

基于可变权重的威布尔 - 广义帕累托模型的巨灾损失分布拟合

吕梦璐

南京信息工程大学数学与统计学院, 江苏 南京

收稿日期: 2023年2月24日; 录用日期: 2023年4月14日; 发布日期: 2023年4月23日

摘要

我国是世界上台风灾害频发的国家之一, 由此造成的人员伤员和财产损失逐年上升, 影响经济发展。本文以台风风暴潮数据为例, 采用1989年~2021年我国沿海地区台风风暴潮直接经济损失数据作为样本, 对数据进行预处理后同时引入以威布尔分布、帕累托分布和广义帕累托分布为基础构建的组合分布模型进行拟合, 结果表明: 可变权重的威布尔 - 广义帕累托模型的拟合效果最好, 并以此为依据为我国台风巨灾损失提供分析方法。

关键词

组合分布模型, 巨灾损失, 广义帕累托分布

Catastrophe Loss Distribution Fitting Based on Variable Weighted Weibull-Generalized Pareto Model

Menglu Lyu

School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing Jiangsu

Received: Feb. 24th, 2023; accepted: Apr. 14th, 2023; published: Apr. 23rd, 2023

Abstract

China is one of the countries in the world with frequent typhoon disasters, and the resulting casualties and property losses have increased year by year, affecting economic development. Taking typhoon storm surge data as an example, this paper uses the direct economic loss data of typhoon storm surge in coastal areas of China from 1989 to 2021 as a sample, preprocesses the data, and

introduces the combined distribution model constructed on the basis of Weibull distribution, Pareto distribution and generalized Pareto distribution for fitting, and the results show that the Weibull-Generalized Pareto model with variable weight has the best fitting effect, and provides an analysis method for typhoon catastrophe loss in China based on this.

Keywords

Combinatorial Distribution Model, Catastrophe Loss, Generalized Pareto Distribution

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

我国是世界上台风灾害频发的国家之一,由此造成的人员伤亡和财产损失逐年上升,对经济的发展、社会的协调、人们的生活等各方面都产生了重要的影响。巨灾不可避免,但巨灾造成的损失又非常大,在巨灾损失分析的研究中,赵智红和李兴绪[1] (2010)在构建模型的过程中给出了极值理论的最大吸引域检验问题,然后利用不同方法讨论了最优门限值的选取问题,并在 POT 模型下利用广义帕累托分布对巨额损失分布进行拟合,然后在复合泊松分布的框架下讨论了险位超赔再保险的纯保费计算问题。Yao [2]等(2013)提出了极值理论的峰值法,运用帕累托分布和威布尔分布来拟合形状参数大于零的损失分布样本,并使用 CVaR 模型来衡量商业银行基本条件下的操作风险。叶小岭[3]等(2011)以台风属性等指标作为输入向量,以台风造成的损失作为输出向量,建立了预测影响浙江地区台风损失的理想人工神经网络模型,实验结果表明运用 BP 神经网络模型预测台风造成的灾害损失是可行的。张晗霄和周月癸[4] (2015)以河南省洪涝巨灾为实际背景,根据极值理论中的超越阈值模型(POT),用广义帕累托分布(GPD)结合三种参数估计对巨灾损失数据进行拟合,得到了不同置信度要求下的 VaR 值,为农业自然灾害预警、粮食安全储备和国家财政预算提供科学依据。Bakar [5]等(2015)提出了一种新的构建双参数组合模型的方法,即组合模型中的混合权重和阈值两个参数由组合模型中的其他参数表示。

近几年,王永茂和杨晓婷[6] (2021)在对数广义误差分布的基础上,运用广义帕累托分布拟合数据的厚尾部分,并加入可变权重组合分布模型,验证了可变权重的对数广义误差-广义帕累托组合分布模型的拟合效果更好。Solari [7]等(2017)在极值理论的基础上,提出了一种使用拟合优度 p 值进行自动阈值估计的方法,为今后极值理论的阈值选取提供了更多的可选方法。郭静和张连增[8] (2021)将截断形式的混合 Erlang 分布与 Pareto 分布进行组合,通过实证分析发现构建的混合 Erlang-Pareto 组合分布可以较好地拟合地震巨灾数据。郝婧和刘强[9] (2022)针对传统神经网络对台风风暴潮灾害的损失测度易陷入局部极值和预测准确度低的问题,提出用天牛须搜索(BAS)算法来优化 Elman 神经网络模型,并用此模型预测受灾人口、直接经济损失、海水养殖受灾面积、海岸工程损毁 4 种灾情评估指标,结果表明该模型预测精度更高。可以看出对巨灾损失方面的研究,大多采用的是单一模型,因此本文利用威布尔分布、帕累托分布和广义帕累托分布构建组合分布模型对我国沿海地区台风风暴潮损失数据进行建模分析。

2. 组合分布模型的构建

2.1. 预备模型

1) 威布尔分布

威布尔的密度函数和分布函数为:

$$f_w(x) = \begin{cases} 0 & x \leq 0 \\ \left(\frac{\tau}{x}\right)\left(\frac{x}{\varphi}\right)^{\tau} \exp\left\{-\left(\frac{x}{\varphi}\right)^{\tau}\right\} & x > 0 \end{cases} \quad (2-1)$$

$$F_w(x) = 1 - \exp\left\{-\left(\frac{x}{\varphi}\right)^{\tau}\right\} \quad (2-2)$$

其中 $\tau > 0, \varphi > 0$, τ 是形状参数, φ 是尺度参数。形状参数控制着密度函数曲线幅度的变化, 当形状参数等于 1 时, 威布尔分布会退化为指数分布, 当形状参数大于 1 时, 形状参数越来越大, 图像的峰值就会越来越高。尺度参数控制密度函数曲线形状的变化, 与形状参数相反, 当尺度参数增大时, 图像的峰值反而会降低。

2) 帕累托分布

早先帕累托分布主要用于研究经济现象, 后面在环境学、灾害学有了广泛的应用, 在应用过程中发现一些数据具有厚尾特征, 因此帕累托分布经常被用来对尾部数据进行建模分析。

帕累托分布的密度函数和分布函数表示为:

$$f_p = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, x > \theta \quad (2-3)$$

$$F_p(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^{\alpha} \quad (2-4)$$

3) 广义帕累托分布

广义帕累托分布是帕累托分布的推广形式, 是一种右偏态分布, 能够基于理论参数为尾部数据进行建模, 其密度函数和分布函数可以表示为:

$$f_{gp}(x) = \frac{1}{\alpha} \left(1 + \frac{\lambda x}{\alpha}\right)^{-\frac{1}{\lambda}-1} \quad (2-5)$$

$$F_{gp}(x) = \begin{cases} 1 - \left(1 + \lambda \frac{x}{\alpha}\right)^{-\frac{1}{\lambda}} & \lambda \neq 0 \\ 1 - \exp\left(-\frac{x}{\alpha}\right) & \lambda = 0 \end{cases} \quad (2-6)$$

2.2. 模型构建

原先组合分布模型构建出来的权重是固定, 而 Scollink [10] (2007) 提出了可变权重, 用可变权重构建组合分布模型, 可以使组合分布模型形式更加灵活。本文将固定权重的组合分布模型和可变权重的组合分布模型进行研究分析。

固定权重的组合分布模型可以表示为:

$$f(x) = \begin{cases} cf_w(x) & 0 < x \leq \theta \\ cf_p(x) & \theta < x < \infty \end{cases} \quad (2-7)$$

对于可变权重的组合分布模型可以表示为

$$f(x) = \begin{cases} rf_w(x)/F_w(\theta) & 0 < x < \theta \\ (1-r)f_p(x) & \theta < x < \infty \end{cases} \quad (2-8)$$

利用这两种组合方式将上述三种分布组合在一起，构建出固定权重下的威布尔分布 - 帕累托组合分布模型、可变权重的威布尔分布 - 帕累托组合分布模型和可变权重的威布尔 - 广义帕累托组合分布模型。

1) 固定权重下的威布尔分布 - 帕累托组合分布模型

为了保证这两个分布在连接点处 θ 具有连续性和可导性，即 $f(\theta-) = f(\theta+)$ ， $f'(\theta-) = f'(\theta+)$ 。根据这两个条件可以计算得到固定权重下的威布尔分布 - 帕累托组合分布模型为：

$$f_{w,p}(x) = \begin{cases} \frac{(m+1)^2 \alpha \left(\frac{x}{\theta}\right)^{\alpha m}}{(2m+1)x \left(\frac{x}{\theta}\right)^{\alpha m}} \exp\left\{-\left(\frac{m+1}{m}\right)\left(\frac{x}{\theta}\right)^{\alpha m}\right\}, & 0 < x \leq \theta \\ \frac{m+1}{2m+1} \left(\frac{\alpha}{x}\right) \left(\frac{\theta}{x}\right)^{\alpha}, & \theta < x < \infty \end{cases} \quad (2-9)$$

其中， $\alpha > 0, \theta > 0$ 。并且在推导过程当中我们发现

$$\exp\left(1 + \frac{1}{m}\right) = m + 1 \quad (2-10)$$

$$1 + \frac{\tau}{\alpha} = \exp\left(1 + \frac{\alpha}{\tau}\right) = \exp\left\{\left(\frac{\theta}{\varphi}\right)^{\tau}\right\} \quad (2-11)$$

联立(2-10)式和(2-11)式可以得到

$$m = \frac{\tau}{\alpha} \quad (2-12)$$

2) 可变权重的威布尔分布 - 帕累托组合分布模型

同样基于连续性和可导性两个条件，可以计算得到可变权重的威布尔分布 - 帕累托组合分布模型为：

$$f(x) = \begin{cases} \frac{\left(\frac{\alpha}{\tau}\right) \exp\left(\left(\frac{\theta}{\varphi}\right)^{\tau}\right)}{\left(\frac{\alpha}{\tau}\right) \exp\left(\left(\frac{\theta}{\varphi}\right)^{\tau}\right) + 1} \left(\frac{\tau + \alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\tau} \exp\left\{-\left(\frac{x}{\theta}\right)^{\tau} \left(1 + \frac{\alpha}{\tau}\right)\right\}, & 0 < x \leq \theta \\ \frac{\left(\frac{\alpha}{\tau}\right) \exp\left(\left(\frac{\theta}{\varphi}\right)^{\tau}\right)}{\left(\frac{\alpha}{\tau}\right) \exp\left(\left(\frac{\theta}{\varphi}\right)^{\tau}\right) + 1} \left(\frac{\alpha \theta^{\alpha}}{x^{\alpha+1}}\right), & \theta < x < \infty \end{cases} \quad (2-13)$$

在推导过程中应用可导性条件可以得到

$$\left(\frac{\theta}{\varphi}\right)^{\tau} = \frac{\alpha}{\tau} + 1 \quad (2-14)$$

3) 可变权重的威布尔 - 广义帕累托组合分布模型

根据广义帕累托和威布尔的密度函数，并应用上述提到的连续性和可导性条件计算得到可变权重的威布尔 - 广义帕累托组合分布模型为：

$$f(x) = \begin{cases} \frac{\left(\frac{\alpha}{\tau}\right) \exp\left(\left(\frac{\theta}{\varphi}\right)^\tau\right) \left(\frac{\tau}{x}\right) \left(\frac{x}{\theta}\right)^\tau \left(\frac{\alpha\theta - \lambda}{(\lambda + \theta)\tau} + 1\right) \exp\left\{-\left(\frac{x}{\theta}\right)^\tau \left(\frac{\alpha\theta - \lambda}{(\lambda + \theta)\tau} + 1\right)\right\}}{\left(\exp\left(\left(\frac{\theta}{\varphi}\right)^\tau\right) - 1\right) \left(\frac{\alpha}{\tau}\right) + \frac{\lambda + \theta}{\theta} \left(\frac{\theta}{\varphi}\right)^\tau}, & 0 < x \leq \theta \\ \frac{\left(\frac{\theta}{\varphi}\right)^\tau (\lambda + \theta) \alpha (\lambda + \theta)^\alpha}{\left(\exp\left(\left(\frac{\theta}{\varphi}\right)^\tau\right) - 1\right) \left(\frac{\alpha}{\tau}\right) \theta + (\lambda + \theta) \left(\frac{\theta}{\varphi}\right)^\tau (\lambda + x)^{\alpha+1}}, & \theta < x < \infty \end{cases} \quad (2-15)$$

也是利用可导性条件得到下面两个等式

$$\left(\frac{\theta}{\varphi}\right)^\tau = \frac{\alpha\theta - \lambda}{(\lambda + \theta)\tau} + 1 \quad (2-16)$$

$$\left(\frac{\theta}{\varphi}\right)^\tau = \frac{\theta^2(\alpha + 1)}{(\lambda + \theta)^2(1 - \tau)\tau} - \frac{1}{\tau} \quad (2-17)$$

综合(2-16)式和(2-17)式可以得到 α 与其他参数的关系

$$\alpha = \frac{\theta^2 - (\tau(\lambda + \theta) + \theta)(1 - \tau)(\lambda + \theta)}{\theta(1 - \tau)(\lambda + \theta) - \theta^2} \quad (2-18)$$

3. 实证分析

3.1. 数据来源

本文的数据来源主要来自于《中国海洋灾害公报》和国家统计局。其中《中国海洋灾害公报》提供了全国沿海地区的海洋灾害情况，详细记录了每年发生的受灾类型、受灾次数。受灾时间、受灾程度、受灾地区和受灾特点。依据《中国海洋灾害公报》所提供的信息，本文整理得到了我国沿海地区 1989 年~2021 年由台风风暴潮灾害造成的直接经济损失数据(亿元)，并将其作为观察样本，共 211 条损失数据。考虑到各年物价水平不同的影响，本文还选取了国家统计局公布的 1989 年~2021 年反映物价水平的居民消费价格指数(CPI)对损失数据进行调整，具体的调整方法是将 1989 年的居民消费价格指数作为定基指数，把每次的台风风暴潮灾害经济损失数据乘以 1989 年的 CPI，再除以年的 CPI，就能得到调整后的经济损失数据(见表 1)。最后对调整后的经济损失数据进行筛选整理，选取直接经济损失额超过 1 亿元的作为本文的研究样本，共计 126 条数据。

Table 1. Adjusted direct economic loss partial data

表 1. 经调整后的直接经济损失部分数据

年份	地区	直接经济损失(亿元)	CPI	调整后的直接经济损失(亿元)
1989	浙江省	13.23	209.9	13.23
1989	浙江省	10.5	209.9	10.5
1989	广东省	12.6	209.9	12.6
1989	三亚一带	19	209.9	19
1990	福建省	16.262	216.4	15.77

Continued

1990	福建省	12.2	216.4	11.83
1990	浙江省	3.61	216.4	3.50
1990	浙江省	1.42	216.4	1.38
1991	海南省	5.2	223.8	4.88
1991	海南省	5.9	223.8	5.53
1991	广东省	4.7	223.8	4.41
1992	福建省	9.15	238.1	8.07
1992	浙江省	31.5	238.1	27.77
1992	江苏省	3.2	238.1	2.82
1992	山东省	41.51	238.1	36.59
1992	天津	3.99	238.1	3.52
1992	河北省	3.2	238.1	2.82

3.2. 样本的描述性统计特征

Table 2. Descriptive statistics of typhoon storm surge loss amount in coastal areas of China

表 2. 我国沿海地区台风风暴潮损失金额描述性统计量

主要统计量	统计量值
个数	126
最小值	1.00
最大值	91.67
均值	12.5445
均值标准误差	1.36892
方差	236.115
偏度	2.821
偏度标准误差	0.216
峰度	9.204
峰度标准误差	0.428

表 2 反映了我国沿海地区台风风暴潮经济损失数据的描述性统计特征。首先从表中可以看出样本方差很大，方差是每个样本值与全体样本值的平均数之差的平方值的平均数，是用来衡量样本数据与期望值相差的度量值，样本方差越大说明样本数据中存在一部分与均值偏离程度较大的数据，即说明样本中存在一些极端损失。偏度主要是用来衡量数据分布的偏斜程度，当偏度小于 0 时称为左偏态，此时表现为左边的尾部比右边的尾部要长，当偏度大于 0 时称为右偏态，表现为右边的尾部比左边的尾部要长，由表 2 可知偏度值为 2.821，说明样本分布偏右，大部分的样本数据位于分布左侧，少部分较大的样本数据位于分布的右侧。峰度反映了峰部的尖度，当峰度值等于 3 时，样本为正态分布；当峰度值大于 3 时，样本为厚尾分布，由表可知峰度值为 9.204，说明该样本的峰度大于正态分布的峰度值，该曲线的峰度更

高更尖锐，曲线的两边尾部较粗。故通过上述描述统计量的分析可知，我国沿海地区的台风风暴潮数据具有尖峰厚尾的特点，但仍需进一步分析。

3.3. 数据的厚尾性检验

厚尾分布是一种重尾分布，它的偏度和峰度都偏大，与常见的分布如正态分布、指数分布相比尾部更厚、峰处会更尖。对于相同数据来说，厚尾分布出现极端值的概率是大于正态分布或者指数分布出现极端值的概率。为了后续的参数估计和模型构建，必须保证样本数据具有尖峰厚尾的特点。厚尾性检验可以通过 QQ 图(图 1)和直方图(图 2)来验证。

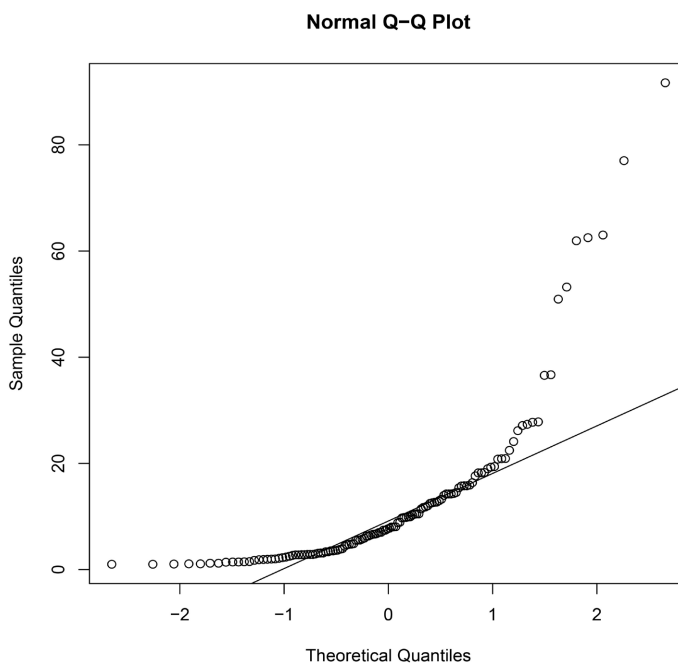


Figure 1. Q-Q map of typhoon storm surge damage from 1989 to 2021
图 1. 1989 年~2021 年台风风暴潮灾害损失 Q-Q 图

1) Q-Q 图

设 x_1, x_2, \dots, x_n 为来自标准正态分布总体 X 的 n 个样本，分位数 q_i 定义如下：

$$P\{X \leq q_i\} = \int_{-\infty}^{q_i} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = p_i \quad (3-1)$$

这里 $p_i = \left(i - \frac{1}{2}\right) / n, i = 1, 2, \dots, n$ ，显然 q_i 由 p_i 所唯一确定。若数据 x_1, x_2, \dots, x_n 的分布与正态分布非常接近，则点应大致构成一条直线。

Q-Q 图是根据变量的分位数对应于理论分布的分位数绘制的散点图，横坐标为某一样本的分位数，纵坐标为另一样本的分位数，横坐标与纵坐标组成的散点图代表同一个累计概率所对应的分位数。若经验分布与理论分布一致，即散点图落在直线 $y = x$ 附近。用 Q-Q 图做厚尾性检验是指若 Q-Q 图的中部为直线，上端向右偏离该直线，呈向下倾斜趋势，则该分布的上尾具有厚尾性；若 Q-Q 图的中部为直线，下端向左偏离该直线，呈向上倾斜趋势，则该分布的下尾具有厚尾性。利用 SPSS 软件画出我国沿海地区台风风暴潮灾害损失数据的 Q-Q 图(图 1)发现两端都是呈倾斜状态，并且中部偏离对角线向上凹，说明

损失数据具有厚尾特征，其尾部比正态分布的尾部更厚。

2) 直方图

从图 2 的直方图可以直观地看出数据明显具有尖峰厚尾特征，说明该数据可以用于后续的研究。

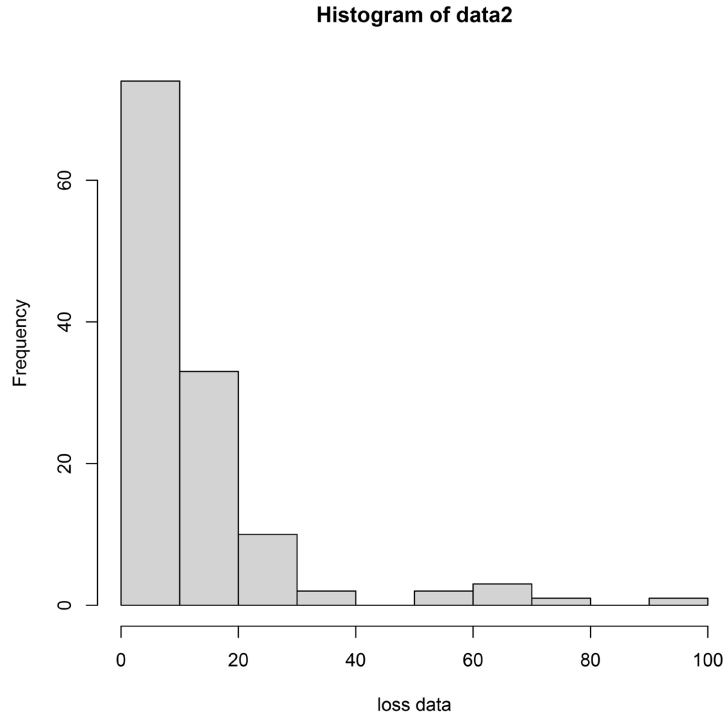


Figure 2. Histogram of typhoon storm surge disaster losses from 1989 to 2021
图 2. 1989 年~2021 年台风风暴潮灾害损失直方图

3.4. 阈值选取

在对组合分布模型进行参数估计之前，得先确定阈值。阈值过高会导致参数估计方差偏高，阈值过低会导致参数估计不够相合。故阈值选取准确与否会影响后面模型参数估计和模型检验。仅仅选择一个方法确定阈值会有失偏颇，本文主要采用平均超额损失函数和 hill 图法两种方法进行阈值选取，具体做法就是将台风风暴潮灾害损失数据导入到 R 软件，利用 hill 函数和 meplot 函数分别画出平均超额损失函数图和 hill 图。

1) 平均超额损失函数

平均超额函数定义为：

$$e_n(\mu) = E(X - U | x > \mu) = \sum_{i=1}^n (X_i - \mu) / n \quad (3-2)$$

其中， $x_i, i=1, 2, \dots, n$ ，表示超过阈值的样本观测值，平均超额损失函数图点 $\{\mu, e_n(\mu), \mu > 0\}$ 集合，以 μ 为横轴， $e_n(\mu)$ 为纵轴，可得到相应的函数图。阈值选取主要是根据样本平均超额损失函数是否呈线性趋势来判断。如果在某个观测值之后平均超额损失函数曲线逐渐呈现线性趋势，则可以选择这个超过的观测值作为所需要确定的阈值。

2) hill 图法

Hill 图法确定阈值主要是假设 $X_1 > X_2 > \dots > X_n$ 表示独立同分布的次序统计量，可得出

$$\gamma_{k,n} = \frac{1}{k} \sum_{i=1}^k [\ln X(i) - \ln X(k)] \quad (3-3)$$

横轴为 k , 纵轴为 $\gamma_{k,n}^{-1}$, 即 hill 图法为坐标 $(k, \gamma_{k,n}^{-1})$ 的点构成的曲线。选取图中尾部指数稳定区域的起始点的横坐标对应的数值作为阈值。

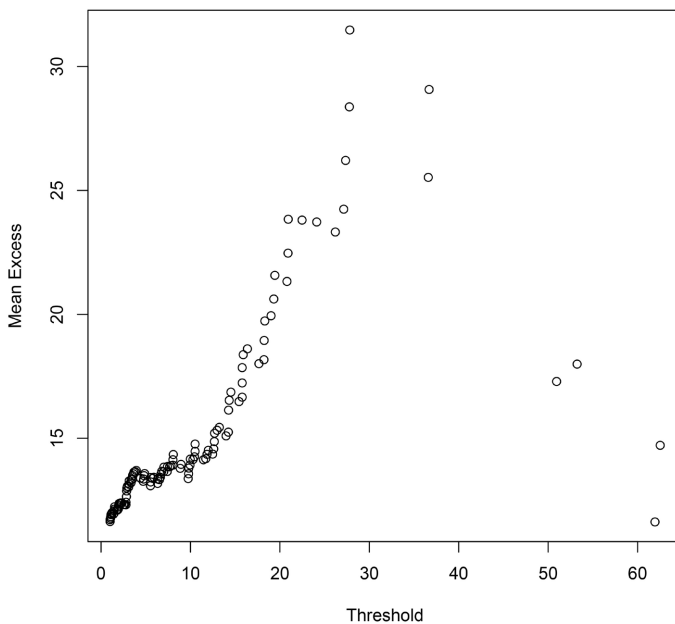


Figure 3. Average excess loss function plot for loss data
图 3. 损失数据的平均超额损失函数图

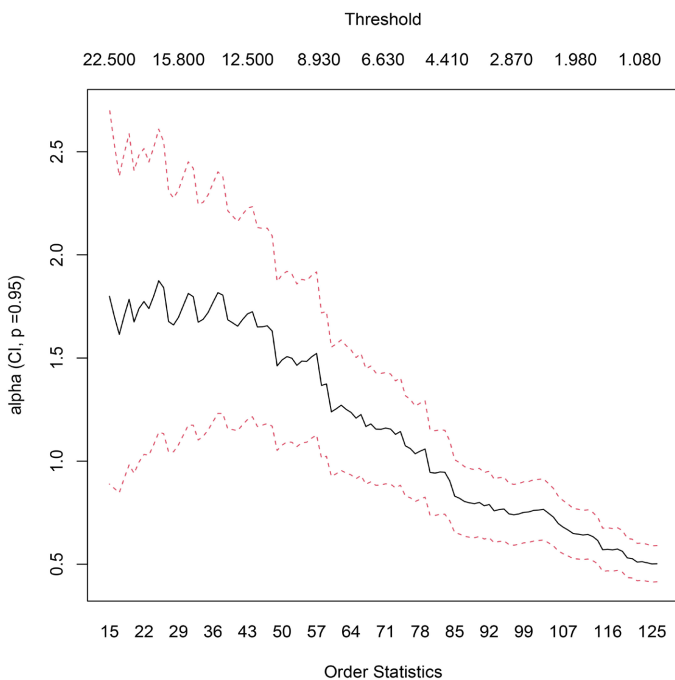


Figure 4. Hill chart of loss data
图 4. 损失数据的 hill 图

由图 3 可以看出图像在区间(5, 15)后, 函数曲线逐渐趋于线性, 故先初步确定阈值在这个区间范围内, 为了更加精确阈值的区间范围, 我们还结合 hill 图做更细致的检验。由图 4 可见, 图形左边呈上下波动且波动较大, 图形右边大致从区间(8.93, 12.5)开始近似为直线。综合以上两种方法, 我们可以确定阈值在区间(8.93, 12.5)内。

3) 峰度法

为了具体确定阈值, 我们还尝试用峰度法。

样本峰度的计算公式为:

$$K_n = \frac{1}{n \sum_{i=1}^n (X_i - m_n)^4} \cdot \frac{1}{(S_n^2)^2} \tag{3-4}$$

其中: $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - m_n)^2, m_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。

首先计算出样本峰度为 9.204, 此时样本峰度小于 3, 故我们将台风风暴潮损失数据中最大的损失值剔除, 重新计算样本峰度。以此循环, 前面七次计算出来的样本峰度均大于 3, 第八次计算出来的样本峰度为 1.371, 此时剩余的损失数据中最大的损失值为 36.69。显然, 根据峰度法计算出来的阈值并不在上述阈值估计区间, 并且在 126 条数据中, 损失值超过 36.69 的只有 7 个。过少的数据量进行参数估计会导致模型的拟合效果不够, 故排除掉峰度法计算出来的阈值。

3.5. 参数估计和模型检验

1) 参数估计

在众多参数估计方法中, 极大似然估计方法最为常用, 且数据量不大时极大似然估计与真实数据最为接近, 因此本文采用极大似然估计对组合分布模型中的参数进行估计。本文所使用的三个组合模型是基于威布尔分布、广义帕累托模型和帕累托模型进行构建的, 经分析可知, 只要求得威布尔分布的参数, 就能求出固定权重的威布尔 - 帕累托分布和可变权重威布尔 - 帕累托分布的参数, 求出广义帕累托分布的参数才能求出可变权重威布尔 - 广义帕累托分布的参数。利用极大似然估计求参数的步骤具体如下:

根据威布尔分布的密度函数建立对数似然函数

$$\ln L = -\tau n \ln \varphi + n \ln \tau + (\varphi - 1) \sum_{i=1}^n \ln x_i - \frac{1}{\varphi^\tau} \sum_{i=1}^n x_i^\tau \tag{3-5}$$

分别对 τ 和 φ 求导, 并令其等于零

$$\frac{\alpha \ln L}{\alpha \tau} = -n \ln \varphi + \frac{n}{\tau} + \sum_{i=1}^n \ln x_i + \frac{\ln \varphi}{\varphi^\tau} \sum_{i=1}^n x_i^\tau - \frac{1}{\varphi^\tau} \sum_{i=1}^n x_i^\tau \ln x_i = 0 \tag{3-6}$$

$$\frac{\alpha \ln L}{\alpha \varphi} = -\frac{\tau \cdot n}{\varphi} + \tau \frac{1}{\varphi^{\tau+1}} \sum_{i=1}^n x_i^\tau = 0 \tag{3-7}$$

经变形得

$$\begin{cases} \frac{1}{\tau} + \frac{1}{n} \sum_{i=1}^n \ln x_i = \frac{\sum_{i=1}^n x_i^\tau \ln x_i}{\sum_{i=1}^n x_i^\tau} \\ \varphi^\tau = \frac{1}{n} \left(\sum_{i=1}^n x_i^\tau \right) \end{cases} \tag{3-8}$$

根据该方程组可解得参数 τ 和 φ 的估计 $\hat{\tau}$ 和 $\hat{\varphi}$ ，并利用 Matlab 数学软件中的 wblbfit 函数进行求解得到 $\tau = 1.8350$ ， $\varphi = 4.5284$ 。将这两个参数估计值代入(2-12)式、(2-14)式就能求出固定权重的威布尔分布 - 帕累托分布和可变权重的威布尔分布 - 帕累托分布组合分布模型的参数估计值。同样也用极大似然估计去求解广义帕累托分布的参数估计值，由于参数估计方程为非线性方程，没有显式解，因此只能用数值法求解，本文采用的是牛顿迭代法求出可变权重的威布尔 - 广义帕累托分布。参数估计表如下表 3 所示：

Table 3. Table of parameter estimates
表 3. 参数估计值表

模型	参数估计值
固定权重的威布尔分布 - 帕累托分布	$\tau = 1.8350$
	$\varphi = 4.5284$
	$\alpha = 4.5446$
	$\tau = 1.8350$
可变权重的威布尔分布 - 帕累托分布	$\varphi = 4.5284$
	$\alpha = 0.6416$
	$\tau = 1.8350$
	$\varphi = 4.5248$
可变权重的威布尔 - 广义帕累托分布	$\alpha = 27.0705$
	$\lambda = -0.1907$

2) 模型检验

在求得参数后还需要对模型进行检验。将所求得的参数带回组合分布模型并进行 K-S 检验和卡方检验。

K-S 检验是一种非参数假设检验，基于累积分布函数用以检验某样本是否服从某一分布，或者两样本是否服从相同分布。K-S 检验统计量定义为： $D_n = \sup_x |F_n(x) - F(x)|$ ，其中： n 为样本量； $F_n(x)$ 为经验分布函数； $F(x)$ 为拟合分布函数。

卡方检验就是统计样本的实际观测值与理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度就决定卡方值的大小，如果卡方值越大，二者偏差程度越大；反之，二者偏差越小；若两个值完全相等时，卡方值就为 0，表明理论值完全符合。其检验统计量为 $\chi^2 = \sum_{i=1}^k (f_i - np_i)^2 / np_i$ 。

分别对三种组合分布模型进行 K-S 检验和卡方检验，检验结果如下表 4 所示。

Table 4. Model checklist
表 4. 模型检验表

分布	K-S 检验	卡方检验
固定权重的威布尔 - 帕累托分布	0.076	6.423
可变权重的威布尔 - 帕累托分布	0.048	12.387
可变权重的威布尔 - 广义帕累托	0.092	3.217

根据表 4 的结果可知，可变权重的威布尔 - 广义帕累托的组合分布模型要优于固定权重的威布尔分布 - 帕累托和可变权重的威布尔分布 - 帕累托组合分布。

4. 小结

在巨灾损失分布方面,常用的厚尾分布有威布尔分布、广义帕累托分布、对数正态分布、伽马分布等分布,本文在这些单一分布的基础上,结合固定权重和可变权重的组合分布方法,构建出固定权重的威布尔-帕累托分布、可变权重的威布尔-帕累托分布和可变权重的威布尔-广义帕累托分布,并进行实证分析发现可变权重的威布尔-广义帕累托分布对台风风暴潮损失数据拟合效果最好,为我国台风巨灾损失分析提供一定的参考依据。

参考文献

- [1] 赵智红, 李兴绪. 非寿险中巨额损失数据的拟合与精算[J]. 数理统计与管理, 2010, 29(2): 336-347.
- [2] Yao, F., Wen, H. and Luan, L. (2013) CVaR Measurement and Operational Risk Management in Commercial Banks According to the Peak Value Method of Extreme Value Theory. *Mathematical & Computer Modelling*, **58**, 15-27. <https://doi.org/10.1016/j.mcm.2012.07.013>
- [3] 叶小岭, 刘程波, 张颖超, 范金平. 基于 BP 神经网络的浙江台风损失预测[J]. 信息技术, 2011, 35(10): 59-61.
- [4] 张晗霄, 周月癸. 基于 POT-GPD 模型的巨灾损失分布及参数估计的比较[J]. 河南科技, 2015(3): 151-153.
- [5] Bakar, S.A., Hamzah, N.A., Maghsoudi, M. and Nadarajah, S. (2015) Modeling Loss Data Using Composite Models. *Insurance: Mathematics and Economics*, **61**, 146-154. <https://doi.org/10.1016/j.insmatheco.2014.08.008>
- [6] 王永茂, 杨晓婷. 基于 LogGED-GPD 模型的巨灾损失分布拟合[J]. 郑州大学学报(理学版), 2021, 53(3): 100-104.
- [7] Solari, S., Eguen, M., Polo, M.J. and Losada, M.A. (2017) Peaks over Threshold (POT): A Methodology for Automatic Threshold Estimation Using Goodness of Fit p-Value. *Water Resources Research*, **53**, 2833-2849. <https://doi.org/10.1002/2016WR019426>
- [8] 郭静, 张连增. 基于 Mixed Erlang-Pareto 组合分布的巨灾风险评估——以中国地震灾害为例[J]. 统计与信息论坛, 2021, 36(3): 119-128.
- [9] 郝婧, 刘强. 基于改进 ELman 神经网络的台风风暴潮损失测度[J]. 中国海洋大学学报(自然科学版), 2022, 52(8): 71-76.
- [10] Scollnik, D.P.M. (2007) On Composite Lognormal-Pareto Models. *Scandinavian Actuarial Journal*, **2007**, 20-33. <https://doi.org/10.1080/03461230601110447>