Pakistan Journal of Statistics and Operation Research

A Novel Chen Extension for Risk Analysis with MOOP and PORT-VAR Assessments under Hydrological Flow Data and Financial Case Study

Mohamed Ibrahim[1,*], Abdullah H. Al-Nefaie[1], Nadeem S. Butt[2], Ahmad M. AboAlkhair[1], Haitham M. Yousof[3] , and Rehab Shehata Mahmoud[3]

* Corresponding Author

[1]Department of Quantitative Methods, School of Business, King Faisal University, Al Ahsa 31982, Saudi Arabia; (M.I.): miahmed@kfu.edu.sa; (A.H.A.): aalnefaie@kfu.edu.sa; (A.M.A.): aaboalkhair@kfu.edu.sa
[2]Department of Family and Community Medicine, King Abdul Aziz University, Jeddah, Kingdom of Saudi Arabia; nshafique@kau.edu.sa
[3]Department of Statistics, Mathematics and Insurance, Faculty of Commerce, Benha University, Benha, Egypt; haitham.yousof@fcom.bu.edu.eg; (R.S.M.): REHAB.MAHMOUD@fcom.bu.edu.eg

Abstract

This paper introduces a new extension of the Chen distribution, designed to better model extreme low-flow events in hydrology and rare events in the medical field. The proposed model incorporates asymmetrical and heavy-tailed behavior, making it particularly useful for analyzing extreme values in complex real datasets. We derive the mathematical properties of the BGC distribution and apply two advanced analytical techniques: the Mean-of-Order-P (MOOP) method to determine the optimal value of P (referred to as Opt-P), and the Peaks Over Threshold Value-at-Risk (PORT-VaR) approach to identify and assess critical extreme events. These methods are applied to real datasets including relief times, minimum river flow data from the Cuiabá River, and U.S. indemnity losses from general liability claims. The MOOP analysis shows that increasing the order P leads to reduced Mean Squared Error (MSE) and Bias, indicating improved estimation accuracy. For example, in the relief times dataset, MSE decreases from 0.64 at P=1 to 0.3844 at P=5. Similarly, for the minimum flow data, MSE drops from 4402.88 to 3684.27 with increasing P, highlighting the benefits of higher-order statistics in capturing central tendencies. Using PORT-VaR, we analyze extreme peaks under varying confidence levels (50%, 70%, 90%, and 99%) and compute key risk indicators such as Value-at-Risk (VaR) , Tail Value-at-Risk (TVaR) , Mean Excess Loss (MEXL) , Tail Variance (TV) , and Tail Mean Variance (TMV) . In the relief times dataset, VaR increases from 1.70 at 50% confidence to 3.055 at 99% confidence, demonstrating growing risk exposure at higher confidence levels. For the minimum flow data, VaR rises from 115.925 at 50% to 157.169 at 99%, underscoring the importance of adaptive risk thresholds in managing water scarcity and dam safety. A financial case study using U.S. indemnity loss data further validates the robustness of the BGC model in capturing tail behavior and estimating extreme risks. At the 99% confidence level, VaR reaches 170400 (in thousands of USD), and MEXL is 203411, illustrating the nonlinear growth of risk in heavy-tailed insurance claims. Finally, a comparative study under a historical financial claims data through an application.

Keywords: Chen distribution; Right skewed extreme values; Asymmetric flow data; Mean-of-order-P; Optimal order of P; Value-at-risk; Peaks over a random threshold.

Mathematical Subject Classification: 62N02; 62E12; 62N01; 62E10.

1. Introduction

Modeling extreme events, such as severe droughts, rare medical outcomes, or catastrophic insurance losses, requires robust statistical tools capable of capturing heavy tails, skewness, and asymmetry in data. While traditional distributions like the lognormal, gamma, and Weibull have been widely used for risk analysis, they often fall short

when applied to real datasets that exhibit extreme variability, asymmetric behavior, or long-tailed distributions. These limitations become especially problematic in fields such as hydrology, public health, and actuarial science, where precise estimation of tail risks is critical for decision-making. The standard Chen distribution, known for its ability to model bathtub-shaped hazard rates and heavy tails, has shown promise in handling extreme values. However, it lacks sufficient flexibility to accurately represent complex patterns found in modern datasets, particularly those with varying degrees of skewness and kurtosis. Existing extensions of the Chen model attempt to improve its performance, but they still struggle to capture the full spectrum of extreme value behavior, especially under high confidence levels where accurate estimation of VaR, TVaR, and related indicators becomes essential. This research addresses the need for a more flexible and adaptive model that can effectively handle asymmetric and heavy-tailed data across multiple domains. The new Chen distribution is introduced as an extension designed to overcome the shortcomings of current models by incorporating additional shape parameters that enhance its ability to fit and predict extreme low-flow events, medical relief times, and large indemnity losses.

The proposed distribution has three parameters and is called the Burr generalized Chen (BGC) distribution. After deriving sufficient statistical and mathematical properties of the new distribution, the new distribution was used in statistical modeling processes through risk analysis and assessment processes.

The risk indicators like TVAR, VAR, MEXL, TV, MOO^p, PORT-VAR and TMV are essential tools for determining maximum possible losses under the extreme event. The BGC distribution is used to improve the precision of these estimations by spotting extreme values (see Figueiredo et al. (2017) and Alizadeh et al. (2024)). According to Klugman et al. (2012), the degrees of exposure are commonly referred to as key risk indicators. These primary risk indicators give actuaries and risk managers knowledge of the level of risk to which the company is exposed (see Wirch (1999), Tasche (2002), Furman and Landsman (2006), and Acerbi (2002) for the VAR, conditional-VAR (CVAR), tail-Variance (TV)) see also Lane (2000), Artzner (1999) and Klugman et al. (2012) for more details about the above-mentioned indicators. Additionally, risk exposure is an actuarial estimation of the potential loss that might develop in the future because of a specific action or occurrence (for more advanced and related works, see Alizadeh et al. (2023), Hamedani et al. (2023), Loubna et al. (2024) and Yousof et al. (2024)).

Assume that $\boldsymbol{Y}$ refers to a loss random variable (RV) with a generalized Chen (GZC) model (see Chaubey and Zhang (2015)), then its corresponding cumulative distribution function (CDF) is given by

$$G_{\theta_2,\theta_3}(y) = \left(1 - \exp\left\{\left[1 - \exp\left(y^{\theta_3{}^{-1}}\right)\right]\right\}\right)^{\theta_2}, \tag{1}$$

where $y > 0, \theta_2 > 0$ and $\theta_3 > 0$. The CDF $G_{\theta_2,\theta_3}(y)$ is monotonically non-decreasing in y. This follows from the fact that the function inside the outer exponent, $1 - \exp\left\{\left[1 - \exp\left(y^{\theta_3{}^{-1}}\right)\right]\right\}$, is increasing as y increases. Consequently, raising this function to a positive power θ_2 preserves monotonicity.

The behavior of $G_{\theta_2,\theta_3}(y)$ as $y \to \infty$ reflects the tail properties of the loss distribution. Specifically, as y increases, the term $\exp\left(y^{\theta_3{}^{-1}}\right)$ decays rapidly, indicating that the CDF approaches one quickly, and the PDF decreases exponentially, suggesting heavy-tail characteristics. For this reason, we are motivated to extend the C model to study the tail behavior of medical and hydrological data. When $\theta_2 = 1$ and $\theta_3 = 1$, the GZC model simplifies to a special case of the generalized extreme value distribution, allowing it to model various types of extreme values. Tail behavior in medical data, such as relief data, is crucial for understanding extreme medical events, such as rare but severe health crises.

Tail behavior has an impact on the calculation of catastrophic claims in medical insurance. Understanding the tail risk allows insurers to properly price policies and set aside suitable reserves to cover severe claims, resulting in more robust financial planning and risk management (Mohamed et al., 2024). Tail behavior is important in hydrology because it helps predict the likelihood of extreme low-flow or high-flow occurrences like droughts and floods. Understanding these extreme tails allows for improved flood risk management and the development of effective mitigation solutions. In both domains, understanding tail behavior aids in the management and mitigation of risks

associated with rare but significant events, resulting in improved preparation, financial stability, and resource management.

In this work, we shall use the Burr G (B-G) family to derive a new version of the C distribution. The CDF of the B-G family is defined as

$$F_{\theta_1,\underline{\boldsymbol{\varpi}}}(y) = \left\{1-\exp\left[-\boldsymbol{W}^2_{\underline{\boldsymbol{\varpi}}}(y)\right]\right\}^{\theta_1} | y \in \mathbf{R}, \tag{2}$$

where

$$\boldsymbol{W}^2_{\underline{\boldsymbol{\varpi}}}(y) = \left(\frac{G_{\underline{\boldsymbol{\varpi}}}(y)}{\overline{G}_{\underline{\boldsymbol{\varpi}}}(y)}\right)^2 | y \in \mathbf{R},$$

and $\underline{\boldsymbol{\varpi}}$ refers to the parameter version of the baseline model. The function $G_{\underline{\boldsymbol{\varpi}}}(y)$ refers to the CDF of the baseline model. However, $\overline{G}_{\underline{\boldsymbol{\varpi}}}(y)=1-G_{\underline{\boldsymbol{\varpi}}}(y)$ the function refers to the survival function corresponding to the baseline model.

By inserting (1) into (2), the CDF of the BGC distribution can be expressed as

$$F_{\underline{\Xi}}(y) = \left\{1-\exp\left[-\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)\right]\right\}^{\theta_1} | y>0,\ \theta_1,\theta_2,\theta_3>0\,, \tag{3}$$

where $\underline{\Xi}=(\theta_1,\theta_2,\theta_3)$,

$$\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y) = \left[\left(1-\boldsymbol{\Phi}_{\theta_3}(y)\right)^{-\theta_2}-1\right]^{-2},$$

and

$$\boldsymbol{\Phi}_{\theta_3}(y) = \exp\left\{\left[1-\exp\left(y^{\theta_3{}^{-1}}\right)\right]\right\}.$$

The corresponding PDF of the BGC can be derived as

$$f_{\underline{\Xi}}(y) = 2\theta_1\theta_2\theta_3{}^{-1}y^{\theta_3{}^{-1}-1}\exp\left(y^{\theta_3{}^{-1}}\right)\boldsymbol{\Phi}_{\theta_3}(y)$$

$$\times\frac{\exp\left[-\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)\right]\left[1-\boldsymbol{\Phi}_{\theta_3}(y)\right]^{2\theta_2-1}}{\left\{1-\left[1-\boldsymbol{\Phi}_{\theta_3}(y)\right]^{\theta_2}\right\}^3\left\{1-\exp\left[-\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)\right]\right\}^{1-\theta_1}} | y>0,\ \theta_1,\theta_2,\theta_3>0. \tag{4}$$

For large values of y, the term $y^{\theta_3{}^{-1}}$ grows, making $\exp\left(y^{\theta_3{}^{-1}}\right)$ dominant. Thus, we can approximate:

$$f_{\underline{\Xi}}(y)\sim\frac{2\theta_1\theta_2\theta_3{}^{-1}y^{\theta_3{}^{-1}-1}\exp\left(y^{\theta_3{}^{-1}}\right)}{\left\{1-\exp\left[-\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)\right]\right\}^{1-\theta_1}}, \tag{5}$$

since $\boldsymbol{\Phi}_{\theta_3}(y)\to 1$ as $y\to\infty$.

As $y\to 0$, the behavior of the PDF and CDF can be studied by analyzing the small y approximations of $\boldsymbol{\Phi}_{\theta_3}(y)$ and $\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)$. For small y, $\boldsymbol{\Phi}_{\theta_3}(y)$ behaves approximately linearly, and $\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)$ can be approximated. The asymptotic PDF behaves as

$$f_{\underline{\Xi}}(y)\approx 2\theta_1\theta_2\theta_3{}^{-1}\frac{y^{\theta_3{}^{-1}-1}}{\left\{1-\left[1-\boldsymbol{\Phi}_{\theta_3}(y)\right]^{\theta_2}\right\}^3\left\{1-\exp\left[-\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y)\right]\right\}^{1-\theta_1}}. \tag{6}$$

As

$$\boldsymbol{\Phi}_{\theta_3}(y) \approx y^{\theta_3^{-1}},$$

and

$$\boldsymbol{W}^{-2}_{\theta_2,\theta_3}(y) \approx \left[\frac{1-\boldsymbol{\Phi}_{\theta_3}(y)}{\theta_2}\right]^{-2},$$

the PDF can be approximated as

$$f_{\Xi}(y) \approx 2\theta_1\theta_2\theta_3^{-1}\frac{y^{\theta_3^{-1}-1}}{\left\{1-\left[1-\boldsymbol{\Phi}_{\theta_3}(y)\right]^{\theta_2}\right\}^3}. \tag{7}$$

Finally, the CDF $F_{\Xi}(y)$ approaches 0 as $y \to 0$. For small y, using the approximation $\boldsymbol{\Phi}_{\theta_3}(y) \approx y^{\theta_3^{-1}}$ we note that:

$$F_{\Xi}(y) = \left[1-\boldsymbol{\Phi}_{\theta_3}(y)\right]^{\theta_2}. \tag{8}$$

Since the tail behavior is critical for risk management and financial modeling, then the tail of the new model in (5), (6), (7), and (8) can be approximated by analyzing the large y behavior of the PDF. These approximation results will provide a simple expression for extreme values spotting and analysis and will support risk assessment for extreme scenarios as illustrated below. With an emphasis on both the peaks and the tail features, Figure 1 below shows how the new model behaves under different parameter values. The model shows a variety of peak shapes, ranging from very narrow and pointed to broad and flat, as can be seen in the image. The particular values used for the parameters determine this variance in peak width. Furthermore, the model's varied tail behavior demonstrates that tails can be heavy and take on various forms. This indicates that depending on the parameter settings, the model's tails might stretch more or less severely and have different weights.

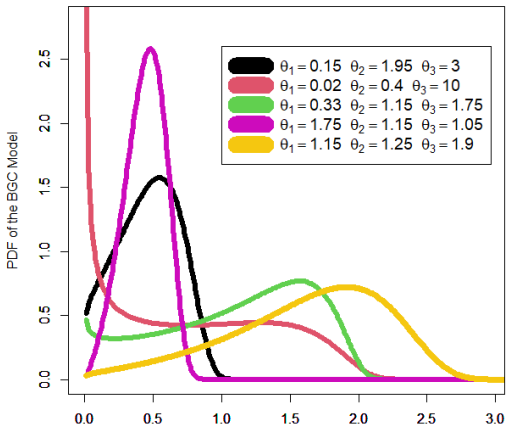


Figure 1: The peaks and the tail behavior of the new model under different parameters values.

Numerous significant improvements have occurred recently in probability distribution modelling and risk analysis. The Burr X Pareto distribution was introduced by Korkmaz et al. (2018) for VAR estimation in financial risk. In order to improve forecast accuracy, Rasekhi et al. (2022) presented a new G family of distributions. A compound-Lomax model for negatively skewed insurance claims was created by Hamed et al. in 2022. Shrahili and colleagues (2021) concentrated on asymmetric density functions for data of claim sizes. Models for extreme values were supplied by Yousof et al. (2023b) and Ibrahim et al. (2023), who also compared Bayesian techniques. Elbatal et al. (2024) investigated novel models for precipitation data, while Hashempour et al. (2023) investigated models for loss probability. For extreme stress data, Alizadeh et al. (2024) investigated the extended Gompertz model. The motivation for developing the new distribution stems from the growing demand for accurate risk quantification in critical

applications such as dam safety, ecological flow management, and financial risk modeling. By offering improved tail behavior modeling and greater adaptability to asymmetric data structures, the new model provides a powerful framework for enhancing predictive accuracy and supporting informed decision-making under uncertainty.

2. Main statistical Properties

2.1 Simple expansions

First, consider the below power series

$$\left(1-\frac{\ell_1}{\ell_2}\right)^{\ell_3}=\sum_{d_1=0}^{+\infty}\frac{\Gamma(1+\ell_3)}{d_1!\,\Gamma(1+\ell_3-d_1)}(-1)^{d_1}\,\left(\frac{\ell_1}{\ell_2}\right)^{d_1}\,|\left|\frac{\ell_1}{\ell_2}\right|<1,\ \ \upsilon_3>0. \tag{9}$$

Applying (9) to the quantity $A(\mathcal{y})$ in equation (4), where

$$A(\mathcal{y})=\left\{1-\exp\left[-\boldsymbol{\mathcal{W}}^{-2}_{\theta_2,\theta_3}(\mathcal{y})\right]\right\}^{\theta_1-1}=\sum_{d_1=0}^{+\infty}\frac{(-1)^{d_1}\,\Gamma\left((\theta_1-1)+1\right)}{d_1!\,\Gamma\left((\theta_1-1)+1-d_1\right)},$$

Then, inserting this expansion in (4), we have

$$\begin{aligned}f_{\Xi}(\mathcal{y})=2\theta_1\theta_2{\theta_3}^{-1}\mathcal{y}^{{\theta_3}^{-1}-1}\frac{\exp\left(\mathcal{y}^{{\theta_3}^{-1}}\right)\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})}{\left(1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right)^{1-\theta_2}}\frac{\left(1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right)^{\theta_2}}{\left\{1-\left[1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right]^{\theta_2}\right\}^3}\\ \times\sum_{d_1=0}^{+\infty}\frac{(-1)^{d_1}\,\Gamma(\theta_1)}{d_1!\,\Gamma(\theta_1-d_1)}\exp\left[-(d_1+1)\,\boldsymbol{\mathcal{W}}^{-2}_{\theta_2,\theta_3}(\mathcal{y})\right].\end{aligned} \tag{10}$$

Then, consider the well-known exponential power series and apply it to $A(\theta_2,\theta_3)$ where

$$A(\theta_2,\theta_3)=\exp\left\{-(d_1+1)\,\boldsymbol{\mathcal{W}}^{-2}_{\theta_2,\theta_3}(\mathcal{y})\right\},$$

equation (10) becomes

$$\begin{aligned}f_{\Xi}(\mathcal{y})=\ 2\theta_1\theta_2{\theta_3}^{-1}\mathcal{y}^{{\theta_3}^{-1}-1}\frac{\exp\left(\mathcal{y}^{{\theta_3}^{-1}}\right)\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})}{\left(1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right)^{1-\theta_2}}\\ \times\sum_{d_1,d_2=0}^{+\infty}\frac{(-1)^{d_1+d_2}\,(d_1+1)^{d_2}\Gamma(\theta_1)}{d_1!\ d_2!\,\Gamma(\theta_1-d_1)}\ \frac{\left[1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right]^{(2d_2+1)\,\theta_2}}{\left\{1-\left[1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right]^{\theta_2}\right\}^{2d_2+3}}.\end{aligned} \tag{11}$$

Consider the series expansion

$$\left(1-\frac{\upsilon_1}{\upsilon_2}\right)^{-\upsilon_3}=\sum_{d_3=0}^{+\infty}\frac{\Gamma(\upsilon_3+d_3)}{d_3!\,\Gamma(\upsilon_3)}\left(\frac{\upsilon_1}{\upsilon_2}\right)^{d_3}\,|\left|\frac{\upsilon_1}{\upsilon_2}\right|<1,\ \ \upsilon_3>0. \tag{12}$$

Using the expansion in (12) for the term $B(\mathcal{y})=\left[1-\left(1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right)^{\theta_2}\right]^{-2d_2-3}$, where

$$B(\mathcal{y})=\sum_{d_3=0}^{+\infty}\frac{\Gamma(2d_2+d_3+3)}{d_3!\,\Gamma(2d_2+3)}\left[\left(1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right)^{\theta_2}\right]^{d_3}.$$

Then, equation (7) becomes

$$f_{\underline{\Xi}}(\mathcal{y}) = \sum_{d_2,d_3=0}^{+\infty} \overbrace{2\theta_1(-1)^{d_2}\frac{\Gamma(\theta_1)\,\Gamma(2d_2+d_3+3)}{d_2!\,d_3!\,\Gamma(2d_2+3)\,\theta_2^{\cdot}} \sum_{d_1=0}^{+\infty}\frac{(-1)^{d_1}\,(d_1+1)^{d_2}}{d_1!\;\Gamma(\theta_1-d_1)}}^{\varsigma_{d_2,d_3}} \underbrace{a g_{\theta_3}(\mathcal{y})\left[G_{\theta_3}(\mathcal{y})\right]^{a-1}}_{\pi_{\boldsymbol{\nabla}}(\mathcal{y})}.$$

Then, equation (7) can be simplified as

$$f_{\underline{\Xi}}(\mathcal{y}) = \sum_{d_2,d_3=0}^{+\infty} \varsigma_{d_2,d_3}\;\pi_{\boldsymbol{\nabla}}(\mathcal{y})|(\boldsymbol{\nabla} = (2d_2+1)\,\theta_2 + d_3 + 1), \tag{13}$$

where

$$\varsigma_{d_2,d_3} = 2\theta_1(-1)^{d_2}\frac{\Gamma(\theta_1)\,\Gamma(2d_2+d_3+3)}{d_2!\,d_3!\,\Gamma(2d_2+3)\,\theta_2^{\cdot}} \sum_{d_1=0}^{+\infty}\frac{(-1)^{d_1}\,(d_1+1)^{d_2}}{d_1!\;\Gamma(\theta_1-d_1)},$$

and

$$\pi_a(\mathcal{y}) = a g_{\theta_3}(\mathcal{y})\left[G_{\theta_3}(\mathcal{y})\right]^{a-1}|a>0.$$

Similarly,

$$F_{\underline{\Xi}}(\mathcal{y}) = \sum_{d_2,d_3=0}^{+\infty} \varsigma_{d_2,d_3}\;\Pi_{\boldsymbol{\nabla}}(\mathcal{y})|\boldsymbol{\nabla} = (2d_2+1)\,\theta_2 + d_3 + 1, \tag{14}$$

where $\Pi_{\boldsymbol{\nabla}}(\mathcal{y}) = \left[G_{\theta_3}(\mathcal{y})\right]^{\boldsymbol{\nabla}}$ refers to the corresponding CDF of the baseline GZC model with power parameter $\boldsymbol{\nabla} > 0$.

2.2 Two theorems

We now have a better knowledge of the properties and uses of the GZC distribution because of the important contributions made by Dey et al. (2017). Their research provided insights into the behavior and applications of the GZC distribution in several settings, while also expanding upon its theoretical foundations.

Theorem 1:

Let $\boldsymbol{\mathcal{Y}}$ be a loss RV with the GZC distribution. Then using the transformation $t = \left[G_{\theta_2,\theta_3}(\mathcal{y})\right]^{\frac{1}{\theta_2}}$, the r^{th} raw moment of $\boldsymbol{\mathcal{Y}}$ can be expressed as

$$\mu'_r = \mathbf{E}[\boldsymbol{\mathcal{Y}}^r] = \theta_2{\theta_3}^{-1}\sum_{\ell,m=0}^{+\infty}\theta_2(\ell;\,r\theta_3)\;\theta_2(m;\,r\theta_3+\ell)\;\frac{(-1)^{2r\theta_3+\ell}}{\left[{\theta_3}^{-1}(\theta_2+\ell+m)+r\right]},$$

where $\theta_2(\ell;\,r\theta_3)$ is the mathematical coefficient of $[\log(1-t)]^{\frac{2r}{\theta_3}+\ell}$ in the expansion of $\left\{\sum_{d_1=1}^{+\infty}\frac{1}{d_1}[\log(1-t)]\right\}^{\frac{r}{\theta_3}}$ and $\theta_2(m;\,r\theta_3+\ell)$ is the coefficient of $t^{\ell+m+r\theta_3}$ in the expansion of $\left(\sum_{d_2=1}^{+\infty}\frac{t^{d_2}}{d_2}\right)^{r\theta_3+\ell}$ (see Balakrishnan and Cohen (2014) and Dey et al. (2017) for more details).

Theorem 2:

Let $\boldsymbol{\mathcal{Y}}$ refer to a loss RV having the GZC distribution. Then, the r^{th} conditional moment can be derived as

$$\mathbf{E}(\boldsymbol{\mathcal{Y}}^r|\mathcal{y}>\mathcal{y}) = \theta_2{\theta_3}^{-1}\sum_{\ell,m=0}^{+\infty}\theta_2(\ell;\,r\theta_3)\;\theta_2(m;\,r\theta_3+\ell)\;\frac{(-1)^{2r\theta_3+\ell}\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})}{\left[{\theta_3}^{-1}(\theta_2+\ell+m)+r\right]\left\{1-\left[1-\boldsymbol{\Phi}_{\theta_3}(\mathcal{y})\right]^{\theta_2}\right\}}.$$

2.3 The Ordinary moments

Based on Theorem 1, the r^{th} ordinary moment are

$$\mu'_{r,\boldsymbol{y}} = E[\boldsymbol{y}^r] = \boldsymbol{\nabla}{\theta_3}^{-1} \sum_{d_2,d_3,\ell,m=0}^{+\infty} \varsigma_{d_2,d_3}\, \boldsymbol{\nabla}_{\ell}(r\theta_3)\boldsymbol{\nabla}(m; r\theta_3+\ell)\, \frac{(-1)^{2r\theta_3+\ell}}{\left[{\theta_3}^{-1}(\boldsymbol{\nabla}+\ell+m)+r\right]}. \tag{15}$$

Then,

$$\mu'_r = \mathbf{E}[\boldsymbol{y}] = \theta_2{\theta_3}^{-1} \sum_{\ell,m=0}^{+\infty} \theta_2(\ell;\, \theta_3)\; \theta_2(m;\, \theta_3+\ell)\; \frac{(-1)^{2\theta_3+\ell}}{\left[{\theta_3}^{-1}(\theta_2+\ell+m)+1\right]},$$

$$\mu'_r = \mathbf{E}[\boldsymbol{y}^2] = \theta_2{\theta_3}^{-1} \sum_{\ell,m=0}^{+\infty} \theta_2(\ell;\, 2\theta_3)\; \theta_2(m;\, 2\theta_3+\ell)\; \frac{(-1)^{4\theta_3+\ell}}{\left[{\theta_3}^{-1}(\theta_2+\ell+m)+2\right]},$$

$$\mu'_r = \mathbf{E}[\boldsymbol{y}^3] = \theta_2{\theta_3}^{-1} \sum_{\ell,m=0}^{+\infty} \theta_2(\ell;\, 3\theta_3)\; \theta_2(m;\, 3\theta_3+\ell)\; \frac{(-1)^{6\theta_3+\ell}}{\left[{\theta_3}^{-1}(\theta_2+\ell+m)+3\right]},$$

and

$$\mu'_r = \mathbf{E}[\boldsymbol{y}^4] = \theta_2{\theta_3}^{-1} \sum_{\ell,m=0}^{+\infty} \theta_2(\ell;\, 4\theta_3)\; \theta_2(m;\, 4\theta_3+\ell)\; \frac{(-1)^{8\theta_3+\ell}}{\left[{\theta_3}^{-1}(\theta_2+\ell+m)+4\right]}.$$

2.4 The Conditional moments

Based on the results of Theorem 2, the conditional moments can be derived as follows

$$\mathbf{E}(\boldsymbol{y}|\boldsymbol{y}>\boldsymbol{y}) = \boldsymbol{\nabla}{\theta_3}^{-1} \sum_{d_2,\,d_3,\,\ell,m=0}^{+\infty} \varsigma_{d_2,\,d_3}\, \frac{\boldsymbol{\nabla}(\ell;\,\theta_3)\,\boldsymbol{\nabla}(m;\,\theta_3+\ell)\,(-1)^{2\theta_3+\ell}\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})}{\left[{\theta_3}^{-1}(\boldsymbol{\nabla}+\ell+m)+1\right]\left[1-\left(1-\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})\right)^{\boldsymbol{\nabla}}\right]},$$

$$\mathbf{E}(\boldsymbol{y}^2|\boldsymbol{y}>\boldsymbol{y}) = \boldsymbol{\nabla}{\theta_3}^{-1} \sum_{d_2,\,d_3,\,\ell,m=0}^{+\infty} \varsigma_{d_2,\,d_3}\, \frac{\boldsymbol{\nabla}(\ell;\,2\theta_3)\,\boldsymbol{\nabla}(m;\,2\theta_3+\ell)\,(-1)^{4\theta_3+\ell}\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})}{\left[{\theta_3}^{-1}(\boldsymbol{\nabla}+\ell+m)+2\right]\left[1-\left(1-\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})\right)^{\boldsymbol{\nabla}}\right]},$$

$$\mathbf{E}(\boldsymbol{y}^3|\boldsymbol{y}>\boldsymbol{y}) = \boldsymbol{\nabla}{\theta_3}^{-1} \sum_{d_2,\,d_3,\,\ell,m=0}^{+\infty} \varsigma_{d_2,\,d_3}\, \frac{\boldsymbol{\nabla}(\ell;\,3\theta_3)\,\boldsymbol{\nabla}(m;\,3\theta_3+\ell)\,(-1)^{6\theta_3+\ell}\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})}{\left[{\theta_3}^{-1}(\boldsymbol{\nabla}+\ell+m)+3\right]\left[1-\left(1-\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})\right)^{\boldsymbol{\nabla}}\right]},$$

and

$$\mathbf{E}(\boldsymbol{y}^4|\boldsymbol{y}>\boldsymbol{y}) = \boldsymbol{\nabla}{\theta_3}^{-1} \sum_{d_2,\,d_3,\,\ell,m=0}^{+\infty} \varsigma_{d_2,\,d_3}\, \frac{\boldsymbol{\nabla}(\ell;\,4\theta_3)\,\boldsymbol{\nabla}(m;\,4\theta_3+\ell)\,(-1)^{8\theta_3+\ell}\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})}{\left[{\theta_3}^{-1}(\boldsymbol{\nabla}+\ell+m)+4\right]\left[1-\left(1-\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})\right)^{\boldsymbol{\nabla}}\right]}.$$

2.5 The mean residual life

The mean residual life (MRL) can be derived using the formula:

$$M_{1,\,\boldsymbol{y}} = \mathbf{E}(\boldsymbol{y}-\boldsymbol{y}|\boldsymbol{y}>\boldsymbol{y}) = \frac{1}{1-F_{\Xi}(\boldsymbol{y})}\left[\int_{\boldsymbol{y}}^{+\infty} \boldsymbol{y} f_{\Xi}(\boldsymbol{y})\, d\boldsymbol{y}\right] - \boldsymbol{y},$$

Then,

$$M_{1,\,\boldsymbol{y}} = \boldsymbol{\nabla}\theta_3 \sum_{d_2,\,d_3,\,\ell,m=0}^{+\infty} \varsigma_{d_2,\,d_3}\, \frac{\varsigma_{d_2,\,d_3}\,\boldsymbol{\nabla}(\ell;\,\theta_3)\,\boldsymbol{\nabla}(m;\,\theta_3+\ell)\,(-1)^{2\theta_3+\ell}\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})}{\left[{\theta_3}^{-1}(\theta_2^{\cdot}+\ell+m)+1\right]\left[1-\left(1-\boldsymbol{\Phi}_{\theta_3}(\boldsymbol{y})\right)^{\boldsymbol{\nabla}}\right]} - \boldsymbol{y}.$$

The MRL is a fundamental concept in reliability and risk analysis, reflecting the expected time remaining for a component or system before failure occurs. In reliability analysis, the MRL provides vital information about the

resilience of systems or components, helping to determine the feasibility of maintenance and replacement planning. In risk analysis, it can help assess the probabilities of failure occurring over a given period, enabling companies to take appropriate preventive measures to mitigate potential risks.

3. Main risk indicators under the BGC model

3.1 The VAR indicator

The VAR of the BGC distribution at the $100q\%$ level, say VAR($\boldsymbol{y}$) or $\pi(q)$, is the $100q\%$ quantile (or percentile). Then, we can simply write.

$$\text{VAR}\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)=\Pr\left(\boldsymbol{y}>Q(U)\right)=\begin{cases}1\%|_{q=99\%}\\5\%|_{q=95\%}\\\vdots\end{cases}, \tag{16}$$

where

$$Q(U)=F_{\underset{\sim}{\Xi}}^{-1}(y)=\left(1-\ln\left\{1-\left[1+\sqrt{-\ln\left(1-q^{\theta_1{}^{-1}}\right)}\right]^{\theta_2{}^{-1}}\right\}\right)^{\theta_3}.$$

3.2 The TVAR indicator

The TVAR, or conditional value-at-risk (CVAR), enhances the VAR indicator by considering the average loss that exceeds the VAR threshold. Unlike VAR, which focuses solely on losses at a specific confidence level, TVAR evaluates the average loss in the tail of the distribution. This provides deeper insights into extreme losses and tail risk. TVAR is extensively used in risk measurement, risk budgeting, and portfolio optimization to assess and manage downside risk effectively. The TVAR of $\boldsymbol{y}$ at the $100q\%$ confidence level is the expected loss given that the loss exceeds the $100q\%$ of the distribution of $\boldsymbol{y}$, then the TVAR$\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)$ of $\boldsymbol{y}$ can be expressed as

$$\text{TVAR}\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)=\mathbf{E}\left(\boldsymbol{y}|\boldsymbol{y}>\pi(q)\right),$$

which can be derived form

$$\text{TVAR}\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)=\frac{1}{1-F_{\underset{\sim}{\Xi}}\left(\pi(q)\right)}\int_{\pi(q)}^{\infty}y\ f_{\underset{\sim}{\Xi}}(y)\,dy.$$

Then,

$$\text{TVAR}\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)=\frac{1}{1-q}\int_{\pi(q)}^{\infty}y\,f_{\underset{\sim}{\Xi}}(y)dy. \tag{17}$$

Then,

$$\text{TVAR}\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)=\frac{\boldsymbol{\nabla}}{1-q}\theta_3{}^{-1}\sum_{d_2,\,d_3,\,\ell,\mathrm{u}=0}^{+\infty}\varsigma_{d_2,d_3}\ C_1(\ell,\mathrm{u})\ \frac{\varsigma_{d_2,d_3}\,(-1)^{2\theta_3+\ell}}{\left[\theta_3{}^{-1}(\boldsymbol{\nabla}+\ell+m)+1\right]}\boldsymbol{\Phi}_{\theta_3}(q;\boldsymbol{\nabla}),$$

where

$$\boldsymbol{\Phi}_{\theta_3}(q;\boldsymbol{\nabla})=\boldsymbol{\Phi}_{\theta_3}(q)\frac{1}{1-\left(1-\boldsymbol{\Phi}_{\theta_3}(q)\right)^{\boldsymbol{\nabla}}},$$

and

$$C_1(\ell,\mathrm{u})=\boldsymbol{\nabla}(\ell;\,\theta_3)\,\boldsymbol{\nabla}(m;\,\theta_3+\ell).$$

The quantity TVAR$\left(\boldsymbol{y};\underset{\sim}{\Xi}\right)$, which gives further details about the tail of the BGC distribution, is, therefore, the average of all the VAR values mentioned above at the confidence level q. The term $\boldsymbol{\Phi}_{\theta_3}(q;\boldsymbol{\nabla})$ can be interpreted as an adjusted version of the CDF, where the adjustment is made based on the parameter $\boldsymbol{\nabla}$. This adjustment is crucial for understanding how extreme values (tail behavior) are treated in the model. In risk analysis, especially for extreme events or losses, it's essential to understand the tail behavior of the distribution. The provided formula helps in analyzing how extreme quantiles (e.g., higher percentiles) are affected by the parameter $\boldsymbol{\nabla}$. This can provide a more accurate assessment of the likelihood and impact of rare but significant events.

3.3 The MEXL indicator

The TVAR$\left(\boldsymbol{y};\underline{\Xi}\right)$ can also be expressed as

$$\text{TVAR}\left(\boldsymbol{y};\underline{\Xi}\right) = e\left(\boldsymbol{y};\underline{\Xi},q\right) + \text{VAR}\left(\boldsymbol{y};\underline{\Xi}\right),$$

where $e\left(\boldsymbol{y};\underline{\Xi},q\right)$ is the MEXL function evaluated at the $100q\%^{th}$ quantile (see Wirch (199), Tasche (2002), Furman and Landsman (2006), and Acerbi (2002)). When the $e\left(\boldsymbol{y};\underline{\Xi},q\right)$ value vanishes, then $\text{TVAR}\left(\boldsymbol{y};\underline{\Xi}\right) = \text{VAR}\left(\boldsymbol{y};\underline{\Xi}\right)$ and for the very small values of $e\left(\boldsymbol{y};\underline{\Xi},q\right)$, the value of $\text{TVAR}\left(\boldsymbol{y};\underline{\Xi}\right)$ will be very close to $\text{VAR}\left(\boldsymbol{y};\underline{\Xi}\right)$. This approximation highlights that TVAR and VaR are closely related to distributions with mild tail behavior. In financial and insurance contexts, knowing that TVAR converges to VAR when $e\left(\boldsymbol{y};\underline{\Xi},q\right)$ is small can simplify risk management practices. This implies that for light-tailed distributions, focusing on VaR might be sufficient, whereas TVAR provides additional information on the expected magnitude of losses beyond VaR.

3.4 The TV indicator

The TV$\left(\boldsymbol{y};\underline{\Xi}\right)$ risk indicator, which Furman and Landsman (2006) developed, calculates the loss's deviation from the average along a tail. Explicit expressions for the TV risk indicator under the multivariate normal distribution were also developed by Furman and Landsman (2006). The TV risk indicator $\text{TV}\left(\boldsymbol{y};\underline{\Xi}\right)$ can then be expressed as

$$\text{TV}\left(\boldsymbol{y};\underline{\Xi}\right) = \mathbf{E}(\boldsymbol{y}^2|\boldsymbol{y} > \pi(q)) - \left[\text{TVAR}\left(\boldsymbol{y};\underline{\Xi}\right)\right]^2, \tag{18}$$

where

$$E(\boldsymbol{y}^2|\boldsymbol{y} > \pi(q)) = \boldsymbol{\nabla}\theta_3 \sum_{d_2,d_3,\ell,m=0}^{+\infty} C_2(\ell,m)\frac{\varsigma_{d_2,d_3}(-1)^{4\theta_3+\ell}}{\left[\theta_3{}^{-1}(\theta_2^{\cdot}+\ell+m)+2\right]}\boldsymbol{\Phi}_{\theta_3}(q;\boldsymbol{\nabla}),$$

and

$$C_2(\ell,m) = \boldsymbol{\nabla}_\ell(2\theta_3)\boldsymbol{\nabla}_m(2\theta_3+\ell).$$

3.5 The TMV indicator

As a statistic for the best portfolio choice, Landsman (2010) developed the TMV risk indicator, which is based on the TV risk indicator. Consequently, the TMV risk indicator may be written as

$$\text{TMV}\left(\boldsymbol{y};\underline{\Xi}\right) = \text{TVAR}\left(\boldsymbol{y};\underline{\Xi}\right) + \pi\text{TV}\left(\boldsymbol{y};\underline{\Xi}\right)|0 < \pi < 1. \tag{19}$$

Then, for any continuous RV, $\text{TMV}\left(\boldsymbol{y};\underline{\Xi}\right) > \text{TV}\left(\boldsymbol{y};\underline{\Xi}\right)$ and, for $\pi = 1$, $\text{TMV}\left(\boldsymbol{y};\underline{\Xi}\right) = \text{TVAR}\left(\boldsymbol{y};\underline{\Xi}\right)$.

3.6 The MOO$^{\text{P}}$ analysis

Due to Figueiredo et al. (2017) and Alizadeh et al. (2024), the MOO$^{\text{P}}$ indicator for risk analysis can be defined as the mean (average) of the P$^{\text{th}}$ order statistics from a certain sample of claims or any other data. For a given dataset $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$, the MOO$^{\text{P}}$ indicator is calculated as follows

$$\text{MOO}^{\text{P}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i^P, \tag{20}$$

The problem optimization can be solved depending on some statistical criteria by the following formula:

$$\text{Opt} - \text{P} = \arg\left\{\min_P[\text{Objective Function(P)}]\right\}. \tag{21}$$

Due to Alizadeh et al. (2024), the mean squared error, and Bias can be considered in optimization process.

Each risk indicator plays a unique and context-dependent role in assessing extreme event risks across hydrology, medicine, and finance. In hydrology, VaR is essential for identifying critical thresholds for low-flow events, particularly at high confidence levels (e.g., 99%), where it helps dam operators and water resource planners prepare

for prolonged droughts by setting operational limits and ensuring ecological sustainability. As confidence levels increase, the number of exceedances also rises, allowing for more robust modeling of rare but impactful flow conditions. TVaR, which measures the expected loss beyond the VaR threshold, becomes especially valuable in flood risk management and infrastructure planning, as it provides an average severity estimate of losses in the tail, supporting capital allocation and mitigation strategies. In medical applications, such as relief times data, VaR serves as a benchmark for understanding treatment durations at various confidence levels, while MEXL reveals how much longer-than-expected relief times extend beyond certain thresholds, directly informing hospital triage and resource planning. The TV highlights the dispersion of losses in the tail and is particularly useful in insurance and financial risk modeling, where volatility can significantly affect solvency and investment decisions. For example, in U.S. indemnity loss data, MEXL increases sharply at higher confidence levels, showing that the average loss beyond VaR becomes significantly larger than the threshold itself, underscoring the nonlinear growth of tail risk and justifying higher premium loadings and reinsurance participation. TMV, which combines both mean and variance of tail losses, offers a balanced perspective, making it a powerful tool in enterprise risk management where decision-makers need to consider both the magnitude and variability of extreme outcomes. Across all domains, higher confidence levels generally lead to more conservative estimates, emphasizing the need for dynamic threshold selection and model adaptability, especially when using advanced methods like PORT-VaR, to ensure accurate risk quantification. These indicators collectively enhance the understanding of extreme value behavior, enabling stakeholders, from dam engineers and healthcare providers to insurers and financial regulators, to make informed, data-driven decisions that balance operational efficiency, financial stability, and systemic resilience in the face of rare but high-impact events.

3.6 The PORT-VAR analysis

Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ be a sample of real data. Then, define the exceedances over the threshold T as

$$\boldsymbol{Z}_i = \boldsymbol{Y}_i - T | \boldsymbol{Y}_i > T.$$

The exceedances $\boldsymbol{Z}_i$ follow a probability distribution conditioned on $\boldsymbol{Y}_i > T$. The tail behavior of this model is important for analyzing extreme events. Often, exceedances values are modeled and analyzed using the generalized Pareto distribution, which describes the tail of the distribution. However, in this study, we will consider the BGC model in (3) and (4). The PORT-VAR analysis for the BGC distribution also involves deriving the VAR and TVAR metrics. (see Figueiredo et al. (2017), Mohamed et al. (2024) and Alizadeh et al. (2024)).

4. MOOP analysis and Opt-P

4.1 MOOP for the relief times

In this subsection, we will explore the MOOP approach applied to the dataset provided by Gross and Clark (1975). The specific relief times in the dataset are as follows: 1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3.0, 1.7, 2.3, 1.6, and 2.0. Table 1 below displays the results of the MOOP analysis applied to the relief times dataset, specifically for different values of the parameter P. The true mean value (TMV) of the Gross and Clark data and Bias are provided for assessing the MOOP indicator.

The P values in the range of 1 to 5 are analyzed. The table shows the computed MOOP value for each order P, with each row corresponding to a distinct order. The table offers a thorough overview of the evolution of the MOOP metric as P values increase. This progression facilitates comprehension of the effects of varying degrees of stress on longer release times on the overall central tendency score. By examining these values, we may learn more about the relief times' distributional properties, like skewness and dispersion, and evaluate how the central tendency changes depending on how much focus is placed on extreme values.

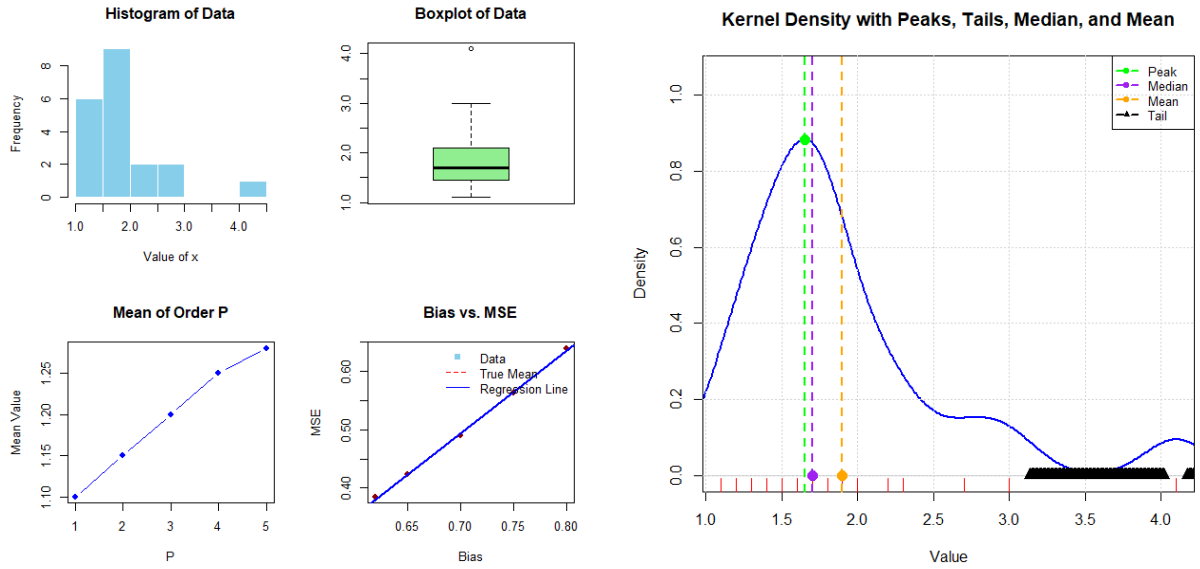
The relief times data's TMV of 1.9 is used as a standard by which to compare how well other MOOP estimations perform. An evolving estimate of the data's central tendency is given by the MOOP estimates for various values of P (1–5). The MOOP estimations indicate a gradual increase from 1.1 to 1.2800 as P increases. As a result of MOOP's ability to capture all facets of the data distribution, this pattern suggests that larger values of P correspond to higher estimates. As P rises, the MSE values drop from 0.64 to 0.3844. This implies that the estimation error decreases as P rises because the MOOP estimations approach closer to the actual mean value. Better estimation precision performance is shown by lower MSE values. As P increases, the Bias falls from 0.80 to 0.6200. This suggests that as P grows, the

MOO^P estimates become less biassed and converge closer to the actual mean value. A declining bias is indicative of improving estimations of accuracy, which is a good thing.

Consider utilizing larger values of P (e.g., 4 or 5) in medical applications where accuracy is critical, such as assessing treatment efficacy or setting dosage amounts. Lower MSE and Bias in the MOOP estimations for these variables suggest that the true mean of relief times or other relevant metrics are more accurately reflected. Greater P values result in more precise estimations, but they also add to the complexity of the process. Lastly, if our objective is to minimize both MSE and Bias, then P=5 is ideal based on the MSE and Bias values shown in Table 1. A graphic that describes the relief times is shown in Figure 2. The MOOP, MSE, and Bias for the relief times are displayed throughout the order of P in Figure 3.

Table 1: MOO^P analysis under the relief times data.

P	1	2	3	4	5
True mean value			1.9		
MOO^P	1.1	1.1500	1.20	1.2500	1.2800
MSE	0.64	0.5625	0.49	0.4225	0.3844
Bias	0.80	0.7500	0.70	0.6500	0.6200



(a) (b)

Figure 2: The Histogram, the box plot, the mean of order and the Bias vs. MSE (a) for the relief times; Kernel plot for the relief times (b).

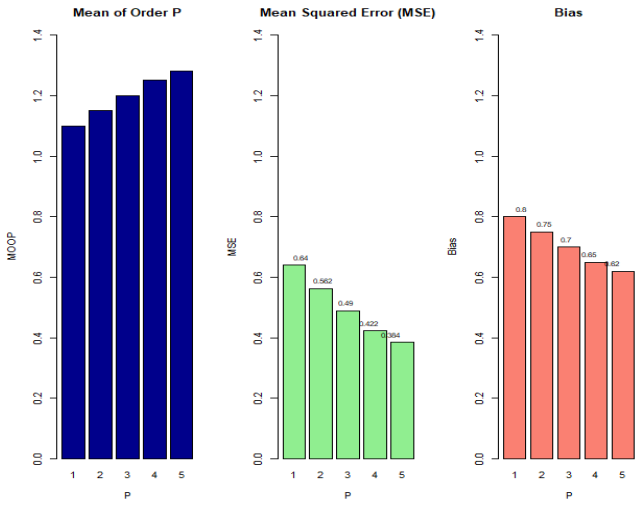


Figure 3: MOO^P, MSE and Bias across the order of P for the relief times.

4.2 MOO^P for the minimum flow

We concentrate on the MOO^P for the Cuiabá River's lower discharge values in this investigation. According to Andrade et al. (2007), the dataset used covers 38 years, from January 1962 to October 1999. In order to effectively plan and manage water resources, it is essential to analyze long-term trends and patterns in river discharge, and this historical data offers a strong foundation for doing so. The data set is the following: 43.86, 44.97, 46.27, 51.29, 61.19, 61.20, 67.80, 69.00, 71.84, 77.31,85.39, 86.59, 86.66, 88.16, 96.03, 102.00, 108.29, 113.00, 115.14, 116.71, 126.86, 127.00, 127.14, 127.29, 128.00, 134.14, 136.14, 140.43, 146.43, 146.43,148.00, 148.43, 150.86, 151.29, 151.43, 156.14, 163.00, 186.43 (see also, Cordeiro *et al.* (2012)). An examination of the MOOP based on the data from the minimum flow observations is shown in Table 2. It has multiple metrics for various P values, as well as estimations for MOOP, MSE, and bias.

Table 2: MOO^P analysis under the minimum flow observations.

P	1	2	3	4	5
TMV			110.2142		
MOO^P	43.86	44.415	45.03333	46.5975	49.516
MSE	4402.881	4329.536	4248.547	4047.086	3684.273
Bias	66.35421	65.79921	65.18088	63.61671	60.69821

Table 2 shows that MSE and Bias both drop with increasing P, suggesting that the MOOP estimations are now more accurate. It is desired for more accurate modelling if higher-order MOOP estimates can more closely resemble the true mean, as this trend implies. The true mean value (TMV = 110.2142) is consistently higher than the MOOP values, which could point to an underestimating issue. This can be because of the particulars of the dataset or the MOOP computations. Higher-order MOO^P estimates appear to be more successful at capturing the underlying mean value, as evidenced by the decline in MSE and Bias as P grows. This would provide a more up-to-date perspective on discharge patterns and improve the robustness of the estimates. Explore other hydrological indicators and metrics to gain a comprehensive understanding of discharge behaviors. This could include analyzing seasonal variations and long-term trends. Finally, based on the results of MSE and Bias provided in Table 2, P=5 is optimal if minimizing both MSE and Bias is our goal. Figure 4 presents a describing the minimum flow of the Cuiabá River. Figure 5 shows the MOO^P, MSE and Bias across the order of P for the minimum flow of the Cuiabá River.

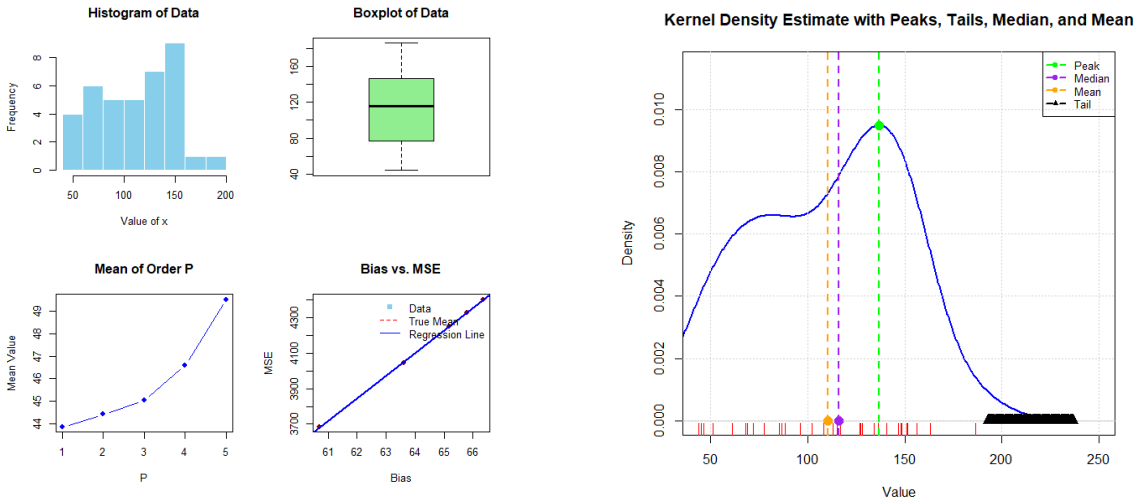


Figure 4: The Histogram, the box plot, the mean of order and the Bias vs. MSE (a) for the Cuiabá River data; Kernel plot for the relief times (b).

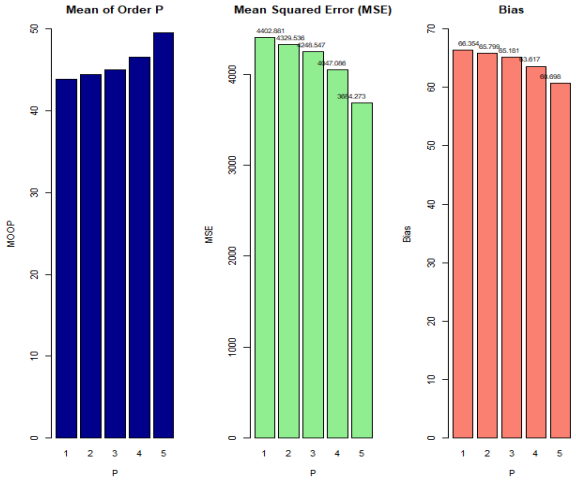


Figure 5: MOO^P, MSE, and Bias across the order of P for the minimum flow of the Cuiabá River.

5. PORT-VAR analysis

5.1 Extreme risks in relief times

Table 3 provides a comprehensive PORT-VAR analysis for the relief times data across different confidence levels (CL=50%, 70%, 90%, 99%). The analysis is summarized with key metrics including the VAR threshold, the number of peaks exceeding this threshold, and various statistical measures of these peaks.

Table 3: PORT-VAR analysis under the relief times data.

CL	VAR Threshold	N. Peaks	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
50%	1.70	9	1.800	1.900	2.200	2.422	2.700	4.100
70%	1.57	14	1.600	1.700	1.850	2.150	2.275	4.100
90%	1.29	18	1.300	1.600	1.750	1.983	2.150	4.100
99%	1.12	19	1.200	1.550	1.700	1.942	2.100	4.100

The VAR thresholds decrease as the confidence level increases. For instance, the VaR threshold is 1.7 at a 50% confidence level, and it drops to 1.12 at a 99% confidence level. This reflects that higher confidence levels correspond to lower thresholds, capturing more extreme values. The number of peaks exceeding the VaR threshold increases with higher confidence levels, ranging from 9 peaks at the 50% level to 19 peaks at the 99% level. This increase indicates that as we lower the threshold, more data points are identified as extreme peaks. The minimum (Min)values observed at each confidence level show that extremely low values become more apparent at higher confidence levels, with the minimum value ranging from 1.800 at 50% CL to 1.200 at 99% CL.

The 1st quartile (1st Qu.) values also decrease with higher confidence levels, showing a shift towards lower relief times as the threshold is reduced. The median values are relatively consistent but show a slight decrease as the confidence level increases, reflecting the central tendency of the extreme peaks. The mean values exhibit a similar trend to the median, with a slight decrease as the threshold becomes lower, indicating the average value of peaks decreases with higher confidence levels. The 3rd quartile (3rd Quartile) values also decrease, suggesting that the upper range of the peaks is lower when considering higher confidence levels. The Maximum (Max) values remain the same across all confidence levels, indicating that the most extreme peak observed is consistent regardless of the threshold.

Figure 6 provides the histogram of relief data with peaks above VAR & CL=50%, 70%, 90%, and 99%. Figure 6 effectively visualizes the distribution of relief data relative to different VAR thresholds and confidence levels. By highlighting the peaks that exceed these thresholds, the histogram offers insights into the frequency and magnitude of significant relief events. This analysis aids in understanding extreme value behavior, supporting data-driven decision-making in fields where relief data is a critical factor.

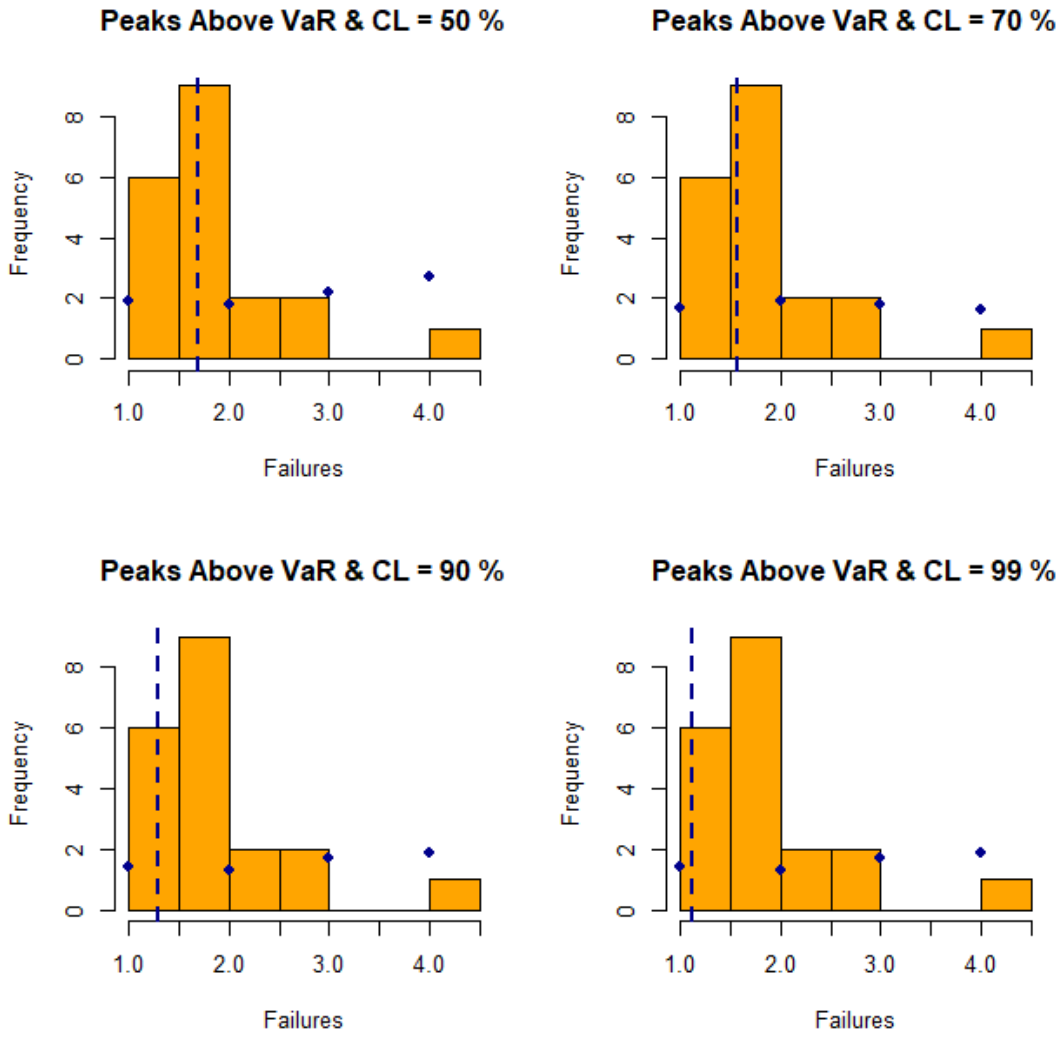


Figure 6: Histogram of relief data with peaks above VaR & CL=50%, 70%, 90%, and 99%, where the blue dashed line is the threashold and blue dots are the peaks.

5.2 Extreme low-flow events

For managing water resources, predicting extreme low-flow conditions is essential for planning water supply, especially in regions prone to droughts. Tail behavior analysis helps ensure adequate water availability during rare but critical periods. So, minimum flow values are essential for maintaining aquatic habitats. Rivers and streams support diverse species of fish, plants, and other wildlife. Adequate flow ensures that these habitats are not compromised, which is critical for the survival and reproduction of aquatic species. PORT-VAR can assess the risk of extreme low-flow events, which can threaten ecological balance. By analyzing these rare but critical low-flow situations, water managers can better understand and mitigate the risks to natural habitats and water supplies. PORT-VAR helps determine optimal flow thresholds for maintaining ecological health. This involves setting thresholds that, when exceeded, would signal a potential threat to ecosystems, prompting timely interventions. Table 4 below provides an analysis of the PORT-VaR for minimum flow data from the Cuiabá River. This table details the relationship between different CL and various statistical measures of peaks exceeding the VAR thresholds.

Table 4: PORT-VAR analysis under the minimum flow data of the Cuiabá River.

CL	VAR Threshold	N. Peaks	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
50%	115.92	19	116.7	127.6	146.4	142.7	151.1	186.43
70%	86.6	26	86.66	115.53	131.07	131.59	148.32	186.43
90%	58.22	34	61.19	87.03	126.93	117.70	146.43	186.43
99%	44.27	37	44.97	85.39	116.71	112.01	146.43	186.43

The VAR thresholds decrease as the confidence level increases. For example, the threshold is 115.92 at the 50% confidence level and drops to 44.27 at the 99% confidence level. This trend reflects that as we consider more extreme confidence levels, the threshold value becomes lower, encompassing more of the observed extreme values. The number of peaks exceeding the VaR threshold increases with higher confidence levels, ranging from 19 peaks at the 50% level to 37 peaks at the 99% level. This indicates that as the threshold decreases, more values from the dataset are classified as peaks, highlighting the presence of more extreme low-flow conditions. The minimum values for peaks increase as the confidence level decreases, showing that lower confidence levels capture a broader range of data points, including lower values. For instance, the minimum peak increases from 116.7 at the 50% CL to 44.97 at the 99% CL. The 1st quartile values also decrease with higher confidence levels, reflecting a shift towards lower values of the peaks. This suggests that more extreme lower flows are identified at higher confidence levels. The median value of the peaks decreases with higher confidence levels, indicating that the central value of the extreme peaks is lower as the threshold is reduced.

The minimum flow histogram, shown in Figure 7, has maxima above VAR & CL=50%,70%,99%, and 99%. The distribution of minimum flow values in relation to various VAR thresholds and confidence levels is effectively depicted in Figure 7. Stakeholders can manage water resources more effectively and mitigate potential dangers related to extreme low-flow conditions by looking at the histogram, which provides a better understanding of the frequency and extremes of low-flow episodes.

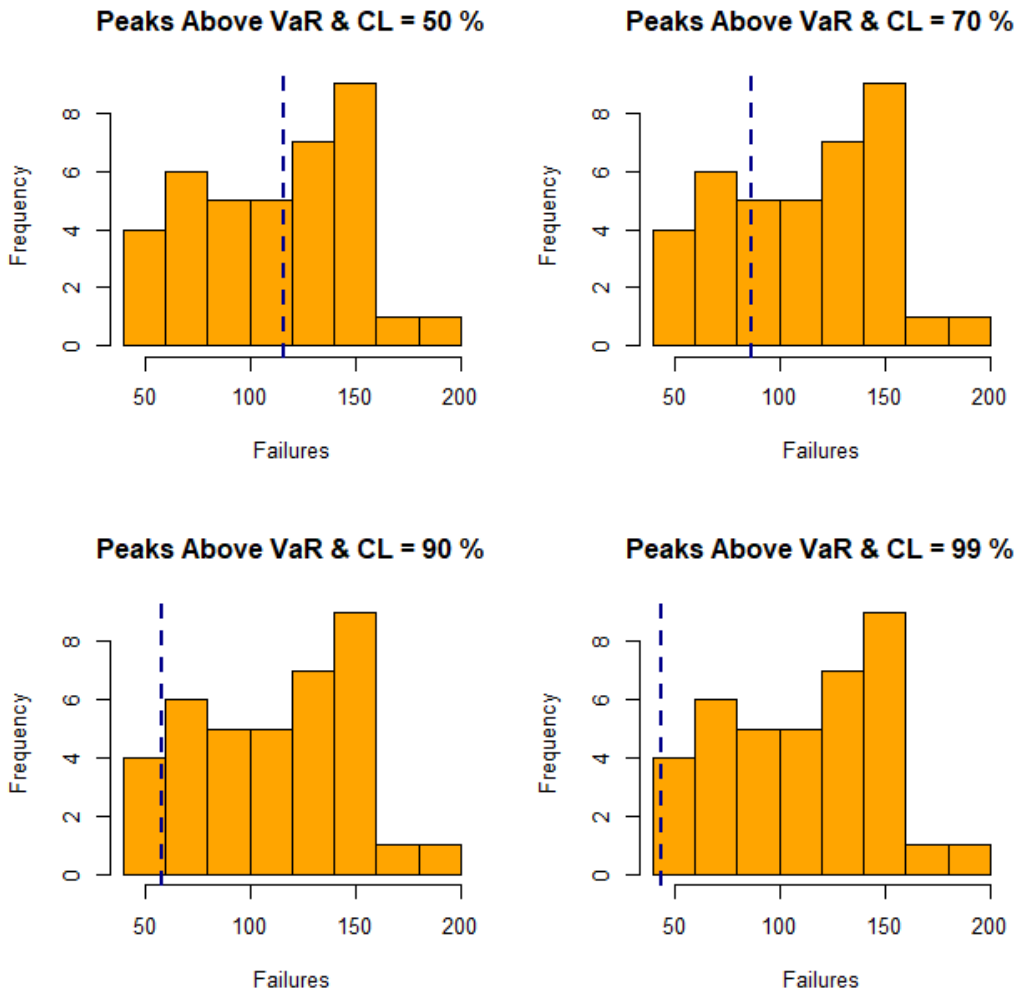


Figure 7: Histogram of minimum flow with peaks above VaR & CL=50%, 70%, 90% and 99% where the blue dashed line is the threshold and blue dots are the peaks.

From a hydrological modeling perspective, the findings of this study have direct implications for engineering decisions related to dam safety, drought risk management, ecological flow maintenance, and flood forecasting, particularly through the application of the BGC distribution and advanced analytical tools like MOOP and PORT-VaR. The MOOP analysis reveals that increasing the order P leads to reduced MSE and Bias, improving the accuracy of central tendency estimation for skewed and heavy-tailed datasets such as river flow data—this enables engineers to better model extreme low-flow conditions, refine reservoir release policies, and enhance long-term drought preparedness by setting more accurate operational thresholds. Using PORT-VaR, the study identifies how VaR increases with higher confidence levels, capturing more exceedances and allowing for robust modeling of extreme events; for instance, in the Cuiabá River dataset, VaR rises from 115.92 at 50% confidence to 157.169 at 99% , highlighting the growing exposure to extreme low-flow scenarios, which is crucial for setting minimum environmental flow requirements and ensuring dam safety under increasingly uncertain climatic conditions. These insights support adaptive water resource planning by enabling engineers to dynamically adjust thresholds based on real-time hydrological behavior, improve infrastructure resilience against prolonged droughts, and maintain aquatic ecosystem integrity through science-based ecological flow standards, all while accounting for the nonlinear growth of tail risk. Ultimately, the BGC model provides a flexible and precise framework for modeling extreme hydrological events, enhancing predictive accuracy and supporting informed, data-driven engineering decisions in the face of climate variability and increasing water scarcity.

6. Risk analysis and assessment

In the field of risk management, we shall employ a suite of risk metrics including VAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$, TVAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$, TV$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ and TMV$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$. By applying these measures to both relief data and minimum flow data, this chapter aims to highlight the varying risk profiles and the implications for dam safety and management. The findings will offer valuable insights into how risk levels evolve with increasing confidence and how different risk measures can be leveraged to improve risk management strategies in the context of hydrological extremes. Table 5 gives all risk analyses under relief times. Table 6 presents all risk analyses under minimum flow. Figure 8 gives the risk indicators across the CLs for relief times. Figure 9 shows the Kernel plots for the risk indicators across the CLs for relief times. Figure 10 provides the risk indicators across the CLs for minimum flow. Figure 11 presents the Kernel plots for the risk indicators across the CLs for minimum flow. Based on Table 5, we note that:

I. The VAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ increases with higher confidence levels (from 1.700 at 50% to 3.055 at 99%). This indicates that the potential risk, or the worst-case loss, grows as the confidence level increases.

II. TVAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ follows a similar trend to VAR, increasing from 0.722 to 1.045. It represents the expected loss, given that the loss exceeds the VAR threshold. The rise in TVAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ reflects higher extreme risks at higher confidence levels.

III. TV$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ also increases but drops to 0 at the 99% level. This could indicate that while losses are significant up to the 90% confidence level, they become less predictable or zero out in the extreme case.

IV. TMV$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ behaves similarly, increasing with confidence levels but dropping to zero at 99%, suggesting a high variability or extreme values that are not present at the highest confidence level.

Due to these main results, we have the following recommendations:

1. Ensure that the VAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ and TVAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ values are regularly updated and monitored, as they reflect potential risk exposure. Implement strategies to mitigate high-risk scenarios, especially at higher confidence levels.
2. Establish robust contingency plans and financial buffers for high-risk scenarios, especially as confidence levels increase.
3. Educate staff on risk management and emergency response procedures to handle situations indicated by high VAR and TVAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ values.

Table 5: Risk analysis under relief times.

CL	VAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$	TVAR$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$	TV$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$	TMV$\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$
50%	1.700	0.7222222	0.5644444	0.5017284
70%	1.930	0.7866667	0.5896667	0.4913889
90%	2.730	0.82	0.6050	0.3025
99%	3.055	1.045	0.0001	0.0003

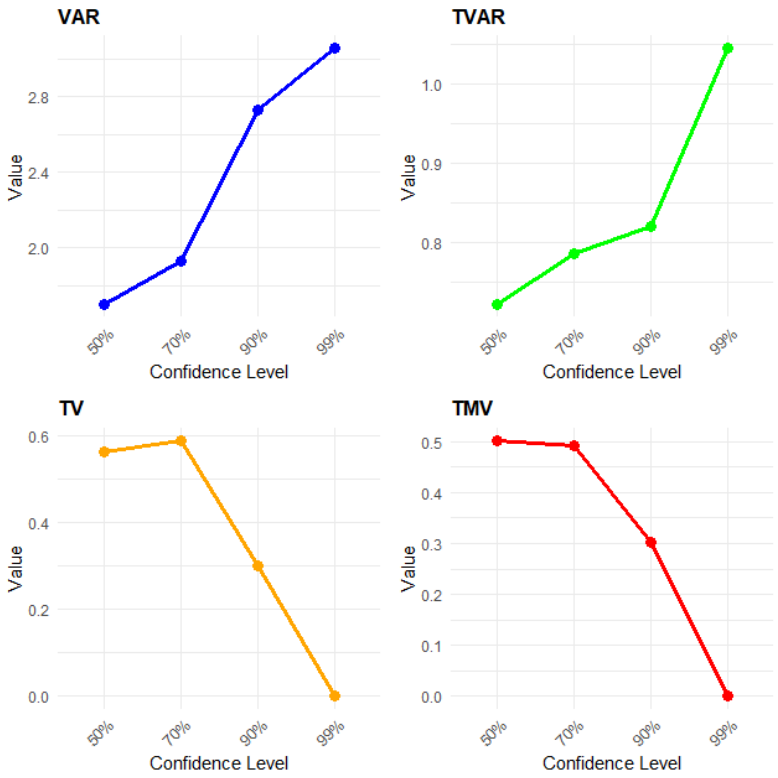


Figure 8: Risk indicators across the CLs for relief times.

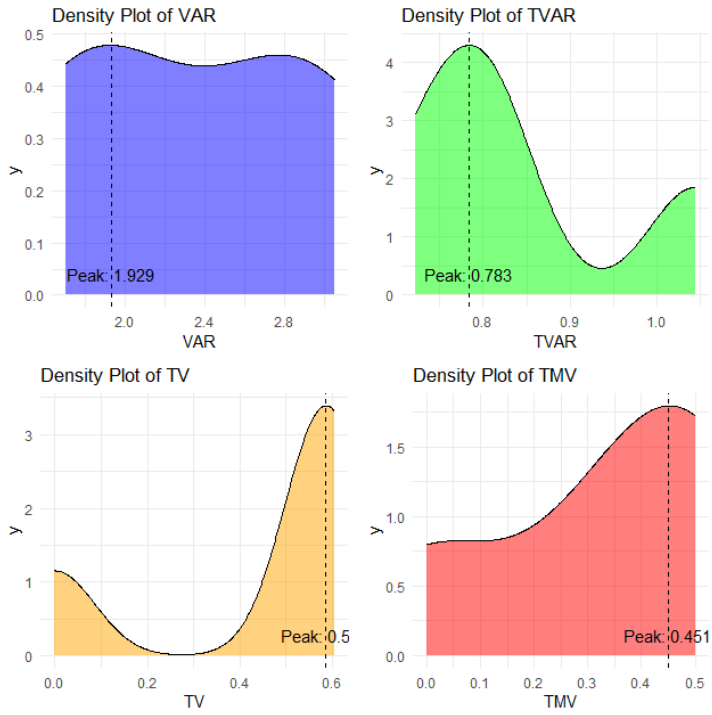


Figure 9: Kernel plots for the risk indicators across the CLs for relief times.

Due to Table 6, we can say:

I. $\text{VAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ increases with the confidence level, from 115.925 at 50% to 157.169 at 99%. This suggests that as we aim to be more confident about risk estimates, the potential maximum loss increases. This is typical as higher confidence levels account for more extreme scenarios.

II. $\text{TVAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ also increases from 26.82 at 50% to 17.55 at 99%. This indicates that the expected loss in the tail of the distribution becomes more pronounced at higher confidence levels but shows a decrease in extreme cases, which could imply improved risk mitigation or less variability in extreme scenarios.

III. $\text{TV}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ also increases with the confidence level, peaking at 274.48 at the 99% level. This indicates that total losses or costs are higher at more extreme confidence levels, reflecting the broader risk exposure.

IV. $\text{TMV}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ shows a similar pattern to $\text{TV}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$, with values increasing and then decreasing slightly at the highest confidence level (137.24). This fluctuation could suggest changes in risk distribution or anomalies at extreme confidence levels.

V. The variation in $\text{TVAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ and $\text{TMV}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ at the highest confidence level (99%) compared to lower confidence levels suggests that while overall risks increase, there might be specific cases where the risk distribution stabilizes or changes. This could reflect changes in data behavior or risk management effectiveness.

Table 6: Risk analysis under minimum flow.

CL	$\text{VAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$	$\text{TVAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$	$\text{TV}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$	$\text{TMV}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$
50%	115.925	26.81974	266.6856	252.6495
70%	135.94	16.14417	163.8007	150.1506
90%	151.332	12.91800	241.2131	180.9099
99%	157.169	17.54600	274.4825	137.2412

Therefore, we advise with the following recommendations:

1. Regularly update risk assessment models to incorporate the latest data, reflecting changes in minimum flow conditions. This helps predict and manage flood risks more accurately.
2. Implement or upgrade early warning systems for high-risk scenarios, especially at higher confidence levels (90% and 99%) where potential risks are more significant.
3. Consider reinforcing or upgrading dam infrastructure to manage increased risk levels and handle extreme flow conditions effectively.
4. Create and periodically review contingency plans for extreme flood events indicated by high $\text{VAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ and $\text{TVAR}\left(\boldsymbol{y};\hat{\underline{\Xi}}\right)$ values. Ensure that emergency response teams are trained and equipped to handle these scenarios.

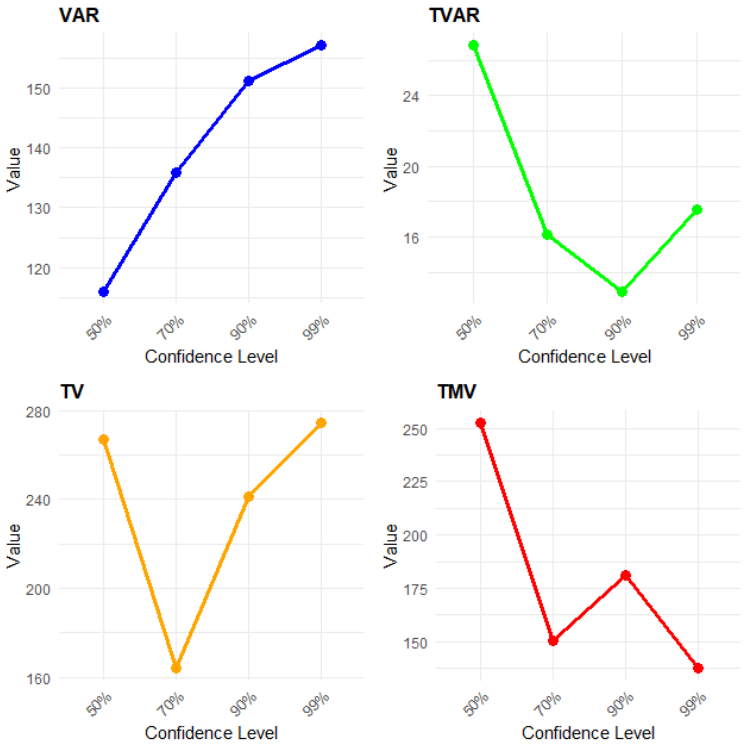


Figure 10: Risk indicators across the CLs for minimum flow.

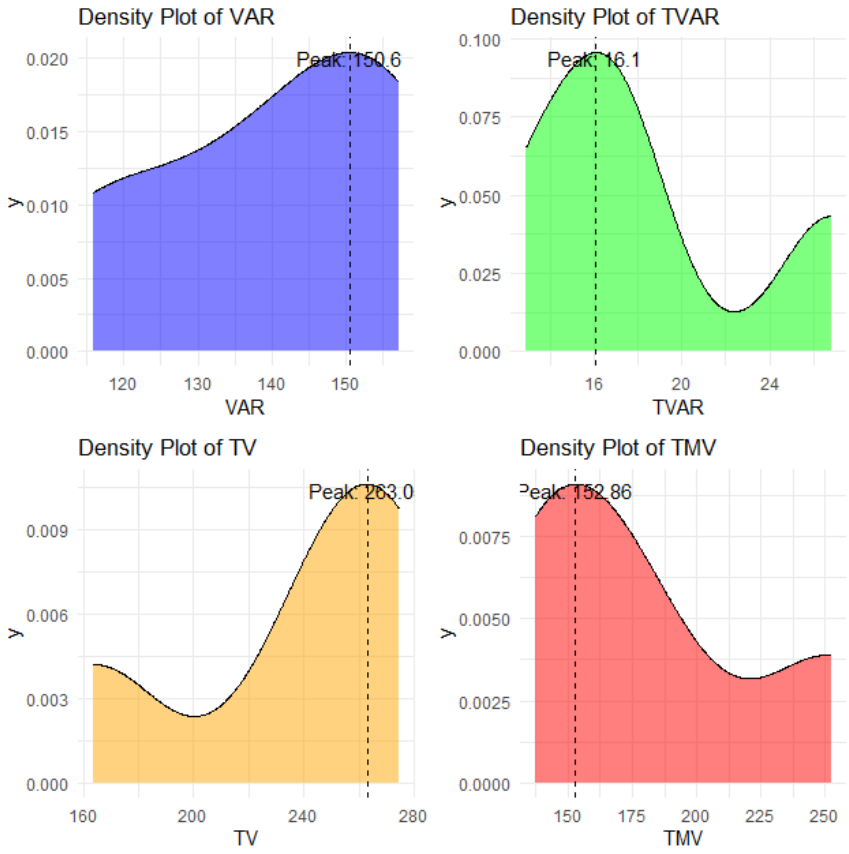


Figure 11: Kernel plots for the risk indicators across the CLs for minimum flow.

7. Financial case study under the U.S. indemnity losses

This dataset contains 1500 records of U.S. indemnity losses from general liability claims, with amounts expressed in thousands of U.S. dollars. It is part of the CASdatasets package in R, a well-known resource for actuaries and risk analysts. The data reflects real claim payments, offering valuable insights into loss distributions commonly encountered in insurance. One notable feature is its right-skewed distribution, most claims are relatively small, but a few large losses extend the upper tail. This pattern is typical in insurance datasets and has important implications for risk modeling and capital allocation. Understanding the frequency and severity of claims helps insurers better estimate potential losses and set appropriate premium rates. The dataset is especially useful for studying heavy-tailed distributions and testing various actuarial models. Its practical relevance makes it a popular choice in both academic research and industry applications. By analyzing this data, risk professionals can improve their predictive accuracy and enhance risk management strategies. Due to these characteristics, the U.S. indemnity losses from general liability claims is selected for this case study under the new Chen distribution.

The Cullen and Frey graph provides strong evidence that the U.S. indemnity losses from general liability claims dataset exhibits right-skewed and heavy-tailed characteristics, consistent with real insurance loss patterns. The empirical data not fully aligns closely with lognormal, gamma, and Weibull distributions, making these theoretical models promising candidates for further analysis. By selecting an appropriate distribution, actuaries and risk analysts can improve their predictive accuracy, enhance risk management strategies, and ensure financial stability in the face of extreme indemnity losses. This graphical approach serves as a valuable first step in the modeling process, guiding subsequent parameter estimation and validation efforts.

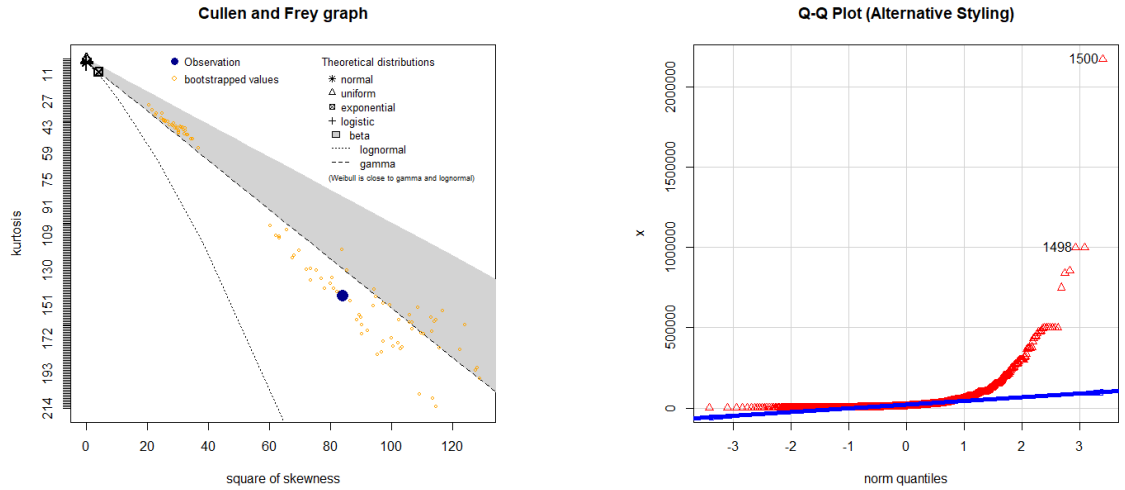


Figure 12: The Cullen and Frey plot (left) and quantile-quantile plot (right) for the U.S. indemnity losses from general liability claims.

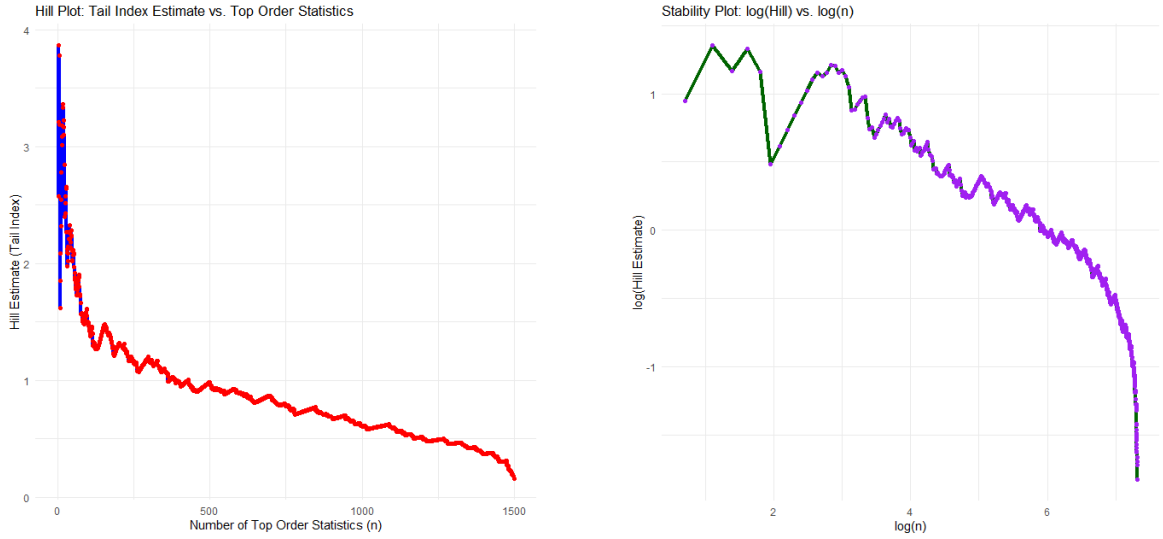


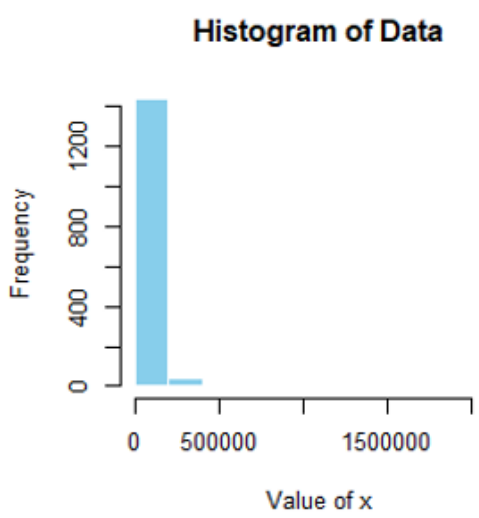
Figure 13: the Hill estimator (left) and the stability plot (right) the U.S. indemnity losses from general liability claims.

As shown in Table 7, the MOOP analysis reveals how increasing the number of P from 1 to 5 leads to a progressive refinement in model performance, reflected in decreasing MSE and bias values, MSE drops from 41208.7 at P=1 to 41172.8 at P=5, while bias follows a similar downward trend from 41198.42 to 4170.42, indicating that higher P values yield better alignment between estimated and actual loss distributions; notably, the corresponding MOOP values rise steadily from 10 to 38, suggesting that optimal thresholds for minimizing error shift upward as model complexity increases, which has practical implications for setting retention levels, reinsurance strategies, and capital reserves; although the improvements in MSE and bias continue through P=5, the marginal gains become smaller beyond P=4, hinting at potential overfitting or diminishing returns in model enhancement; financially, these findings support the use of the new Chen distribution with P=4 or P=5 for more accurate estimation of claim severity, which directly impacts premium pricing, solvency capital requirements, and risk-adjusted profitability; by capturing the tail behavior more effectively, insurers can better anticipate rare but costly events, allocate capital efficiently, and design

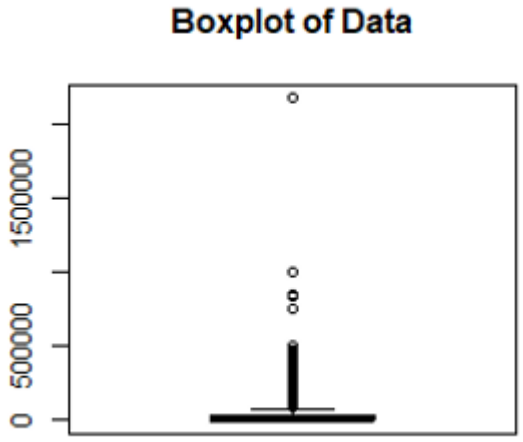
reinsurance programs that balance cost and protection; ultimately, this MOOP analysis under the new Chen distribution demonstrates strong empirical relevance and methodological value in handling real indemnity loss data, reinforcing its applicability in both academic research and industry practice for robust risk modeling and informed financial strategy. Figure 14 gives the Histogram (a), the box plot (b), the mean of order P (c) and the Bias vs. MSE (d) for the U.S. indemnity losses from general liability claims.

Table 7: MOO^P analysis under the U.S. indemnity losses from general liability claims.

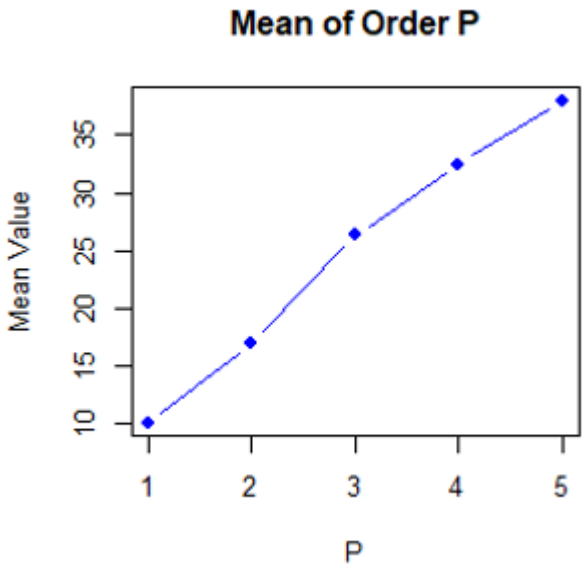
P	1	2	3	4	5
TMV			41208.42		
MOO^P	10.00	17.00	26.33333	32.5 38	38.00
MSE	41208.7	41191.4	41182.1	41175.9	41172.8
Bias	41198.42	41191.42	41182.09	41175.92	41170.42

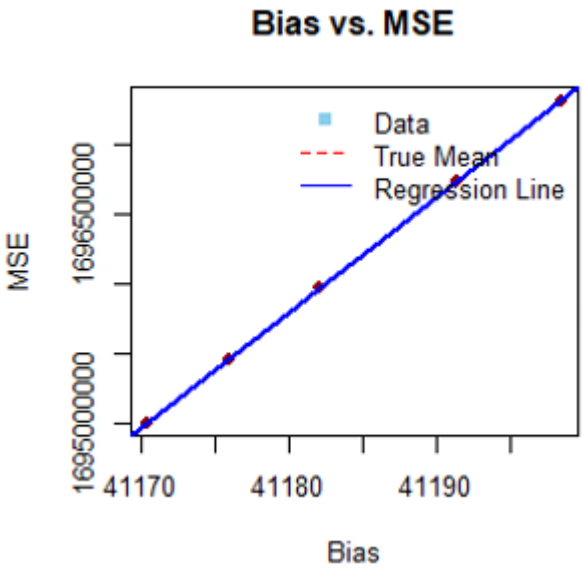


(a)



(b)





(c) (d)

Figure 14: The Histogram (a), the box plot (b), the mean of order P (c) and the Bias vs. MSE (d) for the U.S. indemnity losses from general liability claims.

The PORT-VaR analysis presented in Table 8 offers a detailed view of the tail behavior of the U.S. indemnity losses from general liability claims, a dataset known for its heavy-tailed and right-skewed nature—characteristics that are critical in actuarial modeling, risk management, and solvency planning within the insurance industry; as the CL increases from 50% to 99% , the corresponding VaR threshold decreases, starting from 12000 at 50% and falling to just 147 at 99%, reflecting a progressively more stringent filter for identifying extreme losses; this inverse relationship between CL and threshold is expected in VaR-based approaches, where higher confidence levels require lower thresholds to capture a larger portion of the loss distribution's tail; correspondingly, the number of exceedances or "peaks" over the threshold increases from 748 at 50% to 1,484 at 99% , indicating that more data points qualify as "extreme" under looser thresholds associated with higher confidence levels; examining the statistical summaries of these peaks reveals key financial insights: at the 50% confidence level , where only the most severe losses are considered, the mean claim size is 77970 , with a median of 35000, highlighting the influence of extremely large losses skewing the average upward, while the third quartile (Q3) sits at 75000 and the maximum remains consistent across all levels at 2173595 , representing the largest indemnity payment in the dataset; as we move to the 99% confidence level , the mean claim size declines slightly to 41652 , the median drops to 12500, and the first quartile (Q1) falls to 4028 , illustrating how lowering the threshold captures a broader spectrum of claims, including smaller but still significant losses; this pattern supports the use of extreme value theory (EVT) techniques like PORT-VaR, which allow insurers to model the tail of the loss distribution flexibly by adjusting thresholds according to desired confidence levels; from a financial risk perspective , these results have several important implications: first, the increasing number of peaks at higher CLs provides more data for fitting tail models, improving the robustness of estimates for high-impact events; second, the wide dispersion in peak values, from minimums just above the threshold to extreme losses exceeding 2 million, underscores the need for careful capital provisioning and reinsurance protection; third, the decline in mean and median values at higher CLs suggests that many smaller, yet still material, losses exist in the upper tail, which may be overlooked if only very high thresholds are used; therefore, selecting an appropriate threshold involves balancing model stability, data sufficiency, and risk sensitivity, especially when calculating capital-at-risk , premium loadings , or reinsurance retention levels ; ultimately, this PORT-VaR analysis on the U.S. indemnity losses dataset illustrates how threshold selection impacts the characterization of extreme claims, offering actuaries and risk professionals a practical framework to assess and manage downside risk in a financially meaningful way. Figure 15 presents the Histogram of the U.S. indemnity losses from general liability claims with peaks above VaR & CL=50%, 70%, 90% and 99%.

Table 8: PORT-VAR analysis under the U.S. indemnity losses from general liability claims.

CL	VAR Threshold	N. Peaks	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
50%	12000	748	12100	20000	35000	77970	75000	2173595
70%	5000	1017	5003	11280	23000	59573	55000	2173595
90%	1500	1347	1522	5450	15000	45791	40750	2173595
99%	147	1484	165	4028	12500	41652	35000	2173595

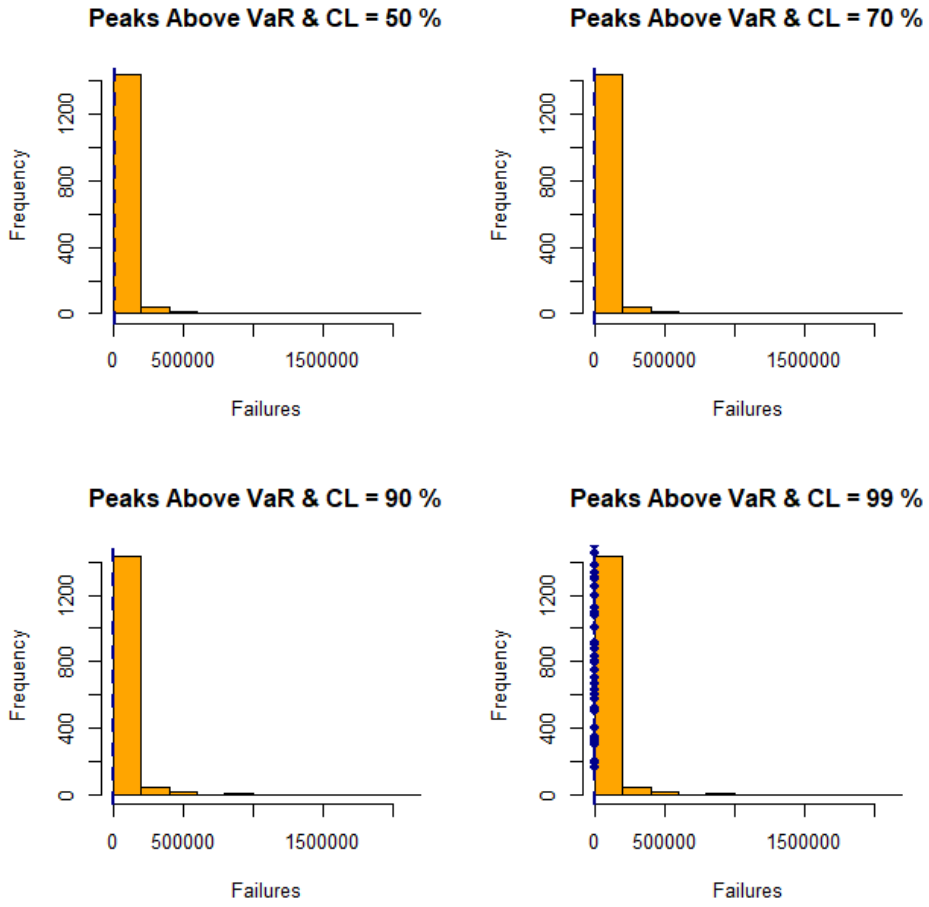


Figure 15: Histogram of the U.S. indemnity losses from general liability claims with peaks above VaR & CL=50%, 70%, 90% and 99%.

Table 9 presents a risk analysis of the U.S. indemnity losses from general liability claims, focusing on two key actuarial metrics, VaR and MEXL across different confidence levels (CL), offering critical insights into the tail behavior and financial exposure associated with insurance claims. As the confidence level increases from 50% to 99%, both VaR and MEXL values rise significantly, reflecting the increasing severity of losses that insurers must account for in their risk models, pricing strategies, and capital planning. Figure 16 gives the risk indicators across the CLs (a) and Kernel plots (b) for the risk indicators across the CLs for the U.S. indemnity losses from general liability claims.

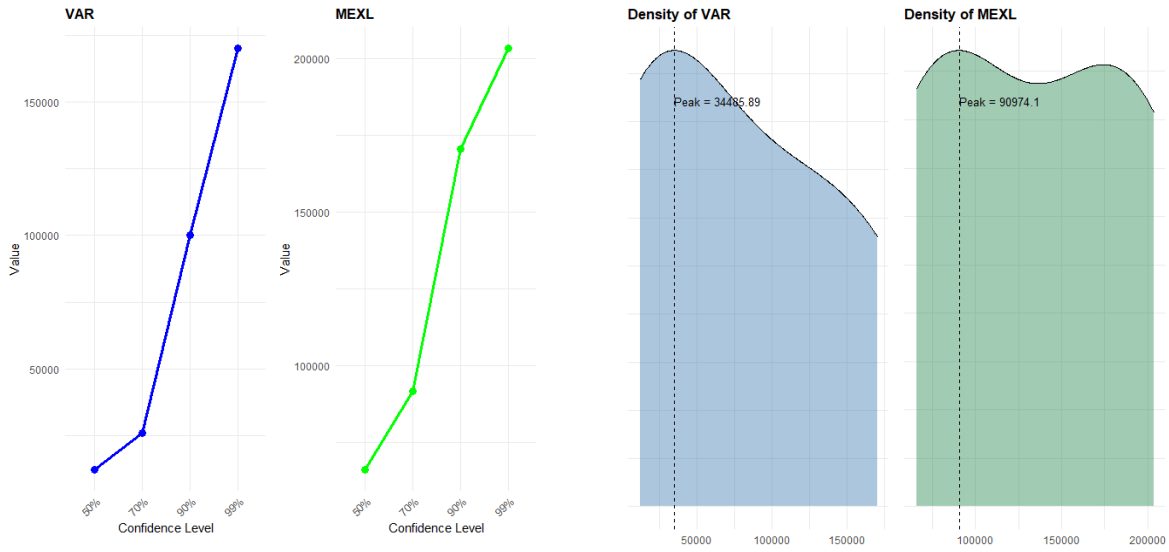
Table 9: Risk analysis under the U.S. indemnity losses from general liability claims.

CL	VAR$\left(\boldsymbol{y};\hat{\Xi}\right)$	MEXL$\left(\boldsymbol{y};\hat{\Xi}\right)$
50%	12000	65969.66
70%	25792.3	91489.48
90%	100000	170567.1
99%	170400	203411.1

At the 50% confidence level, the VaR is estimated at 12000 (i.e., 12 million USD), indicating that there is a 50% chance that losses will not exceed this amount. The corresponding MEXL of 65969.66 suggests that, on average, losses exceeding this threshold are approximately 66 million, more than five times the VaR itself, highlighting the heavy-tailed nature of the data where extreme losses can be vastly larger than typical claim amounts. Moving to the 70% confidence level, the VaR jumps to 25792.3, nearly doubling, while the MEXL increases to 91,489.48, showing that as we look deeper into the tail, both the threshold for extreme events and the expected excess loss grow substantially.

At the 90% confidence level, the VaR reaches 100000, marking a significant jump and signaling a shift toward modeling more extreme outcomes. The MEXL also rises to 170567.1, meaning that once the 100 million threshold is crossed, the average additional loss is around 170.5 million, underscoring the need for robust capital buffers and reinsurance protection to cover such high-severity events. Finally, at the 99% confidence level, which captures the most extreme 1% of losses, the VaR soars to 170,400, and the MEXL hits 203411.1, indicating that once this extremely high threshold is breached, the average loss is even greater than the threshold itself, an important feature of heavy-tailed distributions like the one observed in this dataset.

From a financial risk management perspective , these results have profound implications for insurers: first, the VaR values help establish minimum capital requirements at various confidence levels, ensuring solvency under adverse scenarios; second, the rising MEXL figures highlight the growing marginal cost of insuring against increasingly severe events, justifying higher premium loadings and reinsurance participation; third, the fact that MEXL exceeds VaR at higher confidence levels confirms the nonlinear growth of tail risk , reinforcing the importance of using advanced statistical models, such as those based on extreme value theory or the new Chen distribution, to accurately capture and price such exposures. These findings are particularly relevant for enterprise risk management, solvency assessments (e.g., Solvency II, NAIC RBC), and pricing catastrophe covers, where understanding the dynamics of large claims is essential for maintaining financial stability and competitiveness in the insurance market.



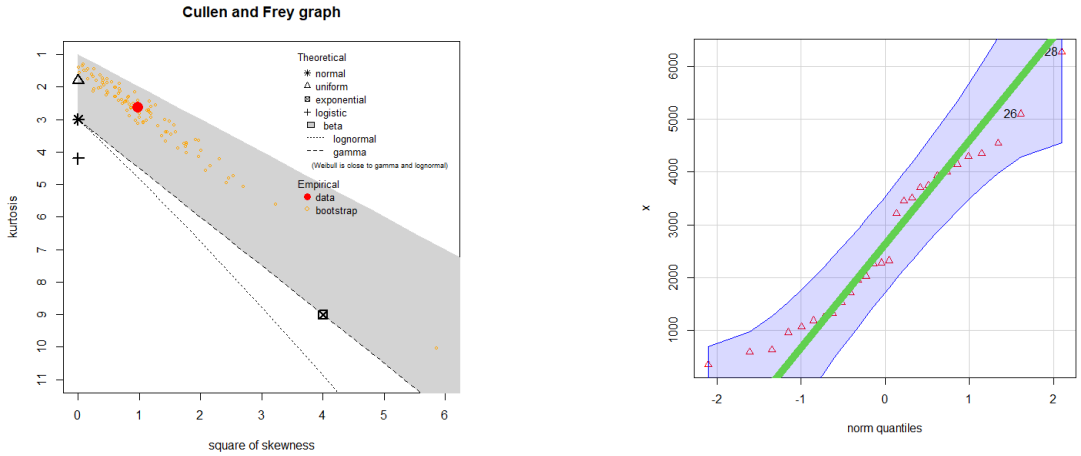
(a) (b)

Figure 16: Risk indicators across the CLs (a) and Kernel plots (b) for the risk indicators across the CLs for the U.S. indemnity losses from general liability claims.

8. A comparative study under historical financial claims data

In this Section, we provide a comparative study under a historical financial claims data through an application. In this application, we analyze the insurance claims payment triangle from a U.K. Motor Non-Comprehensive account. For

convenience, we set the origin period from 2007 to 2013 (see Charpentier (2014)). For this purpose, we consider exponential distribution and some Chen extensions such as the reduced Rayleigh Chen (RRC), exponential Chen (EC), reduced Burr X Chen (RBXC), Burr X exponentiated Chen (RBXC), Weibull Chen (WC), Lomax Chen (LOXC), and Log logistic Chen (LOGLC), among others. For comparing models, we calculate the '− log L' Criteria, Akaike Information Criteria (T1), Consistent Information Criteria (T2), Bayesian Information Criteria (T3), Hannan-Quinn Information Criteria (T4), Cramér-Von Mises (T5), and Anderson-Darling (T6). Additionally, the Kolmogorov–Smirnov (KS) test and its corresponding p-value (P-v) are also computed. Table 10 gives the Completive models, estimates and (standard errors) for the historical claims data. Table 11 lists the test statistics for all models under the historical claims data. Figure 17 presents the Cullen and Frey plot (a) and quantile-quantile plot (b) for the historical claims data. Based on Table 11, the new model is the best distribution according to the selected test statistics.



(a) (b)

Figure 17: The Cullen and Frey plot (a) and quantile-quantile plot (b) for the historical claims data.

Table 10: The Completive models, estimates and (standard errors) for the historical claims data.

Models	estimates and (standard errors)			
	θ_1	θ_2	θ_3	θ_4
Exp	0.130134(0.02462)			
RRC	0.003544(0.00066)			
EC	0.031454(0.00503)			
RBXC	$1.3\times10^{3}(2.2\times10^{2})$	$3.6\times10^{-3}(5.92\times10^{-4})$		
WC	0.028034(0.0035)	0.59339(0.00189)		
LOXC	0.03575(0.00624)	0.593311(0.0015)		
LOGLC	0.05451(0.00835)	0.59375(0.00231)		
GZC	$8.001\times10^{5}(0.3\times10^{-2})$	$2.11\times10^{-1}(1.3\times10^{-3})$	1.000052(0.01)	
BXIIC	0.49286(1.81777)	0.072396(0.2678)	0.5936(0.002)	

BXEC	1.54039(0.34261)	0.64833(0.18652)	3.9933(0.884)	0.1658(0.0584)
MOGC	0.03714(0.01486)	0.97891(0.48343)	0.3165(0.004)	0.71873(0.004)

Table 11: The test statistics for all models under the historical claims data.

Models	log *L*	*T1*	*T2*	*T3*	*T4*	*T5*	*T6*	KS(P-v)
GZC	35.3561	74.71272	75.19274	77.37711	75.52723	0.100202	0.31771	0.13391(0.5742)
MOGC	53.8473	115.6944	117.4331	121.0224	117.3235	0.102793	0.63533	0.3668(0.0012)
BXEC	64.6582	137.3163	139.0554	142.6445	138.9558	0.120825	0.76583	0.5071($<10^{-6}$)
BXIIC	64.8530	135.7059	136.7059	139.7025	136.9277	0.10920	0.68190	0.6588($<10^{-6}$)
LOXC	64.8586	133.7171	134.1971	136.3815	134.5316	0.10923	0.68174	0.5581($<10^{-6}$)
LOGLC	79.1675	162.335	162.815	164.9994	163.1495	0.09687	0.58813	0.7481($<10^{-6}$))
WC	83.6661	171.3321	171.8121	173.9965	172.1466	0.08649	0.50773	0.7532($<10^{-6}$)
Exp	85.1028	172.2056	172.360	173.538	172.613	0.128994	0.82191	0.6316($<10^{-6}$)
EC	228.844	459.6884	459.8422	461.0206	460.0957	0.13448	0.85879	0.9965($<10^{-6}$)
RBXC	493.752	991.504	991.984	994.1684	992.3186	0.13663	0.87319	0.99994($<10^{-6}$)
RRC	690.104	1382.209	1382.363	1383.541	1382.616	0.13647	0.87212	0.98992($<10^{-6}$)

9. Conclusion

In this study, we presented an innovative Chen model tailored for analyzing extreme low-flow events in water systems and rare events in the medical sector, focusing on both maintaining ecological balance in rivers and optimizing risk management strategies. We begin with a detailed mathematical derivation of the new Chen model, which extends traditional Chen distribution properties to more effectively capture extreme values by addressing heavy tails and complex behaviors. Following this, we conduct a Mean-of-Order-P (MOO^P) analysis to determine the optimal order P, known as Opt-P, which helps balance model accuracy and complexity. Then, using the new model, the Peaks over a Random Threshold Value-at-risk (PORT-VAR) analysis is used to set several random thresholds and analyze notable peaks, both highs and lows. This technique aids in identifying the most important extreme values and comprehending the ramifications. Value-at-Risk (VAR), which calculates potential losses over a given period of time; Tail Value-at-Risk (TVAR), which measures the average loss exceeding the VAR threshold; Mean Excess Loss (MEXL), which calculates the average loss above a given threshold; Tail Variance (TV), which evaluates the variability of extreme losses; and Tail Mean Variance (TMV), which combines the mean and variance of extreme losses, are some of the key metrics used for a thorough risk assessment.

Relief data and minimum flow data are used in the research, which helps decision-makers optimize risk management tactics and offers useful insights into the model's applicability in real circumstances. This method improves decision-making and risk reduction in the dam sciences and medical risk management, while also deepening our understanding of extreme occurrences. All things considered, the new Chen distribution provided an adaptable framework for simulating extreme values and tail behavior, which is essential for PORT-VAR evaluations as well as MOOP research. Extreme relief measures and low hydrological flows are examples of rare and extreme events that are well-suited for analysis due to its capacity to capture heavy tails and accommodate different levels of skewness and kurtosis. This is especially important for precisely estimating risk in situations where conventional distributions might not be sufficient. In conclusion, the subsequent useful suggestions are emphasized:

I. By forecasting the medical effects of exceptional relief occurrences, the new Chen distribution enables more efficient resource allocation and budget planning.
II. The revised Chen distribution can help ensure availability and sustainability of water resources during droughts by more precisely modelling extreme low-flow episodes.

Finally, a comparative study under a historical financial claims data through an application, the new model gives the best fits among all the competitive models.

The findings of this study, centered on the development and application of the Burr Generalized Chen (BGC) distribution, offer significant insights across multiple domains, hydrology, medicine, and financial risk management, by enhancing the modeling of extreme events through advanced statistical techniques such as MOOP and PORT-VaR analyses. In hydrology, the BGC model successfully captures extreme low-flow conditions using real river flow data from the Cuiabá River, demonstrating that increasing the order P leads to reduced Mean Squared Error (MSE) and Bias, thereby improving estimation accuracy for central tendencies in skewed and heavy-tailed datasets. The PORT-VaR analysis further reveals that as confidence levels increase from 50% to 99%, the number of exceedances grows, allowing for more robust modeling of drought risks and dam safety thresholds. This has direct implications for adaptive water resource planning, particularly under climate uncertainty, where accurate prediction of low-flow extremes is essential for maintaining ecological health and ensuring sustainable reservoir operations. In the medical domain, the model is applied to relief times data, where it effectively identifies extreme values and improves estimation precision with higher P values, reducing both MSE and Bias significantly critical for assessing treatment efficacy and setting realistic expectations in clinical decision-making. The financial case study using U.S. indemnity loss data from general liability claims highlights the BGC model's ability to capture tail behavior and estimate extreme risks with high fidelity; at the 99% confidence level, VaR reaches 170400 (in thousands of USD), while the MEXL hits 203,411.1, illustrating the nonlinear growth of risk in heavy-tailed insurance claims. These results are crucial for capital adequacy assessments, reinsurance pricing, and solvency planning, especially under regulatory frameworks like Solvency II. A comparative study against other models confirms that the BGC distribution outperforms existing distributions in terms of goodness-of-fit criteria such as Akaike Information Criterio, Bayesian Information Criterion, and Anderson-Darling test statistics. The broader implication of this research lies in its contribution to the growing field of extreme value theory and risk modeling, offering a flexible and robust framework that can be adapted to asymmetric, heavy-tailed data across disciplines. Policymakers, actuaries, environmental scientists, and healthcare professionals can leverage these findings to enhance predictive accuracy, optimize risk mitigation strategies, and support informed decision-making under uncertainty, ultimately contributing to improved resilience in the face of rare but high-impact events.

Acknowledgments: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU260940].
Contributions: All authors participated equally in the preparation of the paper and have equal shares in all types of contributions.
Data Availability: Prof. H.M. Yousof will provide the data upon request. Also, the data can be accessed through the references.
Conflict of interests: The authors declare that there is no conflict of interests.

References

1. Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. Journal of Banking & Finance, 26(7), 1487-1503.
2. Alizadeh, M., Afshari, M., Contreras-Reyes, J. E., Mazarei, D., & Yousof, H. M. (2024). The Extended Gompertz Model: Applications, Mean of Order P Assessment and Statistical Threshold Risk Analysis Based on Extreme Stresses Data. IEEE Transactions on Reliability, doi: 10.1109/TR.2024.3425278
3. Alizadeh, M., Afshari, M., Ranjbar, V., Merovci, F. and Yousof, H. M. (2023). A novel XGamma extension: applications and actuarial risk analysis under the reinsurance data. São Paulo Journal of Mathematical Sciences, 1-31.
4. Artzner, P. (1999). Application of coherent risk measures to capital requirements in insurance. North American Actuarial Journal, 3(2), 11-25.
5. Balakrishnan, N. and Cohen, A. C. (2014). Order statistics & inference: estimation methods. Elsevier.

6. Chaubey, Y. P. and Zhang, R. (2015). An extension of Chen's family of survival distributions with bathtub shape or increasing hazard rate function., 44(19), 4049-4064.
7. Charpentier, A. (2014). Computational Actuarial Science with R, CRC press, 2014.
8. Cordeiro, G. M., Nadarajah, S. and Ortega, E. M. (2012). The Kumaraswamy Gumbel distribution. Statistical Methods and Applications, 21, 139-168.
9. Dey, S., Kumar, D., Ramos, P. L. and Louzada, F. (2017). Exponentiated Chen distribution: Properties and estimation. Communications in Statistics-Simulation and Computation, 46(10), 8118-8139.
10. Elbatal, I., Diab, L. S., Ghorbal, A. B., Yousof, H. M., Elgarhy, M., & Ali, E. I. (2024). A new losses (revenues) probability model with entropy analysis, applications and case studies for value-at-risk modeling and mean of order-P analysis. AIMS Mathematics, 9(3), 7169-7211.
11. Figueiredo, F., Gomes, M. I., & Henriques-Rodrigues, L. (2017). Value-at-risk estimation and the PORT mean-of-order-p methodology. REVSTAT-Statistical Journal, 15(2), 187-204.
12. Furman, E. and Landsman, Z. (2006). Tail Variance premium with applications for elliptical portfolio of risks. ASTIN Bulletin: The Journal of the IAA, 36(2), 433-462.
13. Gross, A. J. and Clark, V. A. (1975). Survival distributions: Reliability Applications in the
14. Biomedical Sciences. John Wiley and Sons, New York.
15. Hamed, M. S., Cordeiro, G. M. and Yousof, H. M. (2022). A New Compound Lomax Model: Properties, Copulas, Modeling and Risk Analysis Utilizing the Negatively Skewed Insurance Claims Data. Pakistan Journal of Statistics and Operation Research, 18(3), 601-631.
16. Hamedani, G. G., Goual, H., Emam, W., Tashkandy, Y., Ahmad Bhatti, F., Ibrahim, M. and Yousof, H. M. (2023). A new right-skewed one-parameter distribution with mathematical characterizations, distributional validation, and actuarial risk analysis, with applications. Symmetry, 15(7), 1297.
17. Hashempour, M., Alizadeh, M. and Yousof, H. M. (2023). A New Lindley Extension: Estimation, Risk Assessment and Analysis Under Bimodal Right Skewed Precipitation Data. Annals of Data Science, 1-40.
18. Ibrahim, M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Yousof, H.M. (2023). Bayesian and Non-Bayesian Risk Analysis and Assessment under Left-Skewed Insurance Data and a Novel Compound Reciprocal Rayleigh Extension. Mathematics 2023, 11, 1593.
19. Klugman, S. A., Panjer, H. H. and Willmot, G. E. (2012). Loss models: from data to decisions (Vol. 715). John Wiley & Sons.
20. Korkmaz, M. Ç., Altun, E., Yousof, H. M., Afify, A. Z. and Nadarajah, S. (2018). The Burr X Pareto Distribution: Properties, Applications and VAR estimation. Journal of Risk and Financial Management, 11(1), 1.
21. Lane, M. N. (2000). Pricing risk transfer transactions1. ASTIN Bulletin: The Journal of the IAA, 30(2), 259-293.
1. Loubna, H., Goual, H., Alghamdi, F. M., Mustafa, M. S., Tekle Mekiso, G., Ali, M. M., ... & Yousof, H. M. (2024). The quasi-xgamma frailty model with survival analysis under heterogeneity problem, validation testing, and risk analysis for emergency care data. Scientific Reports, 14(1), 8973.
2. Mohamed, H. S., Cordeiro, G. M., Minkah, R., Yousof, H. M., & Ibrahim, M. (2024). A size-of-loss model for the negatively skewed insurance claims data: applications, risk analysis using different methods and statistical forecasting. Journal of Applied Statistics, 51(2), 348-369.
3. Rasekhi, M., Altun, E., Alizadeh, M. and Yousof, H. M. (2022). The Odd Log-Logistic Weibull-G Family of Distributions with Regression and Financial Risk Models. Journal of the Operations Research Society of China, 10(1), 133-158.
4. Shrahili, M., Elbatal, I., & M. Yousof, H. (2021). Asymmetric Density for Risk Claim-Size Data: Prediction and Bimodal Data Applications. Symmetry, 13(12), 2357.
5. Tasche, D. Expected Shortfall and Beyond, Journal of Banking and Finance 2002, 26, 1519-1533.
6. Wirch, J. L. (1999). Raising value at risk. North American Actuarial Journal, 3(2), 106-115.
7. Yousof, H. M., Ansari, S. I., Tashkandy, Y., Emam, W., Ali, M. M., Ibrahim, M., Alkhayyat, S. L. (2023a). Risk Analysis and Estimation of a Bimodal Heavy-Tailed Burr XII Model in Insurance Data: Exploring Multiple Methods and Applications. Mathematics. 2023; 11(9), 2179.
8. Yousof, H.M.; Emam, W.; Tashkandy, Y.; Ali, M.M.; Minkah, R.; Ibrahim, M. (2023b). A Novel Model for Quantitative Risk Assessment under Claim-Size Data with Bimodal and Symmetric Data Modeling. Mathematics 2023, 11, 1284.

9. Yousof, H.M.; Tashkandy, Y.; Emam, W.; Ali, M.M.; Ibrahim, M. (2023c). A New Reciprocal Weibull Extension for Modeling Extreme Values with Risk Analysis under Insurance Data. Mathematics 2023, 11, 966.