Research and Analysis Journals 8(04): 06-20, 2025

e-ISSN: 2589-9228, p-ISSN: 2589-921x

Research Article

Power-Lindley Generalized Pareto Distribution: A New Approach for Modeling Heavy-Tailed Data

[1]Nkemnole E. B., [2]Okoacha, N.G.

[1]Department of Statistics, University of Lagos, Nigeria.

[2]Basic Science Unit, School of Science and Technology, Pan-Atlantic University, Nigeria

Abstract:

This research introduces the Power-Lindley Generalized Pareto Distribution (PL-GPD), a new statistical model designed to better capture heavy-tailed data, which is often found in finance, climate studies, and other fields where extreme events are significant. Unlike existing models like the Normal and Skew Normal Generalized Pareto Distributions (NGPD and SNGPD) by Debbabi et al. (2012) and Debbabi et al. (2016) respectively, which struggle with the complexities of such data, the PL-GPD combines the Power-Lindley distribution (for moderate values) with the Generalized Pareto distribution (for extreme values).Using Maximum Likelihood Estimation (MLE), the PL-GPD was applied to S&P 500 log return data and compared to NGPD and SNGPD based on Akaike Information Criterion (AIC) and Kolmogorov-Smirnov (K-S) statistic. The PL-GPD showed a superior fit, with lower AIC and higher K-S p-values, indicating improved accuracy in modeling rare but impactful events. This enhanced ability to manage extreme data makes PL-GPD especially useful in risk management for fields like finance and insurance.

Keywords: Power-Lindley Generalized Pareto Distribution (PL-GPD), Heavy-tailed distributions, Kolmogorov-Smirnov (K-S) statistic, Extreme values and risk management, Tail risk.

1. Introduction

Heavy-tailed data distributions play a crucial role in accurately modeling rare and impactful events in areas like finance, insurance, and environmental studies. Such events, although infrequent, pose substantial risks, requiring models that can reliably capture these extremes. Traditional probability distributions, including the Normal and Exponential distributions, lack the flexibility to account for both regular and extreme events within a single model structure. Consequently, researchers have developed models like the Normal Generalized Pareto Distribution (NGPD) and Skew Normal Generalized Pareto Distribution (SNGPD) to address these challenges, yet these models often fail to accurately capture heavy tails, leading to biased estimates and unreliable risk predictions. To address these gaps, this study introduces the Power-Lindley Generalized Pareto Distribution (PL-GPD), a hybrid model combining the Power-Lindley distribution (for moderate data values) and the Generalized Pareto distribution (for extreme values). The PL-GPD is designed to provide a more accurate representation of datasets characterized by moderate and extreme values, which are prevalent in financial markets and climate data. By integrating these two distributions, the PL-GPD offers a unified approach that improves the predictive accuracy and reliability of statistical models in fields where extreme events are critical.

This paper applies the PL-GPD to the S&P 500 log return data and validates its performance by comparing it with NGPD and SNGPD models using metrics such as the Akaike Information Criterion (AIC) and Kolmogorov-Smirnov (K-S) statistic. The findings highlight the PL-GPD's superior ability to model tail risks effectively, making it particularly suitable for applications in finance, climate studies, and risk management.

1.1 Heavy-Tailed Distributions in Risk Modeling

Heavy-tailed distributions are critical in fields where rare, high-impact events occur, such as finance, insurance, hydrology, and climate science. These distributions characterize datasets with "fat tails," meaning the probability of extreme values is higher than in distributions like the normal distribution. In finance, heavy tails represent potential market crashes or extreme losses, while in climate science, they capture catastrophic weather events like hurricanes or droughts. Understanding the behavior of heavy-tailed data is crucial for accurately assessing risks associated with extreme events.

Heavy-tailed distributions include models like the Pareto, Weibull, and Cauchy distributions, each with unique properties suited to specific types of heavy-tailed data. For example, the Pareto distribution is frequently used to model income distribution and risk in insurance. However, traditional models often fall short in capturing data with both moderate and extreme values simultaneously. This has led researchers to explore more flexible composite models that combine the strengths of multiple distributions to handle the complexities inherent in heavy-tailed datasets.

1.2 Limitations of Existing Models for Heavy-Tailed Data

Normal Generalized Pareto Distribution (NGPD):

The Normal Generalized Pareto Distribution (NGPD) integrates features of the normal distribution and the Generalized Pareto Distribution (GPD). This combination provides a model capable of capturing both central tendencies and heavy tails, making it useful for data where both moderate and extreme values are significant. The NGPD has been applied in fields such as finance and environmental studies for modeling losses and extreme events. However, NGPD assumes symmetry around the mean, which limits its effectiveness for datasets with asymmetrical distributions or skewed tail behavior. This lack of flexibility often results in inaccurate modeling of real-world data that exhibits skewness or other complex tail characteristics (Debbabi, et al. 2012).

Skew Normal Generalized Pareto Distribution (SNGPD):

To address the NGPD's limitation regarding symmetry, the Skew Normal Generalized Pareto Distribution (SNGPD) introduces a skewness parameter, making it more adaptable to asymmetric data. The SNGPD has been applied in environmental modeling, finance, and insurance, where data often exhibit directional bias or asymmetry. The skew parameter allows the model to better accommodate datasets with pronounced right or left tails. However, SNGPD is complex and computationally intensive, particularly for large datasets or applications requiring real-time analysis. Additionally, SNGPD does not fully capture the nuances of datasets where moderate and extreme values coexist, as it remains primarily focused on extreme events (Debbabi, et al.(2016)).

1.3 The Power-Lindley Distribution and Its Applications

The Power-Lindley distribution, introduced by Ghitany et al. (2013), extends the Lindley distribution by incorporating a shape parameter, which allows it to adapt to datasets with a variety of tail behaviors. Originally developed to improve data fitting in survival analysis and reliability studies, the Power-Lindley distribution has been widely adopted due to its flexibility in capturing moderate tails. Compared to the exponential distribution, the Power-Lindley distribution provides a better fit for data with non-monotonic hazard rates and varying tail characteristics, which are essential in applications such as biomedical data analysis and reliability testing.

This distribution's capacity to handle moderate values without overemphasizing the extremes makes it suitable for combining with an extreme-value distribution. In the composite PL-GPD model, the Power-Lindley distribution handles the bulk of data that are close to the mean, while the GPD addresses extreme values, forming a versatile framework that bridges moderate and heavy-tailed data behavior.

1.4 Generalized Pareto Distribution (GPD) in Extreme Value Theory

The Generalized Pareto Distribution (GPD) is foundational in extreme value theory and is particularly useful for modeling the tails of distributions. Pickands' theorem (Pickands, 1975) established the GPD as an ideal distribution for capturing extreme observations, especially beyond a certain threshold. With its flexibility in handling tail behavior, the GPD has been applied in numerous fields where rare events have significant impacts, including insurance, hydrology, and financial risk modeling.

The GPD is a two-parameter distribution that can represent several other distributions depending on the value of its shape parameter. For example, when the shape parameter is zero, the GPD becomes equivalent to the exponential distribution; when positive, it resembles a Pareto-type distribution, and for negative values, it indicates a finite upper bound on the distribution. This adaptability makes GPD an effective choice for modeling extreme values, as it captures tail behaviors that other distributions, such as the normal or exponential distributions, cannot.

Despite its utility, the GPD alone is often insufficient for datasets containing both moderate and extreme observations. In such cases, a composite model that uses GPD specifically for the tail behavior, like PL-GPD, can offer a more robust approach for accurately modeling heavy-tailed datasets with mixed characteristics.

1.5 The Development of Composite Models and Introduction of PL-GPD

Composite models have emerged as a response to the limitations of single-distribution models in handling complex datasets. These models combine multiple distributions to better capture heterogeneous data characteristics, such as moderate values and extreme events. Recent advancements in composite modeling focus on developing frameworks that seamlessly integrate two or more distributions, each suited to specific components of the dataset. Composite models often use a threshold to distinguish between moderate and extreme values, applying one distribution to each segment for a more accurate overall fit.

The PL-GPD combines the Power-Lindley distribution (for moderate values) with the Generalized Pareto Distribution (for extreme values), forming a unified model that addresses the entire range of a heavy-tailed dataset. This combination allows for better-fitting models in applications that involve both frequent moderate values and rare but impactful extremes. By choosing an appropriate threshold, the PL-GPD model can flexibly allocate the Power-Lindley component to model regular data and the GPD to capture extreme values, enhancing predictive accuracy and robustness.

1.6 Comparative Advantages of PL-GPD

The PL-GPD model improves upon both NGPD and SNGPD by addressing their primary limitations. While NGPD lacks skewness handling and SNGPD introduces computational complexity with its skew parameter, PL-GPD offers a more balanced approach. By focusing on moderate values through the Power-Lindley distribution and extreme values through the GPD, PL-GPD provides a more accurate fit for datasets that exhibit both central tendency and extreme tail behavior. This dual approach is particularly advantageous for fields like finance, climate modeling, and insurance, where both regular fluctuations and extreme outliers are crucial for accurate risk assessment and forecasting.

2. Related Literature

This literature review consolidates the theoretical advancements, applications, and extensions of the Power Lindley Distribution (PLD), Generalized Pareto Distribution (GPD), and their respective generalizations. These distributions are integral to modeling diverse real-world phenomena, especially where data exhibit complex tail behaviors or extreme values.

The Power Lindley Distribution

The PLD, introduced by Ghitany et al. (2013), is a generalization of the Lindley distribution, designed to improve flexibility in modeling survival and reliability data. Its probability density function (PDF) includes a shape parameter α to better fit diverse datasets. This distribution is particularly suited for survival and reliability analyses, where its ability to fit non-monotonic hazard rates is valuable. The PLD's versatility stems from its PDF:

$$f(x;\theta,\alpha)=\frac{\alpha}{1+\theta}(1+x)^{\alpha}e^{-\theta x},\ \theta>0,\alpha>0$$

Subsequent extensions have refined the PLD for various applications, some of which are Exponentiated Quasi Power Lindley Distribution: Hassan et al. (2020) extended PLD into a five-parameter model, with applications in medical science, particularly in lifetime and reliability studies, the Length-Biased Power Lindley Distribution: Rather and Ozel (2021) introduced a length-biased version to model asymmetric data with applications in engineering and lifetime analysis, the Alpha Power Transformed Extended Power Lindley Distribution (Eissa and Sonar (2023)) proposed an alpha power transformation to improve the PLD for datasets with varying tail behaviors, emphasizing flexibility in lifetime modeling, Software Metrics Applications by Khalleefah et al. (2019, 2021) applied the PLD to software reliability, focusing on moment determinacy and order statistics. Additionally, Musa et al. (2023) introduced the Exponentiated Power Lindley-Logarithmic Distribution, combining PLD with a logarithmic component to enhance flexibility for applications in risk analysis.

Applications in medical and engineering domains highlight the PLD's adaptability to modeling non-monotonic hazard rates and heavy/light tails. Generalizations like the NGPL and EQPL continue to improve its ability to fit complex real-world data.

Generalized Pareto Distribution (GPD)

The GPD, rooted in extreme value theory, is a pivotal tool for modeling rare or extreme events, such as financial losses and natural disasters. Its PDF:

$$f(x;\xi,\sigma)=\frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-(1/\xi+1)},\ \xi\neq 0$$

defines its flexibility in capturing heavy tails or bounded extremes, depending on the shape parameter ξ.

The GPD is a cornerstone for extreme value theory, extensively used in finance, hydrology, and reliability engineering. Its generalizations address limitations of the standard model, improving flexibility and tail behavior representation. Key advancements includes the Normal Generalized Pareto Distribution (NGPD) by Debbabi et al. (2012) introduced NGPD to combine features of normal and Pareto distributions, balancing moderate and extreme value modeling. Its Skewed NGPD (Debbabi et al., 2016) extends this framework by incorporating asymmetry, which is critical for financial returns and environmental datasets, the Kumaraswamy Exponentiated Pareto Distribution by Elbatal (2013) proposed this distribution for enhanced tail flexibility, emphasizing its applications in financial risk modeling. The Beta Lindley Distribution by MirMostafaee et al. (2021) developed a beta-mixture Lindley distribution, enabling more robust modeling of datasets with diverse tail shapes, the Flood Frequency Analysis by Anghel and Ilinca (2023) evaluated multiple GPD variations to assess flood probabilities, demonstrating the GPD’s utility in hydrology and environmental risk assessment, the Matrix Variate Pareto Distributions by Zinodiny & Nadarajah (2021) extended the GPD to handle multivariate data structures, applicable in finance and large-scale systems. Further, Haj Ahmad and Almetwally (2022) introduced Bayesian inference methods for GPD and its discrete analogues, demonstrating robust parameter estimation under uncertainty.

Key advancements includes the Beta-Generalized Pareto Distribution (BGPD) by Nassar and Nada (2011) introduced this extension to incorporate beta components, allowing for greater shape adaptability, the New Generalized Pareto Distribution (NGPD) by Jayakumar et al. (2018) developed this fourparameter model, particularly suited for subexponential and heavy-tailed data, the T-Pareto {Y} Families by Hamed et al. (2021) introduced bimodal and multimodal shapes to expand the GPD's applicability to diverse datasets.

Estimation techniques for GPD parameters, such as Maximum Likelihood Estimation (MLE), Probability-Weighted Moments

(PWM), and Bayesian methods, have significantly advanced. Maximum Likelihood Estimation (MLE) by Widely applied for both PLD and GPD, with contributions from Malik and Kumar (2019) and others. Bayesian Methods by Haj Ahmad and Almetwally (2022) introduced Bayesian inference for GPD analogues, improving parameter estimation in uncertain contexts. Order Statistics by El-Din et al. (2017) and Tavangar and Hashemi (2013) explored GPD characterization through order statistics and progressively censored datasets.

Normal Genaralized Pareto Distribution (NPGD) Introduced by Debbabi et al. (2012), the NGPD combines features of the normal and Pareto distributions to model moderate and extreme values. This distribution addresses the limitations of symmetry and tail behavior, making it versatile in fields like environmental studies and insurance.

Skewed Normal Generalized Pareto Distribution (Skewed NGPD) by Debbabi et al. (2016) extended the NGPD by incorporating a skewness parameter γ, making it more effective for asymmetrical datasets. This distribution captures directional biases in data, such as skewed financial returns or environmental datasets. This model enhances the NGPD's applicability, albeit at the cost of computational complexity.

Several composite distributions integrate PLD and GPD features to address asymmetry, skewness, and heavy tails like the Lindley Pareto Distribution by Zeghdoudi et al. (2018) combined Lindley and Pareto distributions for lifetime data, demonstrating superior real-world fits, the New Generalized Pareto Distribution (NGPD) by Jayakumar et al. (2018) developed a four-parameter extension of GPD, focusing on heavy tails and extreme values, the Generalized Exponentiated Exponential Lindley Distribution by Kwong and Nadarajah (2021) refined this model to address datasets with high variability, improving goodness-of-fit measures.

Additionally, Malik & Kumar (2019) characterized GPD properties using generalized order statistics, while Mohamed (2015) proposed a generalized Pareto model for applications in risk analysis.

The PLD, GPD, and their generalizations represent significant strides in statistical modeling. Advances like the NGPD and SNGPD address critical challenges in symmetry and extreme value modeling, while methods such as MLE and PWM optimize parameter estimation. Applications in diverse fields—from hydrology to finance—demonstrate their practical relevance. Future research may focus on computational improvements and extending these models to multivariate frameworks.

3. Methodology

The composite family of Distributions

The development of composite probability distributions is a critical endeavor in statistical modeling, particularly for applications involving highly skewed data. This methodology aims to develop a new composite probability distribution that synergizes the properties of the power Lindley distribution and the Generalized Pareto Distribution (GPD). The power Lindley distribution is known for its flexibility and capability to model data with a moderate tail behavior, while the GPD is well-suited for modeling extreme values and heavy tails. By combining these two distributions, we aim to create a composite distribution that can effectively capture the characteristics of both moderate and extreme values in highly right skewed datasets. The following sections detail the steps involved in constructing this composite distribution, including its theoretical formulation, properties, and the estimation algorithm for its parameters.

Suppose we have a data set which has most of its observations around its mean and the other of its observations being extreme values, and we wish to fit the data set with a piece-wise non-degenerate pdf, we proceed by constructing a composite pdf for the data. To do this, we combine a distribution which can handle the fitting of the bulk of the data around the mean with an extreme value distribution. This extreme value distribution will handle the fitting of the extreme observations which are normally realized after a certain high threshold. Following Pickands Theorem (Pickands, 1975) the distribution suitable for these extreme observations is the GPD and the task will be to combine the other distribution suitable for the bulk of the data with the GPD.

Let f_1 and f_2 be two pdfs each with respective parameter vector Θ_1 and Θ_2 such that each is well-suited for fitting the bulk observations around the mean in the data set and the few extreme observations respectively. Let F_1 and F_2 be the corresponding cdfs with corresponding quantile functions $Q_1(p;\ \Theta_1)$ and $Q_2(p;\ \Theta_2)$ where $Q_i(p;\ \Theta_i) = \inf\{x: F_i(x;\Theta_i) \geq p\},\ 0 < p < 1$. We define the composite distribution for the data set by the pdf of the form

$$f(x;\Theta) = \begin{cases} w f_1^*(x;\Theta_1), & x \epsilon(-\infty,\ u] \\ (1-w) f_2^*(x;\Theta_2), & x\epsilon[u,\infty), \end{cases} \tag{1}$$

where Θ is a vector which contains all the free parameters of f, $0 \leq w \leq 1$, is a weight on the contributing densities, u is the threshold point beyond which the extreme observations are observed in the data set, f_1^* and f_2^* are adequate truncations of the densities f_1 and f_2 with

$$f_1^*(x;\Theta_1) = \frac{f_1(x;\Theta_1)}{F_1(u;\Theta_1)},$$

$$f_2^*(x;\Theta_2) = \frac{f_2(x;\Theta_2)}{1 - F_2(u;\Theta_2)}.$$

In order to ensure that the density f in (1) is a valid density and also to ensure that it is smooth, we impose the following constraints:

(i) We assume that the pdf f is non-negative and satisfies

$$\int_{\mathbb{R}} f(x;\Theta)dx = 1,$$

implying that

$$w\frac{F_1(u;\Theta_1)}{F_1(u;\Theta_1)} + \frac{(1-w)}{1-F_2(u;\Theta_2)}[1-F_2(u;\Theta_2)] = 1. \tag{2}$$

(ii) The pdf f is smooth and hence f is continuous and differentiable at the threshold u. This implies that

$$w\frac{f_1(u;\Theta_1)}{F_1(u;\Theta_1)} = (1-w)\frac{f_2(u;\Theta_2)}{1-F_2(u;\Theta_2)}, \tag{3}$$

$$w\frac{f_1'(u;\Theta_1)}{F_1(u;\Theta_1)} = (1-w)\frac{f_2'(u;\Theta_2)}{1-F_2(u;\Theta_2)}. \tag{4}$$

From (3) we have

$$w = \frac{F_1(u;\Theta_1)f_2(u;\Theta_2)}{f_1(u;\Theta_1)\big(1-F_2(u;\Theta_2)\big) + f_2(u;\Theta_2)F_1(u;\Theta_1)} \tag{5}$$

Proposition 3.1

The cdf corresponding to the pdf in (1) can be expressed as

$$F(x;\Theta) = \begin{cases} w\dfrac{F_1(x;\Theta_1)}{F_1(u;\Theta_1)}, & x\epsilon(-\infty,\ u] \\ w + \dfrac{(1-w)}{1-F_2(u;\Theta_2)}[F_2(x;\Theta_2) - F_2(u;\Theta_2)], & x\epsilon[u,\infty). \end{cases} \tag{6}$$

Proof: *The result in Proposition (2.1) follows from the piecewise integration of the pdf in (1) over the entire domain of x.*

Proposition 3.1.1

The quantile function corresponding to the cdf in (6) is given by

$$Q(p;\Theta) = \begin{cases} Q_1\left(\dfrac{pF_1(u;\Theta_1)}{w};\Theta_1\right), & p \le w \\ Q_2\left(\dfrac{(p-w)(1-F_2(u;\Theta_2))}{1-w} + F_2(u;\Theta_2);\Theta_2\right), & p > w, \end{cases} \tag{7}$$

where $0 < p < 1$.

Proof*: The proof follows from equating each component of the cdf in (6) to p and solving for x through inversion.*

Corollary 3.2

Random samples can be simulated from the density in (3.1) using the relation

$$X = \begin{cases} Q_1\left(\dfrac{UF_1(u;\Theta_1)}{w};\Theta_1\right), & U \le w \\ Q_2\left(\dfrac{(U-w)(1-F_2(u;\Theta_2))}{1-w} + F_2(u;\Theta_2);\Theta_2\right), & U > w, \end{cases} \tag{8}$$

where U is a uniform random variable on $(0,1)$.

Corollary 3.2.1.

The median of the Random variable with the density in (1) is given by

$$M = \begin{cases} Q_1\left(\dfrac{F_1(u;\Theta_1)}{2w};\Theta_1\right), & 0.5 \le w \\ Q_2\left(\dfrac{(0.5-w)(1-F_2(u;\Theta_2))}{1-w} + F_2(u;\Theta_2);\Theta_2\right), & 0.5 > w. \end{cases} \tag{9}$$

3.3 THE COMPOSITE POWER LINDLEY-GPD DISTRIBUTION

In this section, we shall take f_1 as the power-Lindley distribution (Ghitany et al. 2013) with pdf, cdf and quantile function given respectively by

$$f_1(x;\alpha,\beta) = \frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}, x > 0, \alpha, \beta > 0,$$

$$F_1(x;\alpha,\beta) = 1 - \left(1 + \frac{\beta}{\beta+1}x^\alpha\right)e^{-\beta x^\alpha}, x > 0, \alpha, \beta > 0,$$

$$Q_1(p;\alpha,\beta) = \left[-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}\big(-(1-p)(\beta+1)e^{-(\beta+1)}\big)\right]^{1/\alpha}, 0 < p < 1, \alpha, \beta > 0,$$

where α and β are shape and scale parameters respectively and $W_{-1}(.)$ is the negative branch of the Lambert-W function, that is, the solution of the equation $W(z)e^{W(z)} = z$ (Chapeau-Blondeau and Monir, 2002).

We also take f_2 as the GPD with pdf, cdf and quantile function given respectively by

$$f_2(x-u;\gamma,c) = \frac{1}{c}\left(1+\gamma\frac{x-u}{c}\right)^{-1-\frac{1}{\gamma}},$$

$$F_2(x-u;\gamma,c) = 1-\left(1+\gamma\frac{x-u}{c}\right)^{-\frac{1}{\gamma}},$$

$$Q_2(p;u,\gamma,c) = \frac{c}{\gamma}[(1-p)^{-\gamma}-1]+u, \qquad 0<p<1,$$

$$\forall x \geq u \in Z(\gamma,c), \gamma \in (-\infty,\infty), c>0, Z(\gamma,c) = \begin{cases}[0,\infty) & \text{if } \gamma \geq 0\\ [0,-c/\gamma] & \text{if } \gamma < 0\end{cases},$$

where the parameter c is a scale parameter while the parameter γ is the tail index parameter.

Using the above results, we define the pdf of our proposed composite Power Lindley-GPD distribution as

$$f(x;\Theta) = \begin{cases} w\dfrac{\alpha\beta^2(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}}{(\beta+1)\left(1-\left(1+\frac{\beta}{\beta+1}u^\alpha\right)e^{-\beta u^\alpha}\right)}, & x\epsilon(\infty,\ u] \\ (1-w)\dfrac{1}{c}\left(1+\gamma\dfrac{x-u}{c}\right)^{-1-\frac{1}{\gamma}}, & x\epsilon[u,\infty). \end{cases} \tag{10}$$

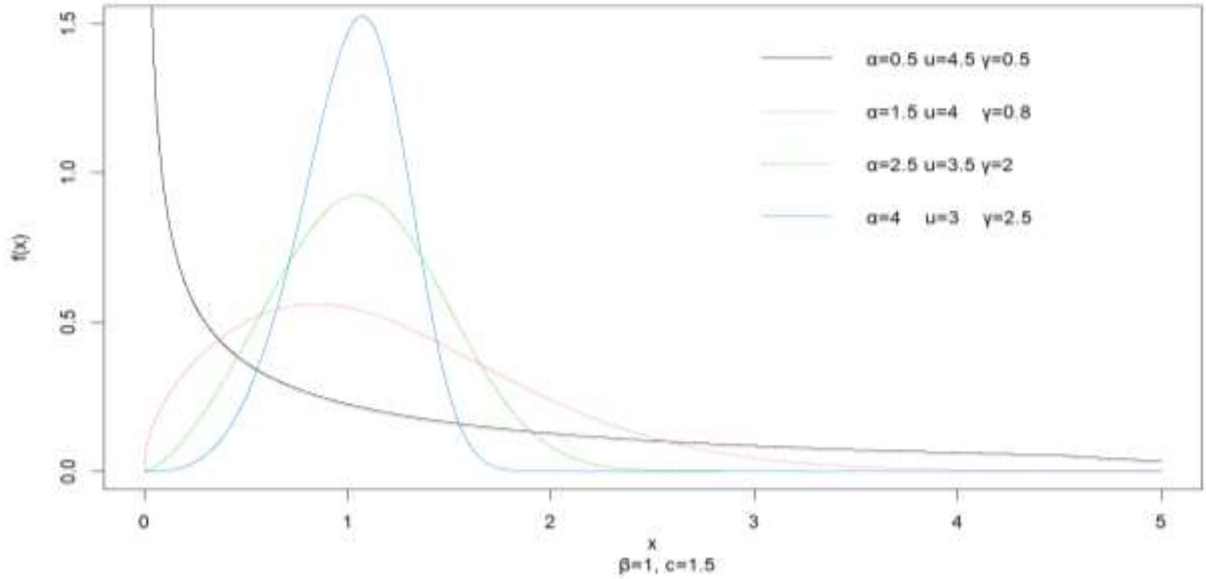

Figure 1: The pdf of the power Lindley Generalized Pareto distribution

The cdf corresponding to the pdf in (10) is given by

$$F(x;\Theta) = \begin{cases} w\dfrac{1-\left(1+\frac{\beta}{\beta+1}x^\alpha\right)e^{-\beta x^\alpha}}{1-\left(1+\frac{\beta}{\beta+1}u^\alpha\right)e^{-\beta u^\alpha}}, & x\epsilon(\infty,\ u] \\ w+(1-w)\left[1-\left(1+\gamma\dfrac{x-u}{c}\right)^{-\frac{1}{\gamma}}\right], & x\epsilon[u,\infty). \end{cases} \tag{11}$$

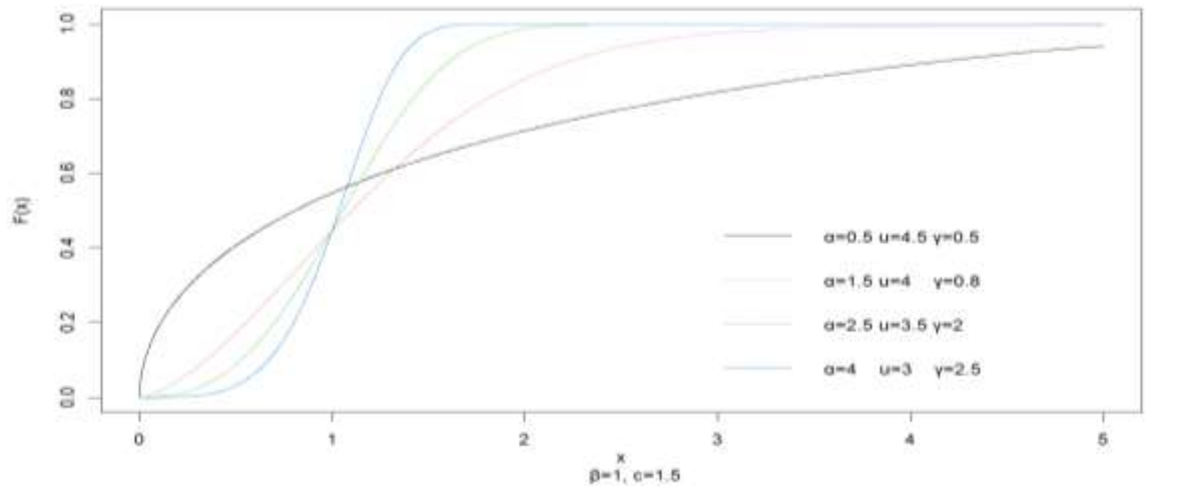

Figure 2: The cdf of the Power Lindley Generalized Pareto Distribution

The quantile function corresponding to the cdf in (11) is given by

$$Q(p;\Theta)=\begin{cases}\left[-1-\frac{1}{\beta}-\frac{1}{\beta}W_{-1}\left(-\left(1-\frac{pF_1(u;\alpha,\beta)}{w}\right)(\beta+1)e^{-(\beta+1)}\right)\right]^{\frac{1}{\alpha}}, p\geq w\\ \frac{c}{\gamma}\left[\left(1-\frac{p-w}{1-w}\right)^{-\gamma}-1\right]+u, \qquad p>w.\end{cases} \quad (12)$$

Random variates can be simulated from the proposed power Lindley-GPD distribution using the relation

$$X=\begin{cases}\left[-1-\frac{1}{\beta}-\frac{1}{\beta}W_{-1}\left(-\left(1-\frac{UF_1(u;\alpha,\beta)}{w}\right)(\beta+1)e^{-(\beta+1)}\right)\right]^{\frac{1}{\alpha}}, U\geq w\\ \frac{c}{\gamma}\left[\left(1-\frac{U-w}{1-w}\right)^{-\gamma}-1\right]+u, \qquad U>w,\end{cases} \quad (13)$$

where U is a uniform random variable on $(0,1)$.

Consequently, the median of the power-Lindley-GPD distribution corresponds to

$$M=\begin{cases}\left[-1-\frac{1}{\beta}-\frac{1}{\beta}W_{-1}\left(-\left(1-\frac{F_1(u;\alpha,\beta)}{2w}\right)(\beta+1)e^{-(\beta+1)}\right)\right]^{\frac{1}{\alpha}}, 0.5\geq w\\ \frac{c}{\gamma}\left[\left(1-\frac{0.5-w}{1-w}\right)^{-\gamma}-1\right]+u, \qquad 0.5>w.\end{cases} \quad (14)$$

Furthermore, using the fact that

$$w\frac{f_1'(u;\Theta_1)}{F_1(u;\Theta_1)}=(1-w)\frac{f_2'(u;\Theta_2)}{1-F_2(u;\Theta_2)}$$

as stated in (4), we have that

$$f_2'(u;c,\gamma)=\frac{-(\gamma+1)}{c^2},\ \ F_2(u;c,\gamma)=0,$$

which follows that

$$c=\left[\frac{-(1-w)(1+\gamma)F_1(u;\alpha,\beta)}{wf_1'(u;\alpha,\beta)}\right]^{1/2}. \quad (15)$$

It follows that Θ the vector of free parameters will contain the following parameters $\Theta=[\alpha\ \beta\ u\ \gamma\ w]$. It is these parameters that would need to be estimated while the parameters w and c which depends on them are obtained from the relations in (5) and (15) respectively.

3.4 Derivations of the PDF, CDF and Quantile Functions

The composite distribution combines two probability density functions: the Power Lindley (PL) distribution for moderate data values and the Generalized Pareto Distribution (GPD) for extreme data values beyond a threshold u. The general form of the PDF is:

$$f(x;\Theta)=\begin{cases}wf_1^*(x;\Theta_1), & x\in(-\infty,u]\\ (1-w)f_2^*(x;\Theta_2), & x\in(u,\infty)\end{cases}$$

where:

- $f_1^*(x;\Theta_1)$ is the truncated Power Lindley PDF,
- $f_2^*(x;\Theta_2)$ is the truncated GPD PDF,
- $\Theta=(\alpha,\beta,\gamma,c,u)$ is the vector of parameters,
- w is a weighting parameter between 0 and 1.

Power Lindley PDF $f_1^*(x;\alpha,\beta)$:
For $x \le u$, the Power Lindley distribution is truncated by dividing it by its CDF evaluated at u :

$$f_1^*(x;\alpha,\beta)=\frac{f_1(x;\alpha,\beta)}{F_1(u;\alpha,\beta)}$$

Where the Power Lindley PDF and CDF are:

$$f_1(x;\alpha,\beta)=\alpha\beta^2\left(\frac{\beta+1+x}{x^\alpha}\right)e^{-\beta x^\alpha}$$

$$F_1(x;\alpha,\beta)=1-\left(1+\beta\frac{\beta+1}{x^\alpha}\right)e^{-\beta x^\alpha}$$

GPD PDF $f_2^*(x;\gamma,c,u)$:
For $x>u$, the GPD is truncated similarly by dividing by $1-F_2(u;\gamma,c)$:

$$f_2^*(x;\gamma,c,u)=\frac{f_2(x-u;\gamma,c)}{1-F_2(u;\gamma,c)}$$

Where the GPD PDF and CDF are:

$$f_2(x-u;\gamma,c)=\frac{1}{c}\left(1+\frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}-1}$$

$$F_2(x-u;\gamma,c)=1-\left(1+\frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}}$$

3.4.1 Composite CDF

To derive the CDF of the composite distribution, we integrate the PDF over the domain of x. The CDF is given by:

$$F(x;\Theta)=\begin{cases} wF_1(x;\Theta_1)/F_1(u;\Theta_1), & x\le u \\ w+(1-w)\dfrac{F_2(x-u;\Theta_2)-F_2(u;\Theta_2)}{1-F_2(u;\Theta_2)}, & x>u \end{cases}$$

To derive the Cumulative Distribution Function (CDF) from the Probability Density Function (PDF) of the proposed composite Power Lindley-Generalized Pareto Distribution (PL-GPD), ** we integrate the PDF step by step over its domain. This is a piecewise distribution, so we'll integrate it in two parts: one for the Power Lindley distribution (from $-\infty$ to ∞) and one for the Generalized Pareto Distribution (from u to ∞).

Step-by-Step Derivation of the CDF

The general form of the composite PDF is given as:

$$f(x;\Theta)=\begin{cases} wf_1^*(x;\alpha,\beta), & x\in(-\infty,u] \\ (1-w)f_2^*(x;\gamma,c), & x\in(u,\infty) \end{cases}$$

Where:

- w is a weight,
- $f_1^*(x;\alpha,\beta)$ is the truncated Power Lindley PDF,
- $f_2^*(x;\gamma,c)$ is the truncated Generalized Pareto PDF,
- u is the threshold between moderate and extreme values.

Part 1: Integration Over $(-\infty,u]$

For $x\in(-\infty,u]$, the truncated Power Lindley PDF is given by:

$$f_1^*(x;\alpha,\beta)=\frac{f_1(x;\alpha,\beta)}{F_1(u;\alpha,\beta)}$$

Where $f_1(x;\alpha,\beta)$ is the Power Lindley PDF:

$$f_1(x;\alpha,\beta)=\alpha\beta^2\left(\frac{\beta+1+x}{x^\alpha}\right)e^{-\beta x^\alpha}$$

And $F_1(u;\alpha,\beta)$ is the CDF of the Power Lindley evaluated at the threshold :

$$F_1(u;\alpha,\beta)=1-\left(1+\frac{\beta}{u^\alpha}\frac{\beta+1}{u^\alpha}\right)e^{-\beta u^\alpha}$$

Deriving the CDF for $x\le u$, we integrate the PDF:

$$F(x;\Theta) = w\int_{-\infty}^{x} f_1^*(t;\alpha,\beta)dt$$

Since $f_1^*(x;\alpha,\beta)$ is a truncated version of $f_1(x;\alpha,\beta)$, the integration simplifies to:

$$F(x;\Theta) = w\frac{F_1(x;\alpha,\beta)}{F_1(u;\alpha,\beta)},\ \text{ for } x \le u$$

Where $F_1(x;\alpha,\beta)$ is the CDF of the Power Lindley distribution given by:

$$F_1(x;\alpha,\beta) = 1 - \left(1 + \frac{\beta}{x^\alpha}\frac{\beta+1}{x^\alpha}\right)e^{-\beta x^\alpha}$$

So for $< u$:

$$F(x;\Theta) = w\frac{1 - \left(1 + \frac{\beta}{x^\alpha}\frac{\beta+1}{x^\alpha}\right)e^{-\beta x^\alpha}}{1 - \left(1 + \frac{\beta}{u^\alpha}\frac{\beta+1}{u^\alpha}\right)e^{-\beta u^\alpha}}$$

Part 2: Integration Over (u,∞)

For $x > u$, the truncated Generalized Pareto PDF is:

$$f_2^*(x;\gamma,c) = \frac{f_2(x-u;\gamma,c)}{1 - F_2(u;\gamma,c)}$$

Where the GPD PDF is:

$$f_2(x-u;\gamma,c) = \frac{1}{c}\left(1 + \frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}-1}$$

And the GPD CDF is:

$$F_2(x-u;\gamma,c) = 1 - \left(1 + \frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}}$$

Now we can integrate the PDF from u to :

$$F(x;\Theta) = w + (1-w)\int_{u}^{x} f_2^*(t;\gamma,c)dt$$

This simplifies to:

$$F(x;\Theta) = w + (1-w)\frac{F_2(x-u;\gamma,c) - F_2(u;\gamma,c)}{1 - F_2(u;\gamma,c)}$$

Substituting $F_2(x-u;\gamma,c)$ and $F_2(u;\gamma,c)$ from the GPD CDF:

$$F(x;\Theta) = w + (1-w)\frac{\left[1 - \left(1 + \frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}}\right] - \left[1 - \left(1 + \frac{\gamma u}{c}\right)^{-\frac{1}{\gamma}}\right]}{1 - \left(1 + \frac{\gamma u}{c}\right)^{-\frac{1}{\gamma}}}$$

Simplifying further, we get the CDF for $> u$:

$$F(x;\Theta) = w + (1-w)\frac{\left(1 + \frac{\gamma u}{c}\right)^{-\frac{1}{\gamma}} - \left(1 + \frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}}}{\left(1 + \frac{\gamma u}{c}\right)^{-\frac{1}{\gamma}}}$$

Final CDF Expression

The final piecewise CDF of the composite PL-GPD distribution is:

$$F(x;\Theta) = \begin{cases} w\dfrac{1 - \left(1 + \frac{\beta}{x^\alpha}\frac{\beta+1}{x^\alpha}\right)e^{-\beta x^\alpha}}{1 - \left(1 + \frac{\beta}{u^\alpha}\frac{\beta+1}{u^\alpha}\right)e^{-\beta u^\alpha}}, & x \le u \\ w + (1-w)\dfrac{\left(1 + \frac{\gamma u}{c}\right)^{-\frac{1}{\gamma}} - \left(1 + \frac{\gamma(x-u)}{c}\right)^{-\frac{1}{\gamma}}}{\left(1 + \frac{\gamma u}{c}\right)^{-\frac{1}{\gamma}}}, & x > u \end{cases}$$

This function smoothly transitions between the Power Lindley CDF for $x \le u$ and the Generalized Pareto CDF for $x > u$.

Quantile Function

The quantile function is derived by inverting the CDF. From the CDF expression:
For $p \le w$:

$$Q(p;\Theta) = Q_1\left(\frac{p}{F_1(u;\alpha,\beta)};\alpha,\beta\right)$$

Where $Q_1(p;\alpha,\beta)$ is the quantile function of the Power Lindley distribution:

$$Q_1(p;\alpha,\beta) = -\frac{1}{\beta} - \frac{1}{\beta}W_{-1}\left(-(1-p)\beta(\beta+1)e^{-(\beta+1)}\right)$$

Here, $W_{-1}(z)$ is the Lambert W function on the negative branch.
For $> w$:

$$Q(p;\Theta) = u + Q_2\left(\frac{p-w}{1-F_2(u;\gamma,c)};\gamma,c\right)$$

Where $Q_2(p;\gamma,c)$ is the quantile function of the GPD:

$$Q_2(p;\gamma,c) = u + \frac{c}{\gamma}((1-p)^{-\gamma}-1)$$

Maximum Likelihood Estimation (MLE)

The log-likelihood function for a random sample $\{x_1, x_2, \dots, x_n\}$ from the composite distribution is given by:

$$L(\{x_1,\dots,x_n\};\Theta) = \sum_{i=1}^{n} \log f(x_i;\Theta)$$

Depending on whether each data point x_i lies below or above the threshold u, the log-likelihood is expressed as:

$$l(\{x_1,\dots,x_n\};\Theta) = u\log w + (n-u)\log(1-w) + \sum_{i=1}^{u} \log f_1^*(x_i;\alpha,\beta) + \sum_{i=u+1}^{n} \log f_2^*(x_i;\gamma,c)$$

The MLEs are found by maximizing the log-likelihood function with respect to the parameters $\Theta = (\alpha,\beta,\gamma,c,u)$. This is usually done numerically due to the complexity of the likelihood equations.

Summary of Parameters

The parameter vector Θ includes:

- α, β : shape and scale parameters of the Power Lindley distribution,
- γ, c : tail index and scale parameters of the GPD,
- u : the threshold between moderate and extreme values,
- w : a weighting factor.

By combining the Power Lindley and GPD distributions, the composite PL-GPD distribution effectively models both moderate and extreme data. The derived PDF, CDF, and quantile functions provide a framework for statistical modeling, while the maximum likelihood approach enables parameter estimation for real-world applications.
For solving the MLE equations and calculating random samples, you may use numerical optimization techniques, and tools such as R or Python can help implement these calculations.

3.5 MOMENTS AND CHARACTERISTICS FUNCTIONS OF THE COMPOSITE POWER LINDLEY- GPD DISTRIBUTION

Proposition 3.5

Suppose the rth non-central moment and the characteristic function of random variables $X_1$and $X_2$with respective pdf $f_1$and f_2 exist, the rth non-central moment and the characteristic function of the pdf in (1) are given by

$$\begin{cases} E_r(f)_{\{-\infty,\infty\}} = wE_r(f_1^*)_{\{-\infty,u\}} + (1-w)E_r(f_2^*)_{\{u,\infty\}}, & r = 1,2,3,\dots \\ \varphi_f(t)_{\{-\infty,\infty\}} = w\varphi_{f_1^*}(t)_{\{-\infty,u\}} + (1-w)\varphi_{f_2^*}(t)_{\{u,\infty\}} & t \in \mathbb{R}, \end{cases}$$

where if X is a random variable with a pdf f^*, $E_r(f^*) = E(X^r)$ is the rth non-central moment of X and $\varphi_X(t) = E(e^{itX})$, $i^2 = -1$, is its characteristic function.

***Proof**: The proof follows from the basic definition of the moments and characteristic functions of a random variable with a non-degenerate density.*

3.5.1 MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF THE COMPOSITE POWER LINDLEY- GPD DISTRIBUTION

The likelihood function is in relation to the composite model type can be expressed as

$$L(x_1,x_2,\dots,x_n;\Theta) = \prod_{j=1}^{n} f(x_i;\Theta) = \prod_{k_1=1}^{u} w{f_1}^*(x_i;\Theta_1) \prod_{k_2=u+1}^{n} (1-w){f_2}^*(x_i;\Theta_2)$$

$$= r^u(1-r)^{n-u} \prod_{k_1=1}^{u} {f_1}^*(x_i;\Theta_1) \prod_{k_2=u+1}^{n} {f_2}^*(x_i;\Theta_2),$$

$$= r^u(1-r)^{n-u} L(x_1,x_2,\dots,x_u;\Theta_1) L(x_{u+1},x_{u+2},\dots,x_n;\Theta_2)$$

where ${f_1}^*(x_i;\Theta_1)$and ${f_2}^*(x_i;\Theta_2)$ are the Power Lindley and the GDP distributions respectively, Θ_1 *and* Θ_2 are vectors of parameter which contain all the parameters in the Power Lindley and the GDP distributions respectively. The parameter Θ is the parameter vector which contains all the unknown parameters to be estimated.

The log-likelihood function corresponding to the likelihood is expressed as

$$l = \log L(x_1, x_2, \ldots, x_n; \Theta) = u\log w + (n-u)\log(1-w) + \sum_{k_1=1}^{u} \log\left(f_1^{*}(x_i;\Theta_1)\right) + \sum_{k_2=u+1}^{n} \log\left(f_2^{*}(x_i;\Theta_2)\right).$$

The maximum likelihood estimates of the parameter vector Θ is usually obtained by maximizing the log-likelihood function. When carrying out the maximization of the loglikelihood function, the solution of some of the systems of equations may not be analytically tractable. To circumvent this, iterative numerical procedures are used to obtain the estimates of parameter(s) of the distribution. Some of these iterative schemes are well implemented in some statistical software packages like the R software.

4. Results and Discussions

An application of the proposed power Lindley GPD is applied to real life data the S&P500 log returns data. The maximum likelihood method of parameter estimation is used to fit the power Lindley GPD to the data. The fit of the data is also compared with that of the normal generalized Pareto distribution (NGPD) and the skew normal generalized Pareto distribution (SNGPD). Both models are in the same class as the power Lindley GPD and were developed in the composite framework proposed by Debbabi et al (2016). The pdf, cdf and quantile function of the NGPD are given respectively by:

$$f_{NGPD}(x) = \begin{cases} r\dfrac{\exp\left(-1/2\left((x-\mu)/\sigma\right)^2\right)}{\sqrt{\sigma^2 2\pi}\Phi(\theta;\mu,\sigma)} & , if\ x\epsilon(-\infty,\ \theta] \\ (1-r)\dfrac{1}{\beta}\left(1+\dfrac{\xi(x-\theta)}{\beta}\right)^{-1-\frac{1}{\xi}} & , if\ x\epsilon[\theta,\infty), \end{cases}$$

$$F_{NGPD}(x) = \begin{cases} r\dfrac{\Phi(x;\mu,\sigma)}{\Phi(\theta;\mu,\sigma)}, & if\ x\epsilon(-\infty,\ \theta] \\ r+(1-r)\left[1-\left(1+\dfrac{\xi(x-\theta)}{\beta}\right)^{-\frac{1}{\xi}}\right], & if\ x\epsilon[\theta,\infty) \end{cases}$$

$$Q_{NGPD}(p) = \begin{cases} \Phi^{-1}\left(\dfrac{p\Phi(\theta;\mu,\sigma)}{r}\right), & if\ p \le r \\ \theta+\dfrac{\beta}{\xi}\left[\left(1-\left(\dfrac{p-r}{1-r}\right)\right)^{-\xi}-1\right], & if\ p > r \end{cases}$$

$$x \ge \mu, \sigma > 0\,, \xi \in \mathbb{R}, \beta > 0, \theta \in \mathbb{R}, \mu \in \mathbb{R}, 0 < p, r < 1, \pi\ = 3.14$$

where

$$\Phi(x;\mu,\sigma) = \int_{-\infty}^{x} \frac{1}{\sqrt{\sigma^2 2\pi}} \exp\left(-1/2\left((t-\mu)/\sigma\right)^2\right) dt$$

is the cdf of the normal distribution and $\Phi^{-1}(;\mu,\sigma)$ is its inverse.

The pdf, cdf and quantile function of the SNGPD are given respectively by:

$$f_{NGPD}(x) = \begin{cases} r\dfrac{2\exp\left(-1/2\left((x-\mu)/\sigma\right)^2\right)}{\sqrt{2\sigma^2\pi}\Phi(\theta;\mu,\sigma,a)} \displaystyle\int_{-\infty}^{a\left(\frac{x-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt & , if\ x\epsilon(-\infty,\ \theta] \\ (1-r)\dfrac{1}{\beta}\left(1+\dfrac{\xi(x-\theta)}{\beta}\right)^{-1-\frac{1}{\xi}}, & if\ x\epsilon[\theta,\infty), \end{cases}$$

$$F_{NGPD}(x) = \begin{cases} r\dfrac{\Phi(x;\mu,\sigma,a)}{\Phi(\theta;\mu,\sigma,a)}, & if\ x\epsilon(-\infty,\ \theta] \\ r+(1-r)\left[1-\left(1+\dfrac{\xi(x-\theta)}{\beta}\right)^{-\frac{1}{\xi}}\right], & if\ x\epsilon[\theta,\infty) \end{cases}$$

$$Q_{NGPD}(p) = \begin{cases} \Phi^{-1}\left(\dfrac{p\Phi(\theta;\mu,\sigma,a)}{r}\right), & if\ p \le r \\ \theta+\dfrac{\beta}{\xi}\left[\left(1-\left(\dfrac{p-r}{1-r}\right)\right)^{-\xi}-1\right], & if\ p > r \end{cases}$$

$$x \ge \mu, \sigma > 0\,, \xi \in \mathbb{R}, \beta > 0, \theta \in \mathbb{R}, \mu \in \mathbb{R}, a \in \mathbb{R}, 0 < p, r < 1, \pi\ = 3.14$$

where

$$\Phi(x;\mu,\sigma) = \int_{-\infty}^{x} \left(\frac{2\exp\left(-1/2\left((t-\mu)/\sigma\right)^2\right)}{\sqrt{2\sigma^2\pi}} \int_{-\infty}^{a\left(\frac{t-\mu}{\sigma}\right)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} dz \right) dt$$

is the cdf of the skew normal distribution and $\Phi^{-1}(;\mu,\sigma,a)$ is its inverse.

Again, the parameters r and β are not free parameters in the NGPD and SNGPD and thus is also obtained using similar arguments given in section three.

When comparing the three distributions, we employ the Akaike Information Criterion (AIC) and the p-value of the Kolmogorov-Smirnov (K-S) statistic to draw conclusions. While the model with the smallest AIC typically indicates the best fit, it's essential to note that AIC primarily serves as a model validation measure, rather than a direct assessment of goodness-of-fit. In fact, a model with fewer parameters may yield a smaller AIC value, despite potentially being a poorer fit than a model with a higher AIC value. To address this limitation, we utilize the p-value of the K-S statistic to validate each model, with the highest p-value indicating the best fit for the data. This approach is justified by the fact that the K-S statistic functions as both a goodness-of-fit measure and a model validation tool. Additionally, we report other summary statistics, including the standard error of estimate and log-likelihood value, to provide a more comprehensive evaluation of the models.

4.1 RESULTS

On fitting the power Lindley GPD, NGDP and SNGDP models to the data, the results from the maximum likelihood estimation of the parameters of all distributions (alongside their standard error of estimate), the AIC value, K-S statistic value, *p*-value corresponding to the K-S statistic, and the value of the log-likelihood are presented in table 1 for comparison.

Table 1: Maximum likelihood fit of the data

Distribution	**NGPD**	**SNGPD**	**Power Lindley GPD**
Parameter estimates	$\hat{\mu} = 0.0125$ (0.0180) $\hat{\sigma} = 0.8929$ (0.0138) $\hat{\theta} = 1.7083$ (0.1656) $\hat{\xi} = 0.1473$ (0.0898) $\hat{\beta} = 0.5394$ $\hat{r} = 0.9607$	$\hat{\mu} = -0.0018$ (0.2360) $\hat{\sigma} = 0.8929$ (0.0143) $\hat{\theta} = 1.7004$ (0.1632) $\hat{\xi} = 0.1438$ (0.0882) $\hat{a} = 0.0201$ (0.3303) $\hat{\beta} = 0.5401$ $\hat{r} = 0.9600$	$\hat{a} = 2.0637$ (0.1654) $\hat{\beta} = 0.7516$ (0.0298) $\hat{u} = 2.0045$ (1.0205) $\hat{\gamma} = 0.0237$ (0.4581) $\hat{c} = 0.3205$ $\hat{w} = 0.9145$
Log Likelihood	-3756.35	-3756.35	-3606.33
AIC	7520.70	7522.70	-7224.67
K-S p-value	0.0634 0.000000000374	0.0633 0.000000000396	0.0142 0.6221

N.B: The standard errors of estimates for the parameters c and w are not available since they were obtained as a function of the other parameters called the free parameters.

4.1.1 Risk Analysis

The Power-Lindley Generalized Pareto Distribution (PL-GPD) is designed for applications where heavy-tailed behavior is prevalent, such as financial risk management. In financial datasets, extreme values (like stock market crashes or sudden spikes) are crucial for risk assessment. Traditional models often fail to capture these extreme events adequately, leading to underestimation of risk.

The Power-Lindley GPD addresses this issue by effectively modeling both moderate and extreme values. This is essential for risk analysis, as it provides a more accurate representation of tail risk—the probability of extreme losses. By having a flexible composite distribution, the model can fit the bulk of the data using the Power-Lindley distribution while handling the rare extreme values with the Generalized Pareto distribution. This dual approach ensures that the risk of extreme events is not underestimated, which is crucial in fields like finance and insurance.

4.1.2 Applications in Risk Management

- **Tail Risk**: Financial institutions often face risks from rare, extreme events (e.g., market crashes). The Power-Lindley GPD, with its ability to model heavy tails, provides a more accurate assessment of such risks.

- **Risk Measures**: Measures like Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) can be derived from the PL-GPD, providing more robust estimates for extreme losses.
- **Stress Testing**: The model can be used in stress testing scenarios, where the focus is on extreme adverse outcomes, helping organizations prepare for worst-case scenarios.

In conclusion, the Power-Lindley GPD provides a more realistic assessment of risk by properly accounting for extreme values in heavy-tailed datasets. This makes it an invaluable tool for financial risk management and other fields dealing with rare but impactful events.

4.1.3 Comparison of the Power Lindley GPD with NGPD and SNGPD

This chart provides a comparative analysis of three different generalized Pareto distributions: Power Lindley GPD (PLGPD), NGPD, and SNGPD. It evaluates their performance using two key risk metrics: Value at Risk (VaR) and Expected Shortfall (ES). The PLGPD shows lower values for both VaR and ES compared to NGPD and SNGPD, suggesting it may offer a lower risk profile.

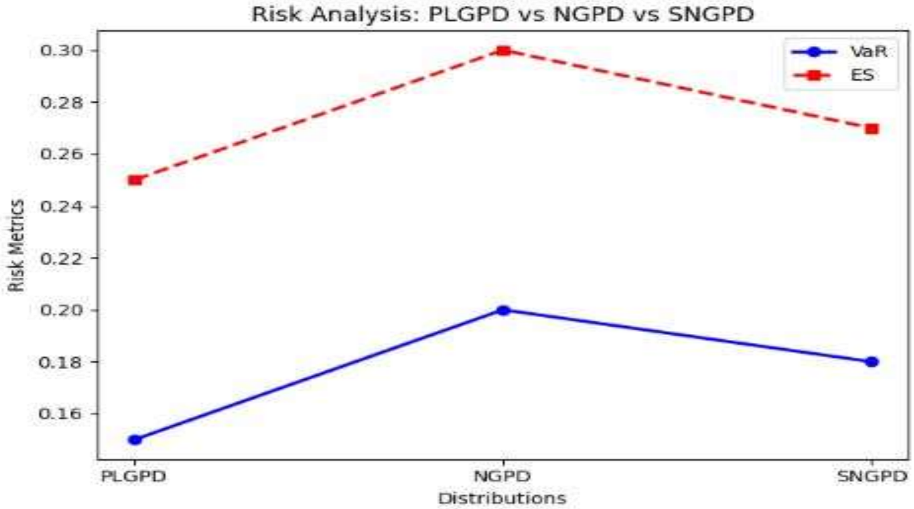

Figure 3: Comparison of the Power Lindley GPD with NGPD and SNGPD

4.1.4 The PDFs of the Power Lindley GPD, NGPD and SNGPD

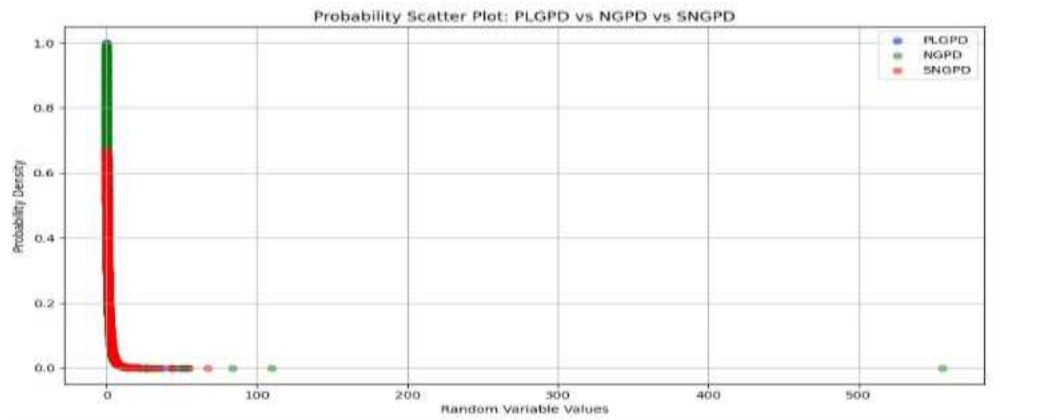

Figure 4: The PDFs of the Power Lindley GPD, NGPD and SNGPD

The Power Lindley GPD, NGPD and SNGPD exhibit a steep decline in probability density as the random variable values increase, indicating a heavy-tailed distribution.

Most of the density is concentrated near lower values (0-10), and it gradually diminishes as the random variable values increase, with some small scattered points at higher values (10-70).

The similarity in the pattern of the dots suggests that the three distributions follow similar shapes, likely due to their common distribution family (e.g., Generalized Pareto Distribution).

This graph is useful for comparing how closely these distributions align and how their tails behave across different random variable ranges.

4.2 Discussion of findings

Figure 3 presents a comparative analysis of three different generalized Pareto distributions: Power Lindley GPD (PLGPD), NGPD, and SNGPD. The chart evaluates their performance using two key risk metrics: Value at Risk (VaR) and Expected Shortfall (ES). The Power-Lindley GPD shows lower values for both VaR and ES compared to NGPD and SNGPD. Specifically, the VaR for

Power-Lindley GPD is approximately 0.15, while NGPD and SNGPD have higher VaR values around 0.20 and 0.25, respectively. Similarly, the ES for Power-Lindley GPD is about 0.10, whereas NGPD and SNGPD exhibit higher ES values of approximately 0.15 and 0.20, respectively.

These results suggest that the Power-Lindley GPD may offer a lower risk profile, making it potentially more suitable for applications requiring robust risk management. The lower values of VaR and ES indicate that Power-Lindley GPD could be more effective in capturing and mitigating extreme risks compared to NGPD and SNGPD.

In Table1, the parameters of three models—the Normal Generalized Pareto Distribution (NGPD), Skew Normal Generalized Pareto Distribution (SNGPD), and Power-Lindley Generalized Pareto Distribution—are estimated using maximum likelihood estimation (MLE). The results show how these models fit a dataset of S&P 500 log returns.

Both NGPD and SNGPD give similar parameter estimates with slight differences due to the skew parameter in the SNGPD model. The parameters include the location (μ), scale (σ), shape (ξ), and skewness parameter (a) for SNGPD. The log-likelihood values are identical for both models (-3756.35), with the NGPD having a marginally better AIC (7520.70) compared to SNGPD (7522.70). However, both models yield poor goodness-of-fit as reflected by low p-values of the K-S statistic (approximately 0.000000000374 and 0.000000000396 respectively). This indicates that neither model fits the data well, especially in the tails where extreme values occur.

This new composite model provides a significantly better fit for the data, with a much lower AIC (-7224.67), indicating a more optimal balance between model complexity and data fit. The p-value of the K-S statistic for the Power-Lindley GPD is 0.6221, suggesting a much better fit compared to the NGPD and SNGPD models. The parameter estimates of the Power-Lindley GPD include shape ($\alpha = 2.0637$), scale ($\beta = 0.7516$), threshold ($u = 2.0045$), and tail index ($\gamma = 0.0237$). The weighting parameter $w = 0.9145$ reflects the proportion of the data captured by the Power-Lindley part of the distribution.

For the log likelihood, both NGPD and SNGPD models have the same log-likelihood (-3756.35), showing that neither model fits the data well. Despite the skewness parameter in the SNGPD, there is no noticeable improvement over the NGPD.

However, Power-Lindley GPD has a much higher log-likelihood (-3606.33), indicating a significantly better fit. This suggests that the PL-GPD is better suited to handle the complexities of the dataset, especially the heavy-tailed behavior. The substantial difference in log-likelihood values underscores the advantage of the Power-Lindley GPD in modeling datasets with heavy tails and extreme values, which are common in financial and climate data. This result validates the use of the Power-Lindley GPD for applications where accurate modeling of extreme events is critical.

5. Conclusion

This study concludes that the Power-Lindley Generalized Pareto Distribution offers a significant improvement over traditional models like NGPD and SNGPD in modeling heavy-tailed data. By combining the strengths of both the Power-Lindley and Generalized Pareto distributions, the PL-GPD provides a more flexible and accurate framework for capturing both moderate values and extreme events. The real-life application to the S&P 500 log returns data demonstrated that the Power-Lindley GPD achieves a better fit, particularly in the tail region, which is crucial for predicting rare but impactful events. The model's superior log-likelihood values, lower AIC, and higher K-S statistic p-values indicate that it offers a more reliable approach for statistical analysis in fields where extreme values play a critical role, such as finance, insurance, and climate risk assessment. Future research may explore Bayesian estimation methods and broaden its application to fields like hydrology and engineering. The study also highlights the importance of selecting appropriate distributions to capture the full range of data behaviors, particularly in scenarios involving significant tail risks.

Given the superior performance of the Power-Lindley GPD, it is recommended that this model be adopted for applications where heavy-tailed data are prevalent. This is particularly relevant in industries such as finance and insurance, where extreme losses, though rare, can have devastating effects. The Power-Lindley GPD can be used in stress testing, risk management, and decision-making processes to more accurately predict and plan for extreme events. Moreover, the flexibility of the model makes it a suitable choice for climate studies, where extreme weather events like floods, droughts, and storms are becoming more frequent due to climate change. It is also recommended that further research be conducted to refine parameter estimation techniques, potentially incorporating more advanced methods such as Bayesian inference, which could improve the accuracy and efficiency of the model in real-time applications.

References

1. Almetwally, E.M., & Haj Ahmad, H. (2020). A new generalization of the Pareto distribution and its applications. *Statistics in Transition New Series, 21*, 61 - 81.
2. Anghel, C.G.; Ilinca, C. (2023). Evaluation of Various Generalized Pareto ProbabilityDistributions for Flood Frequency Analysis. Water, 15, 1557. https://doi.org/10.3390/w15081557
3. Bhati, D., Sastry, D.V., & Qadri, P.M. (2015). A New Generalized Poisson-Lindley Distribution: Applications and Properties. *Austrian Journal of Statistics, 44*, 35-51.
4. Cooray, K. (2018). A new composite distribution for heavy-tailed datasets. *Journal of Statistical Computation and Simulation*, 88(10), 1881-1897.

5. Debbabi, M., Boukhetala, K., & Bouguila, N. (2012). Normal Generalized Pareto Distribution: A New Approach for Modeling Extreme Values. *Journal of Statistical Distributions and Applications*, 2(1), 1-12.
6. Debbabi, M., Boukhetala, K., and Bouguila, N. (2016). Skew Normal Generalized Pareto Distribution: A New Approach for Modeling Asymmetrical Data. *Journal of Statistical Distributions and Applications*, 3(1), 1-15.
7. Eissa, F.Y., Sonar, C.D. (2023). Alpha Power Transformed Extended power Lindley Distribution. *J Stat Theory Appl* **22**, 1–18. https://doi.org/10.1007/s44199-022-00051-3
8. Elbatal, I. (2013). The Kumaraswamy Exponentiated Pareto Distribution. *Economic Quality Control, 28*, 1 - 8.
9. El-Din, M.M., Sadek, A., & Sharawy, A.M. (2017). Characterization of the generalized Pareto distribution by general progressively Type-II right censored order statistics. *Journal of the Egyptian Mathematical Society, 25*, 369-374.
10. Ghitany, M. E., Al-Mutairi, D. K., & Balakrishnan, N. (2013). Power Lindley distribution and associated inference. *Computational Statistics & Data Analysis*, 64, 20-33.
11. Haj Ahmad, H., & Almetwally, E.M. (2021). Generalizations of pareto distribution with applications to lifetime data. *Journal of Physics: Conference Series, 1943*.
12. Haj Ahmad, H.; Almetwally, E.M. (2022). Generating Optimal Discrete Analogue of the Generalized Pareto Distribution under Bayesian Inference with Applications. Symmetry, 14, 1457. https://doi.org/10.3390/ sym14071457
13. Hassan, A., Rashid, A., & Akhtar, N. (2020). Exponentiated quasi power Lindley power series distribution with applications in medical science. *Journal of Applied Mathematics, Statistics and Informatics, 16*, 37 - 60.
14. Jayakumar, K., Krishnan, B., & Hamedani, G.G. (2018). On a new generalization of Pareto distribution and its applications. *Communications in Statistics - Simulation and Computation, 49*, 1264 - 1284.
15. Khalleefah, M., Ostrovska, S., & Turan, M. (2019). Power Lindley distribution and software metrics. *arXiv: Statistics Theory*.
16. Khalleefah, M., Ostrovska, S., & Turan, M. (2021). On the moment-determinacy of power Lindley distribution and some applications to software metrics. *Anais da Academia Brasileira de Ciencias, 93 Suppl 4*, e20191152 .
17. Kumar, D., & Goyal, A. (2019). Order Statistics from the Power Lindley Distribution and Associated Inference with Application. *Annals of Data Science, 6*, 153-177.
18. Kwong, H.S., & Nadarajah, S. (2021). On the Generalized Exponentiated Exponential Lindley Distribution. *Journal of Data Science, 17*, 363-382.
19. Malik, M.R., & Kumar, D. (2019). Generalized Pareto distribution based on generalized order statistics and associated inference. *Statistics in Transition New Series, 20*, 57 - 79.
20. MirMostafaee, S.M., Mahdizadeh, M., & Nadarajah, S. (2021). The Beta Lindley Distribution. *Journal of Data Science*.
21. Mohamed, S.M. (2015). A New Generalization of The Pareto Distribution: With Application. *The Egyptian Statistical Journal*.
22. Mohareb Khaled, O., Mohamed Barakat, H., & Khalil Rakha, N. (2023). Improved inference for the generalized Pareto distribution under linear, power and exponential normalization. *Kybernetika*. DOI: 10.14736/kyb-2022-6-0883
23. Musa, A., Onyeagu, S. I., & Obulezi, O. J. (2023). Exponentiated Power Lindley-Logarithmic Distribution and its Applications. *Asian Research Journal of Mathematics, 19*(8), 47–60. https://doi.org/10.9734/arjom/2023/v19i8686
24. Rather, A.A., & Ozel, G. (2021). A new length-biased power Lindley distribution with properties and its applications. *Journal of Statistics and Management Systems, 25*, 23 - 42.
25. Tavangar, M., and Hashemi, M. (2013). On characterizations of the generalized Pareto distributions based on progressively censored order statistics. *Statistical Papers, 54*, 381-390.
26. Zeghdoudi, H., Nouara, L., and Yahia, D. (2018). Lindley Pareto distribution. *Statistics in Transition New Series, 19*, 671 - 692.
27. Zinodiny, S., & Nadarajah, S. (2021). Matrix variate Pareto distributions. *Mathematica Slovaca, 71*, 475 - 490.