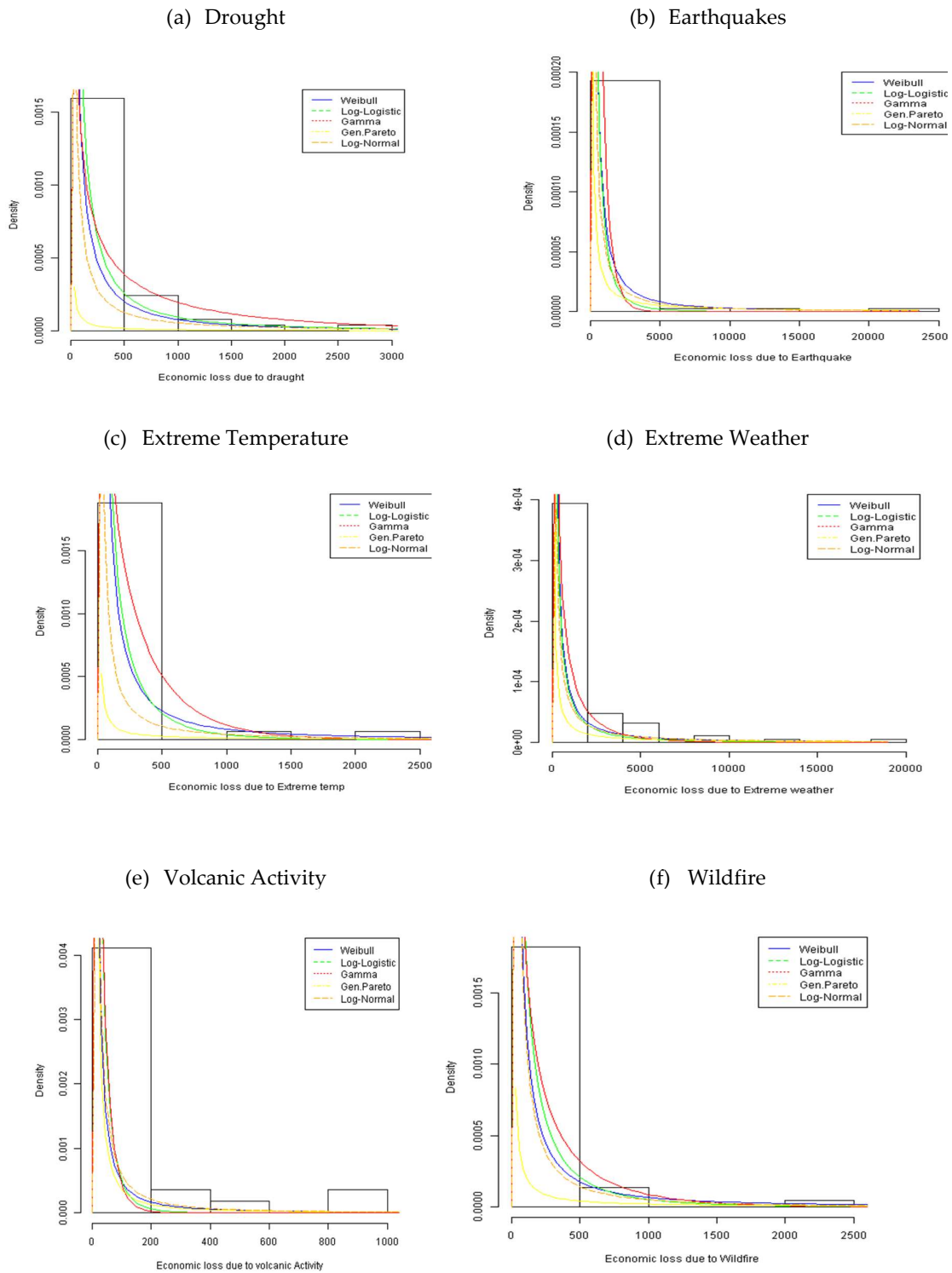(d) Extreme Weather

(e) Volcanic Activity

(f) Wildfire

Figure 3: *Fitted PDF plots for economic losses*

According to Table 7, the economic losses as a result of volcanic activity we discovered out of all the distributions, log-logistic has the least AIC value, followed by Weibull. Among all, Generalized Pareto has the highest AIC values.

Table 7: *MLEs, KS statistics, p-value and AIC for all five distributions for Volcanic data*

Distribution	MLE			P - Value	KS	AIC
Log-Logistic	0.6367	2.2675	0.0200	0.9628	0.0948	168.3601
Weibull(3P)	0.5663	9.6857	0.0199	0.0817	0.2390	171.2018
Gamma	0.3550	0.0253	0.0165	0.2350	0.1955	174.1816
Gen.Pareto	1.0966	1.9923	0.0200	0.6410	0.14019,	177.8270
Lognormal	0.8888	2.8176	0.0199	0.9108	0.1061	174.4506

Next, from Table 8 we discovered that among all distributions, Weibull has a lower AIC value than log-logistic, which has the second lowest value for the distribution of economic losses caused by wildfire. The AIC value of Generalized Pareto is the highest of all.

Table 8: *MLEs, KS statistics, p-value and AIC for all five distributions for Wildfire data*

Distribution	MLE			P - Value	KS	AIC
Log-Logistic	0.5814	40.4627	0.1000	0.5036	0.1244	510.7080
Weibull(3P)	0.5027	82.5439	0.0998	0.3161	0.1446	504.5558
Gamma	0.3547	0.0017	0.0477	0.9977	0.0594	516.5590
Gen. Pareto	1.4600	36.3105	0.0794	0.3054	0.1460	534.6831
Lognormal	2.9891	2.6805	0.0896	0.0507	0.2044	525.7421

V. Conclusion and future work

The goal of the current work is to identify the most appropriate three parametric probability models for datasets of economic losses from natural catastrophes. For modelling economic losses, scholars have previously advocated using the Generalized Pareto or extreme value distribution. Both probability distributions are specified on the real line, and economic losses occur on the positive real line. As a result, these distributions can offer a negative lower bound on the economic losses. In this work, we take into account five significant probability distributions (Weibull, Log-logistics, Gamma, Generalized Pareto, and Lognormal) that are defined on the positive real line to describe the economic scenarios. Utilizing the KS-test, CDF plot, and AIC criterion, the best fitted probability distribution is determined for each dataset. Empirical CDF plots show that Weibull and log-logistic fit pretty well, whereas generalised Pareto fits poorly. According to the KS-test statistic, we discovered that the Log-logistic and Lognormal suit all economic losses resulting from natural catastrophe data for the stated level of significance, 5%. However, draught, earthquake, and extreme weather datasets cannot be fitted by Weibull or Gamma. The earthquake dataset does not match the generalised Pareto model. The Log-logistic distribution offers the greatest fit among all taken distributions for five datasets (drought, earthquake, extreme weather, extreme temperature, wildfire, and volcanic activity) according to the AIC criteria. It is, nonetheless, the second-best fitted distribution for the wildfire dataset. It should be noticed that the Weibull distribution is rejected by the KS-test yet has the minimum AIC for the wildfire dataset. As a result, we may also suggest log-logistic for modelling economic losses from wildfires. Finally, we advise using the log-logistic probability model to fit and analyse economic losses brought on by natural catastrophes in future research.

Regression analysis is employed when the assumption of normality is taken into consideration to develop and investigate the relationship between the response and explanatory variables. In certain applications, the assumption of normalcy is not valid practically; see [27]. Numerous examples are

given in [28] that demonstrate the usage of skewed or non-normal distributions for both random components and response variables. In such circumstances, we advise fitting parametric regression for the economic losses using the log-logistic probability model. The model may be defined as

$$Economic\ loss\ (Y) = \beta X' + \sigma\ \varepsilon$$

where X' are the features matrix (regressions of economic losses), β is a vector of regression coefficients, σ is a scale parameter and ε stands for random component that may follow the log-logistics distribution. Readers are encouraged to take this work into consideration while planning their own future projects on modelling of economic losses due to natural disasters.

Table 9: *Ranks of the fitted distributions based on AIC values.*

Distributions	Draught	Earthquake	Extreme Temperature	Extreme Weather	Volcanic Activity	Wildfire
Log-Logistic	I	I	I	I	I	II
Weibull (3P)	III	V	III	IV	II	I
Gamma	II	IV	V	V	III	III
Gen. Pareto	IV	III	II	III	V	V
Lognormal	V	II	IV	II	IV	IV

Acknowledgement

References

[1] Sedghi, A. (2013). Typhoon Haiyan: how does it compare with other tropical cyclones. *The Guardian*.

[2] Yamano, N., Kajitani, Y., & Shumuta, Y. (2007). Modeling the regional economic loss of natural disasters: the search for economic hotspots. *Economic Systems Research*, *19*(2), 163-181.

[3] Kellenberg, D. K., & Mobarak, A. M. (2008). Does rising income increase or decrease damage risk from natural disasters? *Journal of urban economics*, *63*(3), 788-802.

[4] Pisarenko, V. F., & Rodkin, M. V. (2014). *Statistical analysis of natural disasters and related losses* (p. 82). Dordrecht-Heidelberg-London-New York: Springer.

[5] Panwar, V., & Sen, S. (2019). Economic impact of natural disasters: An empirical re-examination. *Margin: The Journal of Applied Economic Research*, *13*(1), 109-139.

[6] Benson, C., & Clay, E. (2003). Economic and financial impacts of natural disasters: an assessment of their effects and options for mitigation: synthesis report. *Overseas Development Institute, London*.

[7] Pauw, K., Thurlow, J., Bachu, M., & Van Seventer, D. E. (2011). The economic costs of extreme weather events: a hydrometeorological CGE analysis for Malawi. *Environment and Development Economics*, *16*(2), 177-198.

[8] Okuyama, Y. (2007). Economic modeling for disaster impact analysis: past, present, and future. *Economic Systems Research*, *19*(2), 115-124.

[9] Hallegatte, S. (2008). An adaptive regional input-output model and its application to the assessment of the economic cost of Katrina. *Risk Analysis: An International Journal*, *28*(3), 779-799.

[10] Coronese, M., Lamperti, F., Keller, K., Chiaromonte, F., & Roventini, A. (2019). Evidence for sharp increase in the economic damages of extreme natural disasters. *Proceedings of the National Academy of Sciences*, *116*(43), 21450-21455.

[11] Jindrová, P., & Pacáková, V. (2016). Modelling of extreme losses in natural disasters. *International Journal of Mathematical Models and Methods in Applied Sciences, volume 10, issue: 2016*.

[12] Ibrahim, R. A., Sukono, S., & Riaman, R. (2021). Estimation of the Extreme Distribution Model of Economic Losses Due to Outbreaks Using the POT Method with Newton Raphson Iteration. *International Journal of Quantitative Research and Modeling*, 2(1), 37-45.

[13] Moeini, A., Jenab, K., Mohammadi, M., & Foumani, M. (2013). Fitting the three-parameter Weibull distribution with Cross Entropy. *Applied Mathematical Modelling*, *37*(9), 6354-6363.

[14] Ramos, M. W. A., Cordeiro, G. M., Marinho, P. R. D., Dias, C. R. B., & Hamedani, G. G. (2013). The Zografos-Balakrishnan log-logistic distribution: Properties and applications. *Journal of Statistical Theory and Applications*, *12*(3), 225-244.

[15] Khorashadizadeh, M., Rezaei Roknabadi, A. H., & Mohtashami Borzadaran, G. R. (2013). Characterization of life distributions using Log-odds rate in discrete aging. *Communications in Statistics-Theory and Methods*, *42*(1), 76-87.

[16] Kleiber, C., & Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. John Wiley & Sons.

[17] Ashkar, F., & Mahdi, S. (2006). Fitting the log-logistic distribution by generalized moments. *Journal of Hydrology*, *328*(3-4), 694-703.

[18] Thom, H. C. (1958). A note on the gamma distribution. *Monthly weather review*, *86*(4), 117-122.

[19] Aksoy, H. (2000). Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences*, *24*(6), 419-428.

[20] Holmes, J. D., & Moriarty, W. W. (1999). Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, *83*(1-3), 1-10.

[21] Heintzenberg, J. (1994). Properties of the log-normal particle size distribution. *Aerosol Science and Technology*, *21*(1), 46-48.

[22] Hatch, T., & Choate, S. P. (1929). Statistical description of the size properties of non uniform particulate substances. *Journal of the Franklin Institute*, *207*(3), 369-387.

[23] Dufresne, D. (2004). The log-normal approximation in financial and other computations. *Advances in applied probability*, *36*(3), 747-773.

[24] Kennedy, J., & Eberhart, R. (1995). "Particle swarm optimization," Proceedings of ICNN'95–International Conference on Neural Networks, Perth, WA, Australia.

[25] Marinho, P. R. D., Silva, R. B., Bourguignon, M., Cordeiro, G. M., & Nadarajah, S. (2019). AdequacyModel: An R package for probability distributions and general purpose optimization. *PloS one*, *14*(8), e0221487.

[26] Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*, 192-196.

[27] Dunteman, G. H., & Ho, M. H. R. (2006). An introduction to generalized linear models (Vol. 145). Sage.

[28] Kalbfleisch, J. D., & Prentice, R. L. (2011). The statistical analysis of failure time data. John Wiley & Sons.