

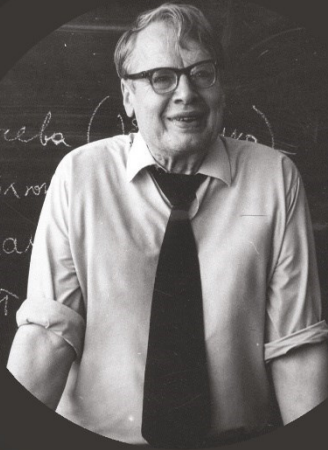
RTA

ISSN 1932-2321

JOURNAL IS REGISTERED
IN THE LIBRARY OF THE
U.S. CONGRESS

**RELIABILITY:
THEORY & APPLICATIONS**

INTERNATIONAL
GROUP ON
RELIABILITY



GNEDENKO FORUM PUBLICATIONS

#3

(92) VOL.21

SEPTEMBER
2026

SAN DIEGO

RELIABILITY

RISK ANALYSIS

MAINTENANCE

SAFETY

ISSN 1932-2321

© "Reliability: Theory & Applications", 2006, 2007, 2009-2026

© " Reliability & Risk Analysis: Theory & Applications", 2008

© I.A. Ushakov

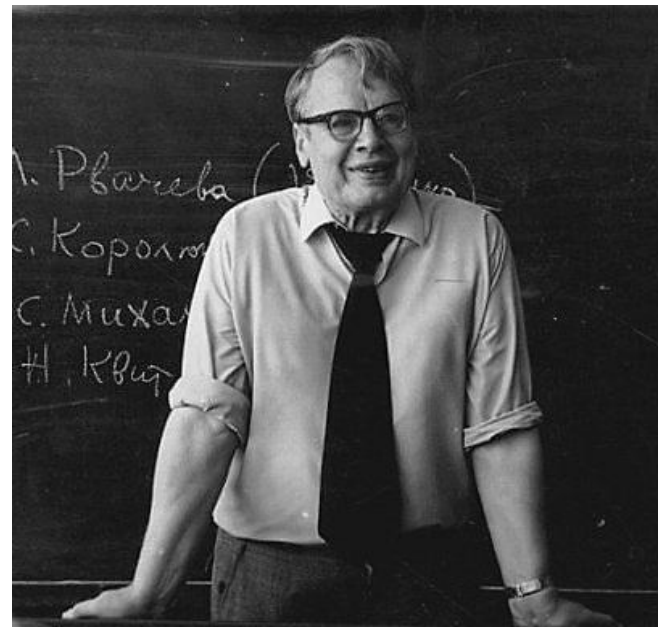
© A.V. Bochkov, 2006-2026

© Kristina Ushakov, Cover Design, 2024

<http://www.gnedenko.net/RTA>

All rights are reserved

The reference to the magazine "Reliability: Theory & Applications"
at partial use of materials is obligatory.



RELIABILITY: THEORY & APPLICATIONS

Vol.21 No.3 (92),
September 2026

San Diego
2026

Editorial Board

Editor-in-Chief

Rykov, Vladimir (Russia)

Doctor of Sci, Professor, Department of Applied Mathematics & Computer Modeling, Gubkin Russian State Oil & Gas University, Leninsky Prospect, 65, 119991 Moscow, Russia.
e-mail: vladimir_rykov@mail.ru

Managing Editors

Bochkov, Alexander (Russia)

Doctor of Sci., Scientific Secretary, JSC NIIAS, Bldg 1, 5 Orlikov Pereulok, Moscow, 107078, Russia. e-mail: a.bochkov@gmail.com

Rumyantsev, Alexander (Russia)

Doctor (DSc) of Physics and Mathematics, head of the SMITS Lab, Institute of Applied Mathematical Research KarRC RAS; Professor of the Institute of Mathematics and Information Technologies, Petrozavodsk State University; IAMR KarRC RAS, 11, Pushkinskaya str., Petrozavodsk, Karelia 185910, Russia,
e-mail: ar.psoid@gmail.com

Deputy Editors

Dimitrov, Boyan (USA)

Dr. of Math. Sci., Professor of Probability and Statistics, Associate Professor of Mathematics (Probability and Statistics), GMI Engineering and Management Inst. (now Kettering).
e-mail: bdimitro@kettering.edu

Gnedenko, Dmitry (Russia)

PhD., Associate Professor, Deputy Head, Department of Probability, Department Mechanics and Mathematics, Moscow State University, Moscow, Russia.
e-mail: dmitry@gnedenko.com

Kashtanov, Viktor A. (Russia)

PhD, M. Sc (Physics and Mathematics), Professor of Moscow Institute of Applied Mathematics, National Research University "Higher School of Economics" (Moscow, Russia).
e-mail: VAKashtan@yandex.ru

Krishnamoorthy, Achyutha (India)

M.Sc. (Mathematics), PhD (Probability, Stochastic Processes & Operations Research), Professor Emeritus, Department of Mathematics, Cochin University of Science & Technology,

Kochi-682022, INDIA.

e-mail: achyuthacusat@gmail.com

Morozov, Evsey (Russia)

Doctor (DSc) of Physics and Mathematics, Professor, Chief Researcher, SMITS Lab, Institute of Applied Mathematical Research KarRC RAS; Professor of the Institute of Mathematics and Information Technologies, Petrozavodsk State University; IAMR KarRC RAS, 11, Pushkinskaya str., Petrozavodsk, Karelia 185910, Russia
e-mail: emorozov@karelia.ru

Shybinsky, Igor (Russia)

Doctor of Sci., Professor, Division manager, JSC NIIAS, expert of the Scientific Council under Security Council of the Russian Federation,
e-mail: igor-shubinsky@yandex.ru

Yastrebenetsky, Mikhail (Ukraine)

Doctor of Sci., Professor. State Scientific Technical Center on Nuclear and Radiation Safety (Dept. NPP Control Systems Safety Analysis)- main researcher. Ukraine. Expert of International Electrotechnical Commission. Ukraine. e-mail: ma.yastreb2013@gmail.com

Associate Editors

Bulinskaya, Ekaterina (Russia)

Doctor of Physical and Mathematical Sciences, Professor, Department of Probability Theory, Lomonosov Moscow State University, Fellow of ISIR, e-mail: ebulinsk@yandex.ru

Chakravarthy, Srinivas (USA)

Ph.D., Professor of Industrial Engineering & Statistics, Departments of Industrial and Manufacturing Engineering & Mathematics, Kettering University (formerly GMI-EMI).
e-mail: schakrav@kettering.edu

Christozov, Dimitar (Bulgaria)

Professor of Computer Science, American University in Bulgaria,
e-mail: dgc@aubg.edu

Efrosinin, Dmitry (Austria)

Doktor der Naturwissenschaften, University Assistant for Research and Teaching Institute for Stochastics Johannes Kepler University, Altenbergerstrasse 69, 4040 Linz, Austria,
e-mail: dmitry.efrosinin@jku.at

Fokoue, Ernest (USA)

PhD, Professor School of Mathematics and Statistics College of Science Rochester Institute of Technology Rochester, NY 14623, USA; PhD in Statistics from University of Glasgow, Scotland, United Kingdom,
e-mail: epfeqa@rit.edu

Kaminsky, Mark (USA)

PhD, Principal reliability engineer at the NASA Goddard Space Flight Center.
e-mail: mkaminskiy@hotmail.com

Kolev, Nikolai (Brazil)

Institute of Mathematics and Statistics, University of Sao Paulo, Brazil,
e-mail: nkolev@ime.usp.br

Krivtsov, Vasiliy (USA)

PhD, Director of Reliability Analytics at the Ford Motor Company. Associate Professor of Reliability Engineering at the University of Maryland (USA).
e-mail: VKrivtsov@Ford.com, krivtsov@umd.edu

Lemeshko, Boris (Russia)

Doctor of Sci., Professor, Novosibirsk State Technical University, Professor of Theoretical and Applied Informatics Department.
e-mail: Lemeshko@ami.nstu.ru

Lesnykh, Valery (Russia)

Doctor of Sci., Professor, Adviser to General Director, Saint Petersburg, Russia, LLC "Gazprom gaznadzor"
e-mail: vvlesnykh@gmail.com

Levitin, Gregory (Israel)

PhD, The Israel Electric Corporation Ltd. Planning, Development & Technology Division. Reliability & Equipment Department, Engineer-Expert; OR and Artificial Intelligence applications in Power Engineering, Reliability.
e-mail: levitin@iec.co.il

Pagano, Michele (Italy)

PhD, Associate Professor, Department of Computer Science, Scientific Discipline Sector: Telecommunications IINF-03/A, University of Pisa, e-mail: michele.pagano@unipi.it

Telek, Miklos (Hungary)

Professor, Budapest University of Technology and Economics, Department of Networked Systems and Services; Corresponding member of the Hungarian Academy of Sciences,
e-mail: telek@hit.bme.hu

Timashev, Sviatoslav (Russia)

Doctor of Sci., Professor, Director and principal scientist the Sci & Engng Center of the Russian Academy of Sciences, Ekaterinburg.
e-mail: timashevs@cox.net

Regional Editors

Aliyev, Vugar (Azerbaijan)

Doctor (DSc) of Physical and Mathematical Sciences, Professor, Chief Researcher of the Institute of Physics of the National Academy of Sciences of Azerbaijan Director of the AMIR Technical Services Company Company,
e-mail: prof.vugar.aliyev@gmail.com

Ciu, Lirong (China)

PhD, Professor, School of Management & Economics, Beijing Institute of Technology, Beijing, P.R. China,
e-mail: lirongcui@bit.edu.cn

Papic, Ljubisha (Serbia)

PhD, Professor, Head of the Department of Industrial and Systems Engineering Faculty at Technical Sciences Cacak, University of Kragujevac, Director and Founder the Research Center of Dependability and Quality Management (DQM Research Center), Prijevor, Serbia, e-mail: dqmcenter@mts.rs

Krougly, Zinovi (Canada)

Doctor of Sci., Western University, Department of Applied Mathematics, London, Ontario, Canada, e-mail: zkrougly@gmail.com

Content Manager

Bushinskaya, Anna (Russia)

Candidate of Tech. Sci. Leading Research Fellow of the Sci & Engng Center of the Russian Academy of Sciences, Ekaterinburg.
e-mail: bushinskaya@gmail.com

e-Journal *Reliability: Theory & Applications* publishes papers, reviews, memoirs, and bibliographical materials on Reliability, Quality Control, Safety, Survivability and Maintenance.

Theoretical papers must contain new problems, finger practical applications and should not be overloaded with clumsy formal solutions.

Priority is given to descriptions of case studies.
General requirements for presented papers.

1. Papers must be presented in English in LaTeX format.
2. The total volume of the paper (with illustrations) can be up to 15 pages.
3. A presented paper must be spell-checked.
4. For those whose language is not English, we kindly recommend using professional linguistic proofs before sending a paper to the journal.

The manuscripts complying with the scope of journal and accepted by the Editor are registered and sent for external review. The reviewed articles are emailed back to the authors for revision and improvement.

The decision to accept or reject a manuscript is made by the Editor considering the referees' opinion and considering scientific importance and novelty of the presented materials. Manuscripts are published in the author's edition. The Editorial Board are not responsible for possible typos in the original text. The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Publication in this e-Journal is equal to publication in other International scientific journals.

Papers directed by Members of the Editorial Boards are accepted without referring.
The Editor has the right to change the paper title and make editorial corrections.
The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

EDITORIAL	11
------------------------	-----------

COMPARATIVE ANALYSIS OF MICROGRID CONTROL TECHNIQUES USING VARIANCE-BASED METHODS	12
--	-----------

Khaled Belhouchet, Abdelhakim Idir, Assam Zorig,
Mohamed Amroune, Ismail Ghadbane, Nouredine Hamouda

This paper presents a rigorous comparative evaluation of three advanced microgrid control strategies, "droop control", "model predictive control", and "fuzzy logic control", under varying renewable penetration levels and grid conditions. A detailed microgrid model incorporating energy storage and realistic load profiles is developed, and key performance metrics, including frequency deviation, voltage stability, energy cost, and loss of load probability, are analyzed. A two-way analysis of variance framework is employed to ensure statistical significance of the results. The findings indicate that Model Predictive Control achieves the best overall performance, particularly in terms of stability and cost efficiency, while Fuzzy Logic Control demonstrates strong robustness under uncertain conditions. Droop control, although simple and reliable, shows comparatively lower performance. Additionally, a regression model is proposed to predict energy cost based on control strategy, renewable integration, and storage capacity, providing practical insights for optimal microgrid operation.

A NEW EXTENDED EXPONENTIAL RAYLEIGH DISTRIBUTION: ESTIMATION AND APPLICATION	25
---	-----------

Dinesh Kumara, Omprakash Prajapatia, Sultan Parveenb, Abhijit Sahooa

In the present paper, we have introduced a new extended exponential Rayleigh distribution with the help of the Transformed-Transformer technique using traditional exponential and Rayleigh distributions, and some important statistical properties of the newly developed distribution have been studied. In addition, the parameters of the distribution have been estimated through the maximum likelihood, least squares, and maximum product of spacing methods of estimation on the basis of complete samples of it. Two real data sets have been considered in order to show the applicability of the newly developed distribution to the real-life problems.

A FLEXIBLE EXTENSION OF THE POWER FUNCTION DISTRIBUTION: PROPERTIES AND APPLICATIONS	44
---	-----------

Uchenna U. Uwadi, George Uchechukwu, Elebe E. Nwezza, Victor C. Ogbuehi

This paper introduces a new extension of the Power Function Distribution, referred to as the Exponentiated Gumbel Power Function (EGuPF) distribution derived from the exponentiated Gumbel family of distributions using the power function as the baseline distribution. The EGuPF distribution extends the classical power function distribution by incorporating additional shape parameters, thereby enabling it to capture J-shaped, reverse J-shaped, and bathtub-shaped hazard rate behaviours. The new distribution is expressed as an infinite linear combination of a power function distribution. The quantile function, moments, incomplete

moments, Trimmed-L Moments, mode, Moment of Residual Life (MRL), Moment of Reversed Residual Life (MRRL), entropy, and order statistics of EGuPF distribution are derived. Parameter estimation is carried out using the maximum likelihood estimation (MLE) method, and the performance of the estimators is evaluated through Monte Carlo simulation. Simulation results show that the estimators of the parameters of the EGuPF distribution are consistent, efficient, and reliable. The application of the EGuPF distribution is demonstrated using two real datasets: survival times of acute myelogenous leukaemia patients and monthly tax revenue data. In both cases, EGuPF distribution outperforms competing distributions with the same baseline, achieving the lowest AIC, BIC, and goodness-of-fit statistics. These findings establish the EGuPF distribution as a flexible and robust model for analyzing bounded and skewed data.

**FUNDAMENTAL OF MODELING FOR THE
ASSESSMENT OF ORGANIZATIONAL RISKS IN THE
INTEGRATED SAFETY SYSTEM AT EXPLOSION AND
FIRE-HAZARDOUS ENTERPRISES 62**

Evgeniy V. Gvozdev, Alexander I. Ovsyanik

The article examines the problem of quantitatively assessing organizational risks within the integrated safety system (ISS) of explosion and fire-hazardous enterprises. An approach is proposed for representing the ISS as an organizational management system, in which organizational risks are considered to be causal factors influencing the probability of hazardous events such as accidents and fires. For a normative evaluation of the contribution of organizational risks, it is proposed to use a conditional information measure, the “conditional bit.” It reflects the contribution of a specific risk to a realized hazardous event. The possibility of applying an additive weighted model and an interval scale for ranking organizational risks is substantiated. The practical significance of the results lies in their potential use by the ISS management body while developing measures aimed at reducing organizational risks and the probability of accidents and fires.

**A SURVIVAL WEIGHTED RESILIENCE FRAMEWORK FOR THE
MARSHALL–OLKIN DISTRIBUTION WITH APPLICATIONS IN
ENVIRONMENTAL AND BIOMEDICAL MODELING 73**

Shameera T., Bindu P.P

This paper develops a survival weighted hazard interpretation of the Marshall–Olkin extension for lifetime distributions. The resulting formulation, termed the Resilience Distribution (RD), introduces a parameter $\theta > 0$ that modulates the baseline hazard through the survival probability, providing a representation of resilience and hazard damping in reliability systems. The framework allows flexible hazard patterns, including increasing, decreasing, and non-monotonic shapes. Key statistical properties, including the density, survival and hazard functions, quantile function, moments, and order statistics, are derived. Theoretical results on stochastic ordering, tail behavior, and hazard elasticity are established. Parameter estimation is performed using maximum likelihood, supported by a simulation study. A special case based on the Lindley distribution is applied to real datasets, demonstrating the flexibility and interpretability of the proposed framework for modeling lifetime data.

RELIABILITY MODELING USING AN EQUILIBRIUM-BASED XGAMMA DISTRIBUTION: PROPERTIES, ESTIMATION, AND APPLICATIONS	91
--	-----------

Molay Kumar Ruidas, Sukanta Pramanik, Sudhansu S. Maiti

The equilibrium distribution approach is applied to the existing Xgamma model in this work to develop a new lifetime model known as the Induced Xgamma (i-Xg) distribution. Because of its adaptable shape properties, the suggested distribution can be used to model heterogeneous and positively skewed lifetime datasets. Probability density and distribution functions, moments, the moment generating function, reliability measures (mean residual life, hazard rate, and reversed hazard), stochastic dominance, entropy measures, inequality curves, and order statistics, including extremes, are among the important analytical properties that we derive. The model's identifiability is rigorously established. The model parameter is estimated using seven traditional estimation methods: Maximum Likelihood (ML), Least Squares (LS), Weighted LSE (WLS), Cramer-von-Mises (CvM), Maximum Product of Spacings (MPS), Anderson-Darling (AD), and Right-tail AD (RAD). Their performance is assessed in terms of bias and MSE for varying sample sizes using a Monte-Carlo simulation study. The suggested i-Xg model provides a better fit than competing distributions, as shown by applications to a variety of real-life datasets, confirming its value in statistical modelling of survival-type and reliability data.

THE AKAIKE INFORMATION CRITERION AND ARTIFICIAL INTELLIGENCE	109
---	------------

Hendrik Schabe

In this paper we use the Akaike information criterion to compare the use of parametric models with the use of very flexible models with a large parameter set, as in artificial intelligence models. It turns out, that information is lost in models with many parameters, compared with using the real model.

From Control to Catastrophe: A Common Thread in Modern Applied Statistics

This issue of Reliability: Theory & Applications presents a rare collection of studies that, at first glance, belong to incommensurable subject domains — microgrid power engineering, reliability theory, actuarial modeling, and organizational risk assessment at explosive facilities. Yet a careful reading reveals a common methodological axis running through all five papers: the problem of justified model selection under uncertainty and the price of that choice.

The opening paper by Belhouchet et al. sets the tone for the entire collection. Comparing three microgrid control strategies — classical droop control, model predictive control (MPC), and fuzzy logic control (FLC) — the authors move beyond a simple comparison of technical indicators. The application of two-way analysis of variance (ANOVA) and a regression model for predicting energy cost transforms an engineering experiment into a full-fledged statistical study. The key finding — that MPC achieves the best absolute performance while FLC demonstrates practically comparable robustness at substantially lower computational cost — reads as an invitation to discuss where the boundary between optimality and expediency lies. It is precisely this question that becomes the leitmotif of the issue.

The papers on distribution theory — the new extended exponential Rayleigh (NEER) distribution by Kumar et al., the exponentiated Gumbel power function (EGuPF) distribution by Uwadi et al., the induced Xgamma distribution by Ruidas et al., and the resilience distribution based on the Marshall–Olkin framework (RD) by Shameera and Bindu — form a second line of meaning. All four works are devoted to extending classical distributions in order to enhance their flexibility: capturing J-shaped, bathtub-shaped, and non-monotonic hazard functions, modeling right-skewed data, and accounting for the "resilience" effects of a system. What they share is meticulous attention to estimation procedures: maximum likelihood, least squares, maximum product of spacings, and their comparison by mean squared error in simulation studies. Notably, the paper by Ruidas et al. includes a dedicated section proving model identifiability — an issue that often remains behind the scenes in applied literature, yet without which any parameter estimate loses its meaning.

The paper by Gvozdev and Ovsyanik transports us to an entirely different plane — organizational risks at explosion- and fire-hazardous enterprises. The "conditional bit" measure proposed by the authors represents an original attempt to quantify the contribution of qualitatively described organizational causes to a realized hazardous event. The use of an additive weighted model and an interval scale, as well as Kendall's coefficient of concordance for assessing expert agreement, demonstrates how a rigorous statistical apparatus can be adapted to weakly structured subject domains. This work is an important reminder that the reliability of technical systems is not reducible to equipment failures; it is to a significant extent determined by the quality of management and the human factor.

The issue closes with a short but conceptually important note by Schabe, in which the Akaike information criterion is applied to compare parametric models with artificial intelligence models. The author's conclusion is disheartening for adherents of "black boxes": as sample size grows, the information loss of AI models increases as $O(n \log n)$, whereas a parametric model, if correct, retains a linear dependence. This does not mean that AI is useless — rather, it points to its proper place in the research process: as a tool for preliminary exploration, not for final inference.

The common thread uniting all the papers can be formulated as follows: every model is a compromise between accuracy and complexity, and statistics is the discipline that makes this compromise explicit. Whether it is the choice between MPC and FLC in a microgrid, between MLE and MPS in parameter estimation, or between a parametric and a non-parametric approach — everywhere it is a matter of the same thing: the need for a justified, verifiable, and reproducible choice.

The editorial board hopes that the materials presented here will resonate both with reliability theory specialists and with researchers working at the intersection of statistics, engineering, and risk management. Interdisciplinarity is not a fashionable slogan but a necessary condition for solving the problems that the modern technosphere poses to us.

Editorial Board

COMPARATIVE ANALYSIS OF MICROGRID CONTROL TECHNIQUES USING VARIANCE-BASED METHODS

KHALED BELHOUCHE¹, ABDELHAKIM IDIR¹, ASSAM ZORIG¹,
MOHAMED AMROUNE², ISMAIL GHADBANE¹, NOUREDDINE HAMOUDA³

¹Laboratory of Electrical Engineering (LGE), Faculty of Technology, University of M'sila, Algeria

²Department of Electrical Engineering, Ferhat Abbas University, Setif – 1, Algeria

³Research Center in Industrial Technologies (CRTI), Algiers, Algeria

khaled.belhouchet@univ-msila.dz

Abstract

This paper presents a rigorous comparative evaluation of three advanced microgrid control strategies, "droop control", "model predictive control", and "fuzzy logic control", under varying renewable penetration levels and grid conditions. A detailed microgrid model incorporating energy storage and realistic load profiles is developed, and key performance metrics, including frequency deviation, voltage stability, energy cost, and loss of load probability, are analyzed.

A two-way analysis of variance framework is employed to ensure statistical significance of the results. The findings indicate that Model Predictive Control achieves the best overall performance, particularly in terms of stability and cost efficiency, while Fuzzy Logic Control demonstrates strong robustness under uncertain conditions. Droop control, although simple and reliable, shows comparatively lower performance. Additionally, a regression model is proposed to predict energy cost based on control strategy, renewable integration, and storage capacity, providing practical insights for optimal microgrid operation.

Keywords: Microgrid, advanced control strategies, ANOVA, droop control, model predictive control, fuzzy logic control, renewable energy integration.

1. INTRODUCTION

Microgrids have emerged as a promising solution to enhance the reliability, resilience, and sustainability of modern power systems, particularly with the increasing penetration of distributed renewable energy sources [1, 2]. Their ability to operate in both grid-connected and islanded modes enables greater flexibility in energy management, but it also introduces significant operational challenges. One of the most critical aspects in this context is the implementation of advanced control strategies capable of maintaining voltage stability, frequency regulation, and power quality under varying load and generation conditions [3]. Traditional control methods, such as droop control and proportional-integral (PI) schemes, offer simplicity but may struggle to ensure optimal performance when microgrids face rapid fluctuations or high renewable energy variability [4, 5]. Recent studies have proposed model predictive control (MPC), adaptive fuzzy controllers, and hybrid optimization-based approaches to address these challenges [6-8]. While these advanced strategies demonstrate improved dynamic response and robustness, their comparative performance under realistic operating scenarios remains insufficiently explored. Moreover, few works have employed rigorous statistical tools, such as, to quantitatively assess the significance of performance differences between competing control strategies [9]. In this study, we extend the state of the art by combining scenario-based evaluation with ANOVA statistical

validation, enabling a systematic comparison of droop control, MPC, and FLC under varying renewable penetration levels and grid configurations. Unlike prior studies, we not only compare dynamic performance but also introduce a regression model that links energy cost to control strategies, renewable integration, and storage capacity. The combined technical-economic analysis delivers robust insights into both stability improvement and cost reduction, offering a valuable decision-support tool for researchers, practitioners, and regulatory authorities.

2. METHODS

The proposed methodology integrates microgrid modeling, scenario-based performance assessment, and ANOVA statistical validation to compare three advanced energy management strategies: Droop Control, MPC, and FLC. A detailed microgrid model with renewable sources, battery storage, and realistic load profiles is developed, and each strategy is implemented under identical conditions. Performance is tested across scenarios with varying renewable penetration and grid connectivity. Key performance indicators, including frequency deviation, voltage stability, energy cost, and Loss of Load Probability (LOLP), are extracted from simulations and analyzed using two-way ANOVA to assess the effects of control strategy, scenario, and their interaction. Post-hoc ranking is then applied to determine the optimal strategy for each operational context. The full workflow is outlined in Figure 1. The microgrid model illustrated in this figure integrates various components including the main grid, Battery Energy Storage System (BESS), local loads, and Renewable Energy Sources (RES). The RES block represents the variable power output from non-dispatchable generation units that introduces stochastic volatility into the system's power balance. The process begins with system modeling and the implementation of three energy management strategies (Droop, MPC, and FLC) under identical conditions. These strategies are tested across four defined operating scenarios, with time-domain simulations generating performance indicators such as frequency deviation, voltage stability index, energy cost, and LOLP. Two-way ANOVA is then applied to determine the statistical significance of observed differences, followed by post-hoc ranking to identify the optimal strategy for each operational context.

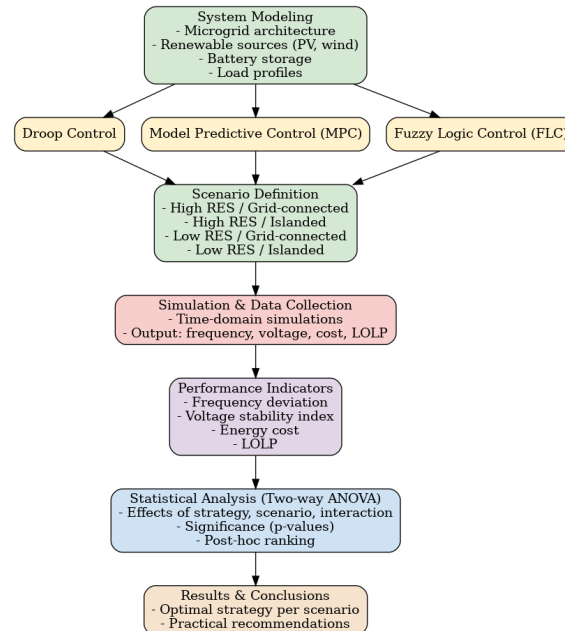


Figure 1: Research methodology for evaluating microgrid control strategies considering Photovoltaic (PV) generation, RES and LOLP as a performance indicator.

The three control strategies evaluated in this study - Droop Control, MPC, and FLC are

summarized in Table 1. These strategies represent three distinct paradigms in microgrid energy management. Droop control, one of the earliest and most widely adopted decentralized approaches, adjusts frequency and voltage setpoints proportionally to active and reactive power deviations, enabling autonomous load sharing without communication links [10]. MPC, in contrast, is a centralized optimization-based method that predicts future system behavior over a defined time horizon and computes control inputs that minimize a cost function subject to system constraints, offering fast dynamic response and optimal resource utilization [11]. FLC leverages linguistic rules and membership functions to mimic human reasoning, allowing for robust decision-making under uncertainty and incomplete modeling, making it particularly effective in renewable-rich and highly variable operating environments [12]. Together, these strategies cover a broad spectrum of control philosophies, enabling a comprehensive comparative evaluation under the scenarios considered in this work.

Table 1: Summary of Microgrid Control Strategies Considered in this Study

Control Strategy	Key Principle	Main Advantages	Limitations	Typical Applications
Droop Control	Decentralized power sharing based on frequency-active power ($f - P$) and voltage-reactive power ($V - Q$) droop characteristics [5].	Simple implementation, no communication needed, scalable.	Sensitive to tuning, reduced performance under high harmonics.	Rural/islanded microgrids, minimal infrastructure.
MPC	Predicts future states using discrete-time models; optimizes over a moving horizon [6].	Fast response, handles constraints, high control accuracy.	High computational burden, requires accurate system modeling.	Dynamic loads, systems needing optimal economic dispatch.
FLC	Decision-making based on fuzzy rules and membership functions [7].	Robust to uncertainties, no precise model required, adaptable.	Dependent on rule base and membership design.	High renewable penetration, variable environments.

Table 2 outlines the four operating scenarios used to assess the performance of Droop Control, MPC and FLC under varying renewable penetration levels and grid connectivity modes, aligning directly with the study's goal of determining which strategy performs best in specific microgrid conditions. Scenarios S1 and S2 feature high renewable penetration ($\geq 70\%$) with grid-connected and island-connected modes, respectively, allowing evaluation of the robustness of the strategy under significant renewable variability, with or without external stability support [13],[14]. It is imperative to note that absolute performance metrics, particularly frequency deviation, are highly sensitive to grid connectivity. While S2 records higher absolute deviations for all methods due to the absence of external grid support, the comparative ranking remains consistent, with MPC performing optimally under constrained islanded conditions compared to Droop and FLC.

Scenarios S3 and S4 represent low renewable penetration (less than 30 percent), again with grid-connected and islanded modes, allowing comparison in more conventional, stability -friendly conditions [15]. In these two scenarios, while MPC maintains a minor absolute advantage, the relative performance benefit of MPC over FLC remains small across key metrics. This indicates that in stable, low-penetration grids (S3 and S4), FLC provides a highly competitive and computationally lightweight alternative to MPC while delivering practically comparable stability indices. Battery storage capacity is scaled to renewable share - 1 MWh for high-renewable cases to buffer variability, and 0.5 MWh for low-renewable cases -while load profiles reflect realistic demand patterns (residential peaks for high renewables, industrial fluctuations for low renewables) [16]. This design ensures that differences detected through ANOVA capture not only the intrinsic merits of each control method but also their interaction with operating conditions, providing targeted recommendations rather than a one-size fits all solution. The difference between grid - connected and islanded operation is illustrated in Figure 2, where the former allows power exchange with the main utility grid, providing external voltage and frequency support, while the latter relies entirely on local generation and storage for stability. The

fundamental power balance of the microgrid is modeled as the sum of local generation and BESS discharge/charge, which must satisfy the instantaneous power demand of the connected loads.

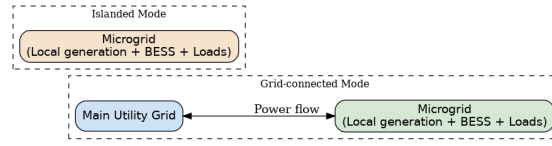


Figure 2: Schematic of the microgrid power balance, illustrating the interaction between local generation (renewable sources), BESS, and total consumer loads.

2.1. Simulation Model and Platform

The microgrid model developed for this study represents a low-voltage distribution network comprising renewable energy sources, dispatchable generation, battery energy storage, and aggregated loads. The renewable subsystem includes a PV array and a wind turbine generator, whose dynamic models incorporate irradiance and wind speed profiles generated from realistic time-series datasets [17],[18]. Dispatchable generation is modeled using a diesel generator with governor and excitation system dynamics, enabling it to provide both active and reactive Power support under varying load and renewable generation conditions (Figure 3). The BESS is modeled with a bidirectional converter and state-of-charge (SoC) management to buffer short-term renewable fluctuations and enhance system stability during islanded operation. The load model reflects mixed residential and commercial demand patterns, with daily variations and random fluctuations superimposed to emulate realistic conditions. All system components are interconnected through a three-phase, four-wire distribution network, operating at a nominal voltage of 400 V and a frequency of 50 Hz, with a total installed capacity of 1 MW. Control strategies Droop Control, MPC, and FLC are implemented within the same microgrid architecture to ensure a fair comparison. The simulation platform used is MATLAB/Simulink with the Simscape Electrical toolbox, which provides accurate time-domain modeling of power electronics, electrical machines, and control systems. A fixed - step solver with a 50 μ s step size is adopted to capture fast transient responses, while simulation horizons of 24 hours are used to assess performance over realistic operating periods. The platform enables the integration of both deterministic load/renewable profiles and stochastic disturbances, allowing the evaluation of control strategies under repeatable yet realistic conditions.

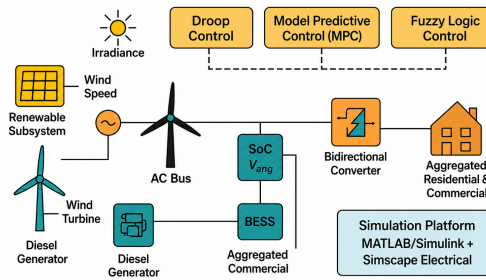


Figure 3: Low-Voltage Microgrid Architecture with Renewable Integration, Dispatchable Generation, And Advanced Control Strategies.

Table 2: Operational Scenarios for Performance Evaluation

Scenario Code	Renewable Penetration	Grid Status	Storage (BESS)	Load Profile Characteristics
S1	$\geq 70\%$	Grid-connected	1 MWh	Variable with evening peak
S2	$\geq 70\%$	Islanded	1 MWh	Variable with evening peak
S3	$\leq 30\%$	Grid-connected	0.5 MWh	Moderate fluctuations, industrial load
S4	$\leq 30\%$	Islanded	0.5 MWh	Moderate fluctuations, industrial load

2.2. Control Strategy Implementation

In order to ensure stable operation and optimal performance of the grid-connected PV-SMES (Superconducting Magnetic Energy Storage) system under varying load and generation conditions, three distinct control strategies were implemented: Droop Control, MPC, and FLC. Each approach was configured with parameters tailored to the dynamic characteristics of the test system, as summarized in Table 3.

Table 3: Summary of Control Parameters

Control Strategy	Key Parameters	Values / Settings
Droop Control	k_P (frequency droop gain), k_Q (voltage droop gain), f_0 (nominal frequency), V_0 (nominal voltage)	$k_P = 0.02$ Hz/kW, $k_Q = 0.04$ V/kVAR, $f_0 = 50$ Hz, $V_0 = 1$ p.u.
MPC	N_p (prediction horizon), N_c (control horizon), T_s (sampling time), w_P, w_Q (weight coefficients), operational constraints	$N_p = 10, N_c = 3, T_s = 1$ ms, $w_P = 0.5, w_Q = 0.5$
FLC	Inputs: power error (e_P) and change in power error (Δe_P); membership functions; rule base	Triangular membership functions (Negative [N], Zero [Z], Positive [P]), 9-rule base, centroid defuzzification

2.2.1 Droop Control

Droop control is widely employed in microgrids and distributed generation systems due to its decentralized nature and simplicity. Active power-frequency and reactive power-voltage droop characteristics are defined as:

$$f = f_0 - k_P(P - P_0), \quad V = V_0 - k_Q(Q - Q_0), \quad (1)$$

where f_0 and V_0 are the nominal system frequency and voltage, P_0 and Q_0 represent the reference active and reactive powers, and k_P and k_Q are the droop gains. Here, the $f - P$ relationship dictates frequency regulation as active power varies, whereas the $V - Q$ relation governs voltage adjustments relative to reactive power fluctuations. The gains were selected to balance fast response with adequate damping, avoiding excessive frequency or voltage deviations during transients.

2.2.2 Model Predictive Control (MPC)

MPC was configured to optimize the control inputs by predicting system behavior over a specified horizon and minimizing a quadratic cost function:

$$J = \sum_{k=1}^{N_p} \left[w_P (P_{\text{ref}} - P_k)^2 + w_Q (Q_{\text{ref}} - Q_k)^2 \right]. \quad (2)$$

This optimization is performed subject to operational limits on the converter voltage, converter current, and the SMES coil current. The prediction horizon (N_p) defines the forward evaluation window in Equation (2), while the control horizon ($N_c = 3$) restricts the number of optimized control adjustments per interval, such that control increments $\Delta u(k) = 0$ for $k > N_c$ over the prediction window. These parameters, along with the weighting factors (w_P and w_Q), are selected to balance rapid dynamic response against computational burden and control effort. The MPC framework is implemented using a discrete-time state-space representation of the PV-SMES system, updated at each sampling interval (T_s).

2.2.3 Fuzzy logic control (FLC)

The FLC schema was designed to handle system nonlinearities and uncertainties without requiring an explicit mathematical model. Two inputs, power error (e_p) and change in power error (Δe_p) were fuzzified into three linguistic terms (negative, zero, positive) using triangular membership functions. The rule base comprised nine rules mapping these inputs to the control action, and the centroid method was employed for defuzzification. This approach provided a smooth and adaptive response under fluctuating PV output and grid disturbances conditions.

3. COMPARATIVE EVALUATION OF CONTROL STRATEGIES

Before applying statistical testing, a descriptive comparison was conducted to visualize the raw performance of droop control, MPC, and FLC across the four operating scenarios. Figure 4 presents the results for frequency deviation, voltage stability index, energy cost, and LOLP. MPC consistently achieves the best absolute performance with the lowest frequency deviation and LOLP across all evaluated contexts, including the more stable, low-renewable penetration scenarios (S3 and S4). However, as illustrated in Figure 4, the absolute performance margins between MPC and FLC remain consistently narrow across S2, S3, and S4 for frequency deviation and LOLP. While MPC consistently achieves the lowest absolute values, the small performance margin confirms that FLC achieves near-optimal regulation under low-volatility conditions (S3, S4). FLC thus presents a strong practical trade-off, delivering near-MPC performance without the online optimization overhead or precise system model requirements of MPC. In voltage stability, all methods remain above 0.89 p.u., with MPC leading slightly, followed by FLC, and droop showing the largest deviations. Energy cost analysis shows MPC as most cost-efficient, particularly under high renewable penetration, with FLC ranking second and droop highest. These trends suggest MPC as the best overall performer, FLC as a strong alternative in less volatile conditions, and droop as reliable but less efficient. However, statistical testing via ANOVA is required to confirm whether these observed differences are significant [19]-[25].

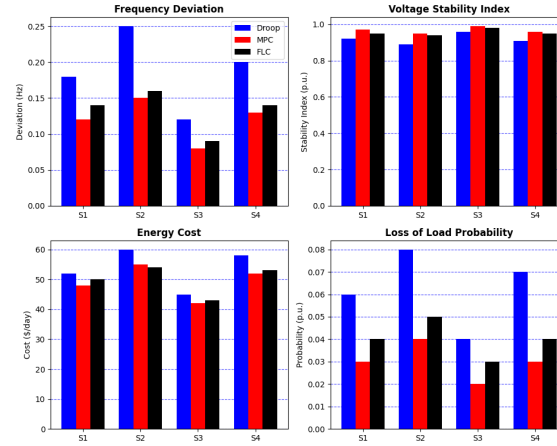


Figure 4: Comparative performance of Droop Control, MPC, and FLC across four operating scenarios.

4. STATISTICAL ANALYSIS (ANOVA INTEGRATION)

While these descriptive results indicate clear performance patterns, they do not confirm whether the differences are statistically significant or due to random variation. The following section applies a two-way ANOVA to formally assess the significance of the main effects (control strategy and scenario) and their interactions for each performance indicator. To ensure that the comparative assessment of the three control strategies - Droop Control, MPC, and FLC - is not only based on raw performance indicators but also statistically robust, Analysis of Variance (ANOVA) was incorporated into the evaluation framework. ANOVA enables the determination of whether observed performance differences across strategies and operating scenarios are statistically significant, rather than the result of random variations inherent in simulation data. The performance indicators extracted from the time-domain simulations included frequency deviation, voltage stability index, energy cost, and LOLP. A two-way ANOVA was selected, with control strategy and operating scenario as independent factors, and each performance metric as the dependent variable. Three control-related factors were selected for optimization using the Taguchi $L9(3^3)$ orthogonal array. Each factor was assigned three levels, as shown in Table 4, and coded numerically to facilitate statistical analysis.

Table 4: Factors and Levels for the Taguchi $L9$ Orthogonal Array Design

Factor	Description	Level 1 (1)	Level 2 (2)	Level 3 (3)
A	Control Strategy	Droop Control	MPC	FLC
B	Renewable Penetration Ratio	Low (30%)	Medium (50%)	High (70%+)
C	BESS Capacity	0.5 MWh	1.0 MWh	1.5 MWh

In this study, the Taguchi $L9(3^3)$ orthogonal array was employed to systematically evaluate the influence of three key control related factors - control strategy, renewable penetration level, and BESS capacity - on system performance. Among the available performance indicators (e.g., frequency deviation, voltage deviation, loss of load probability), the analysis focused exclusively on Energy Cost (euro/MWh) as the primary response variable. The smaller - the-better criterion was applied to Energy Cost, allowing the Taguchi method to identify the most cost-effective combination of factor levels within the tested operating scenarios. The selected factors, their levels, and the corresponding orthogonal experimental plan are summarized in Table 5.

Table 5: Taguchi L9 Orthogonal Array Experimental Results for Energy Cost

Run	Control Strategy (A)	Renewable Penetration (B)	BESS Capacity (C)	Energy Cost (Euro/MWh)
1	Droop	Low (30%)	0.5 MWh	83.5
2	Droop	Medium (50%)	1.0 MWh	81.0
3	Droop	High (70%+)	1.5 MWh	79.0
4	MPC	Low (30%)	1.0 MWh	77.5
5	MPC	Medium (50%)	1.5 MWh	75.0
6	MPC	High (70%+)	0.5 MWh	76.2
7	FLC	Low (30%)	1.5 MWh	78.0
8	FLC	Medium (50%)	0.5 MWh	77.0
9	FLC	High (70%+)	1.0 MWh	75.5

The empirical regression model developed to predict the Energy Cost (Euro/MWh) based on the centered and normalized factor configurations is expressed as follows:

$$y = 76.856 - 2.167A - 1.383B - 0.783C, \quad (3)$$

where y represents the predicted energy cost (Euro/MWh). To ensure statistical uniformity and physical alignment with the interaction patterns, the parameters A , B , and C represent the standardized, coded levels defined strictly within the symmetric range of $\{-1, 0, 1\}$. Specifically, these coded coordinates map onto the physical Taguchi design parameters as follows:

- Control Strategy (A): Coded as -1 for Droop Control, 0 for MPC, and $+1$ for FLC.
- Renewable Penetration (B): Coded as -1 for Low (30%), 0 for Medium (50%), and $+1$ for High (70%+).
- BESS Capacity (C): Coded as -1 for 0.5 MWh, 0 for 1.0 MWh, and $+1$ for 1.5 MWh.

This normalized representation sets 0 as the mid-point baseline configuration, enabling the signs and magnitudes of the regression coefficients to accurately reflect each factor's actual sensitivity contribution.

The interaction plots in Figure 5 show the combined effects of Control Strategy, Renewable Penetration, and BESS Capacity on energy cost. Energy cost decreases as the level of each factor increases, with Control Strategy having the largest impact, followed by Renewable Penetration and BESS Capacity. The nearly parallel lines indicate that the factors mostly act additively, with only minor interactions: higher BESS capacity slightly enhances the cost reduction effect of advanced control strategies and increased renewable penetration. Overall, the plots visually confirm the regression model and highlight the predictable reduction in energy cost with improved system design and operation.

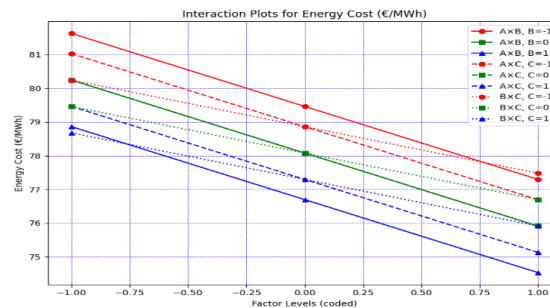


Figure 5: Effect of Control Strategy, Renewable Penetration, and BESS Capacity on Energy Cost (euro/MWh).

The 3D surface plot (Figure 6) and the corresponding contour plot (Figure 7) illustrate the combined effect of Control Strategy and Renewable Penetration on predicted energy cost, while

holding BESS Capacity at its medium level. The surface demonstrates a clear decreasing trend in energy cost as either the control strategy or renewable penetration increases, consistent with the negative regression coefficients for these factors. The smooth gradient observed across the surface indicates mostly additive effects, with minimal interaction between the two variables at the medium BESS capacity. The contour plot provides a top-down view, highlighting regions of lower energy cost corresponding to more advanced control strategies and higher renewable penetration levels. Together, these plots confirm the regression model and provide a visual representation of how operational and design choices can optimize energy cost.

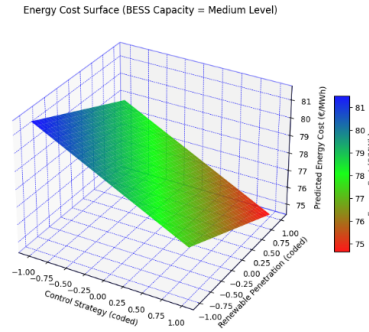


Figure 6: Effect of Control Strategy and Renewable Penetration on Energy Cost (Medium BESS Capacity).

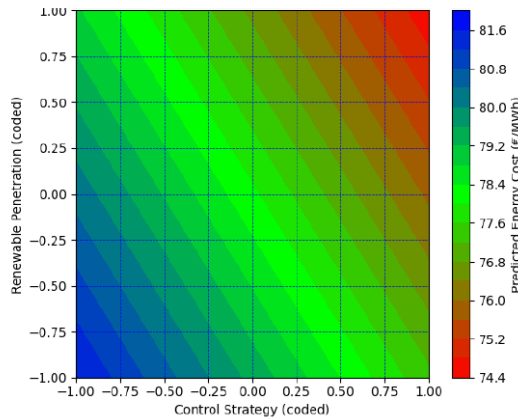


Figure 7: Predicted Energy Cost Contours as a Function of Control Strategy and Renewable Penetration (BESS Capacity = Medium Level).

The combined interaction plot shown in Figure 8 demonstrates how Control Strategy, Renewable Penetration, and BESS Capacity influence the predicted energy cost. While the total energy cost is influenced by the operational scenario (e.g., lower in S3 due to reduced renewable variability), the regression analysis successfully isolates the control strategy contribution. This confirms that MPC remains the most cost-efficient choice for optimizing dispatch and BESS usage across all studied conditions when evaluated relative to the baseline scenario constraints. Different line styles and colors represent the interactions between pairs of factors, while the blue marker highlights the optimal combination (MPC, Medium Renewable Penetration, 1.5 MWh BESS), which yields the lowest energy cost. The plot shows that Control Strategy has the largest effect, followed by Renewable Penetration and BESS Capacity, and that small deviations around the optimal levels result in only minor increases in cost. This indicates a robust and stable optimum, confirming that careful selection of all three factors can reliably minimize energy cost in the system. Graphically, the primary influence of Control Strategy (A) is evident in Figure 8 by comparing the steep downward trajectory along the horizontal x - axis with the vertical

spacing between parameter lines. Varying the control strategy across the x - axis induces an overall energy cost reduction of approximately 4.33 €/MWh, whereas shifting between renewable penetration levels (vertical gap between curves) and BESS capacities produces smaller variations of 2.77 €/MWh and 1.57 €/MWh, respectively. This visual trend directly reflects the magnitude of the regression coefficients in Equation (3), confirming that real-time operational management exerts a stronger sensitivity influence on microgrid economics than physical hardware sizing alone.

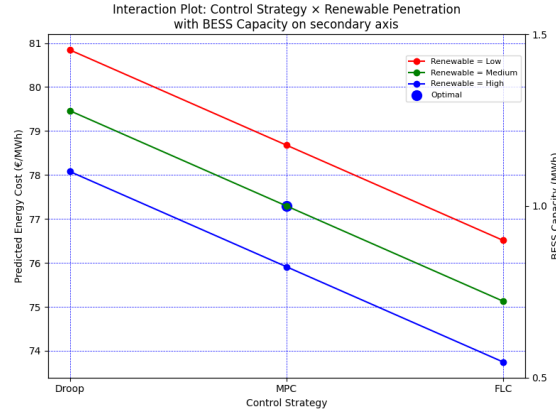


Figure 8: Interaction Plot of Control Strategy, Renewable Penetration, and Bess Capacity on Predicted Energy Cost (euro/Mwh), Highlighting the Optimum Response (MPC, Medium Renewable, 1.5 Mwh Bess).

The primary reason the Control Strategy (A) exhibits the largest sensitivity impact on the system's operational economy - outperforming the absolute contributions of both renewable sharing (B) and storage constraints (C) stems from its active operational nature. Physical components like BESS capacity define the static operational limits of the microgrid, but the control loop represents the active intelligence that commands power dispatch. Advanced strategies like MPC achieve substantial cost reductions because their objective functions continuously optimize active and reactive flows across a moving horizon, minimizing transience and maximizing the economic utilization of renewable surpluses. Conversely, decentralized structures like classic droop control operate on instantaneous proportional rules without foresight, leading to premature battery depletion and a heavier reliance on expensive dispatchable diesel generation. Consequently, optimizing the control paradigm yields a higher performance dividend than simply expanding physical hardware capacity.

Figure 9 illustrates the predicted energy cost as a function of Control Strategy and Renewable Penetration, with BESS Capacity fixed at 1.5 MWh. The contour colors indicate the magnitude of energy cost, decreasing from yellow to dark purple. The yellow marker highlights the optimal response, corresponding to the MPC control strategy, Medium Renewable Penetration, and 1.5 MWh BESS, which achieves the lowest predicted energy cost. The smooth gradient indicates that energy cost decreases consistently with more advanced control strategies and higher renewable integration, while the contours show minimal interaction between the two factors. This visualization confirms The regression model's predictions and clearly identifies the configuration that minimizes energy cost. The observed waveforms reflect the controller's reaction to the step-change in load and renewable availability. As illustrated by the parametric variations in Figure 8, the transient response profiles (specifically, the frequency overshoot and settling time) are highly sensitive to the control configuration. This dynamic behavior is primarily influenced by the MPC's predictive horizon length (N_p). Because the moving horizon defines how far into the future the optimizer projects system state trajectories, a longer horizon equips the controller to anticipate steep stochastic gradients in PV generation or load demand. This look-ahead capability allows the MPC to activate BESS charging or discharging cycles preemptively, smoothing the power transition and minimizing grid transience. Shorter horizons, by contrast,

limit the controller to reactive adjustments, which results in larger peak overshoots and extended recovery oscillations under high-penetration conditions.

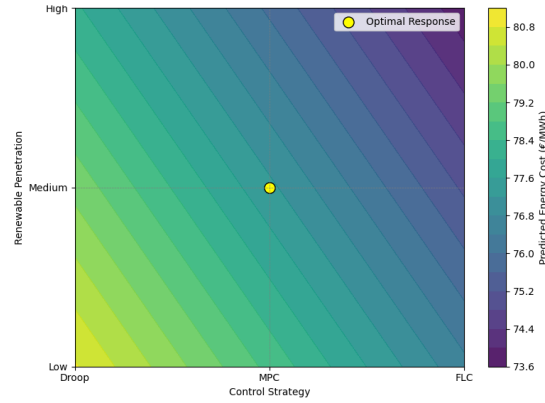


Figure 9: Contour Plot of Predicted Energy Cost (euro/Mwh) As a Function of Control Strategy and Renewable Penetration (Bess Capacity = 1.5 Mwh), Highlighting the Optimal Response.

Figure 10 shows that the residuals are approximately aligned with the reference line, indicating a near - normal distribution and confirming the validity of the normality assumption required for ANOVA and regression analysis. The residuals, defined as the differences between observed and predicted energy cost values, are randomly distributed around the zero reference line without any discernible pattern. This behavior indicates homoscedasticity and confirms that the selected variables - control strategy, renewable energy penetration, and BESS capacity - adequately explain the variability in total energy cost. The total cost represents the true, absolute economic expenditure obtained directly from the physical time-domain microgrid simulation run. In contrast, the predicted cost is the numerical estimate computed via the simplified regression model established in Equation (3).

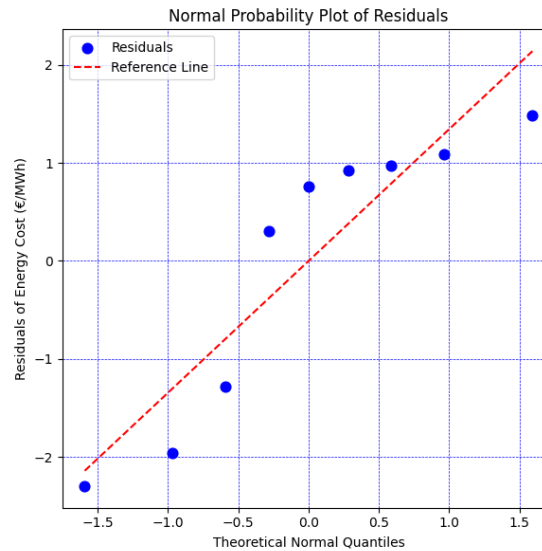


Figure 10: Normal Probability Plot of Residuals.

5. CONCLUSION

This study presented a comparative evaluation of Droop Control, MPC, and FLC in a realistic microgrid framework with renewable generation, storage, and variable demand. The results

demonstrated that MPC consistently achieved the best overall performance, minimizing frequency deviation, energy cost, and loss of load probability, particularly in high-renewable islanded scenarios. FLC provided robust and adaptive behavior in moderately variable conditions, while droop control, though simple and communication-free, showed reduced efficiency and higher costs. The application of two-way ANOVA confirmed that these performance differences are statistically significant, and the regression model revealed that control strategy has the greatest influence on energy cost, followed by renewable penetration and storage capacity. Together, these findings offer both technical and economic insights, guiding operators in selecting control strategies that balance stability, cost, and robustness.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING:

During the preparation of this work, the authors used Google Gemini in order to improve language, readability, and formatting. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

- [1] R. H. Lasseter, "MicroGrids," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, vol. 1, 2002, pp. 305–308. <https://doi.org/10.1109/PESW.2002.985003>
- [2] J. M. Guerrero, M. Chandorkar, T. Lee, and P. C. Loh, "Advanced control architectures for intelligent microgrids–Part I: Decentralized and hierarchical control," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1254–1262, Apr. 2013. <https://doi.org/10.1109/TIE.2012.2194969>
- [3] A. Bidram and A. Davoudi, "Hierarchical structure of microgrids control system," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1963–1976, Dec. 2012. <https://doi.org/10.1109/TSG.2012.2197425>
- [4] C. K. Sao and P. W. Lehn, "Control and power management of converter fed microgrids," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1088–1098, Aug. 2008. <https://doi.org/10.1109/TPWRS.2008.922232>
- [5] Y. A.-R. I. Mohamed and E. F. El-Saadany, "Adaptive decentralized droop controller to preserve power sharing stability of paralleled inverters in distributed generation microgrids," *IEEE Trans. Power Electron.*, vol. 23, no. 6, pp. 2806–2816, Nov. 2008. <https://doi.org/10.1109/TPEL.2008.2005100>
- [6] H. Bevrani, T. Ise, and Y. Miura, "Virtual synchronous generators: A survey and new perspectives," *Int. J. Electr. Power Energy Syst.*, vol. 54, pp. 244–254, Jan. 2014. <https://doi.org/10.1016/j.ijepes.2013.07.009>
- [7] K. Belhouchet, A. Bayadi, N. Alti, and L. Ouchen, "Effects analysis of the pollution layer parameters on a high-voltage porcelain cylindrical insulator using response surface methodology," *Diagnostyka*, vol. 22, no. 2, pp. 21–28, 2021. <https://doi.org/10.29354/diag/134114>
- [8] L. Ouchen, K. Belhouchet, A. Bayadi, et al., "A finite element with statistical analysis study to investigate the electrical performance of composite insulators under water droplet impact," *Sci. Rep.*, vol. 15, p. 7364, 2025. <https://doi.org/10.1038/s41598-025-90863-x>
- [9] K. Belhouchet, A. Zemmit, A. Bayadi, L. Ouchen, and A. Zorig, "A novel application of artificial intelligence technology for outdoor high-voltage composite insulator," *Measurement*, vol. 238, p. 115372, 2024. <https://doi.org/10.1016/j.measurement.2024.115372>
- [10] A. Zorig, K. Belhouchet, and L. Ouchen, "Performance evaluation of cap and pin insulator under pollution conditions using Finite element method (FEM)," *Przegląd Elektrotechniczny*, vol. 100, no. 9, pp. 142–145, 2024. <https://doi.org/10.15199/48.2024.09.26>
- [11] Y. Yan, D. Shi, D. Bian, B. Huang, Z. Yi, and Z. Wang, "Small-Signal Stability Analysis and Performance Evaluation of Microgrids Under Distributed Control," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4848–4858, Sep. 2019. <https://doi.org/10.1109/TSG.2018.2869566>
- [12] H. S. Salama, A. Ali, and K. Mahmoud, "Adaptive control for microgrid frequency stability integrating battery energy storage and photovoltaic," *Sci. Rep.*, vol. 15, p. 43839, 2025. <https://doi.org/10.1038/s41598-025-28321-x>

- [13] D. C. Montgomery, *Design and Analysis of Experiments*, 9th ed. Hoboken, NJ, USA: Wiley, 2017.
- [14] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. de Vicuña, and M. Castilla, "Hierarchical control of droop-controlled AC and DC microgrids—A general approach toward standardization," *IEEE Trans. Ind. Electron.*, vol. 58, no. 1, pp. 158–172, Jan. 2011. <https://doi.org/10.1109/TIE.2010.2066534>
- [15] K. Belhouichet, A. Bayadi, and M. E. Bendib, "Artificial neural networks and genetic algorithm modelling and identification of arc parameter in insulators flashover voltage and leakage current," *Int. J. Comput. Aided Eng. Technol.*, vol. 11, no. 1, pp. 1–13, 2019. <https://doi.org/10.1504/IJCAET.2019.096708>
- [16] A. Villalón, M. Rivera, Y. Salgueiro, J. Muñoz, T. Dragičević, and F. Blaabjerg, "Predictive control for microgrid applications: A review study," *Energies*, vol. 13, no. 10, p. 2454, May 2020. <https://doi.org/10.3390/en13102454>
- [17] O. Feddaoui, R. Toufouti, J. Labed, and S. Meziane, "Fuzzy logic control of hybrid systems including renewable energy in microgrids," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 10, no. 6, pp. 5559–5569, Dec. 2020. <https://doi.org/10.11591/ijece.v10i6.pp5559-5569>
- [18] M. A. Hossain, H. R. Pota, M. J. Hossain, and F. Blaabjerg, "Evolution of microgrids with converter-interfaced generations: Challenges and opportunities," *Int. J. Electr. Power Energy Syst.*, vol. 109, pp. 160–186, Jul. 2019. <https://doi.org/10.1016/j.ijepes.2019.01.038>
- [19] K. Belhouichet, I. Ghadbane, A. Zemmit, L. Ouchen, and A. Zorig, "Measurement and evaluation of the flashover voltage on polluted cap and pin insulator: An experimental and theoretical study," *Electr. Power Syst. Res.*, vol. 236, p. 110979, 2024. <https://doi.org/10.1016/j.epsr.2024.110979>
- [20] J. C. Vasquez, J. M. Guerrero, M. Savaghebi, J. Eloy-Garcia, and R. Teodorescu, "Modeling, analysis, and design of stationary-reference-frame droop-controlled parallel three-phase voltage source inverters," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1271–1280, Apr. 2013. <https://doi.org/10.1109/TIE.2012.2194951>
- [21] K. T. Tan, X. Y. Peng, P. L. So, Y. C. Chu, and M. Z. Q. Chen, "Centralized control for parallel operation of distributed generation inverters in microgrids," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1977–1987, Dec. 2012. <https://doi.org/10.1109/TSG.2012.2205952>
- [22] F. Mohammadi, G. A. Shahgholian, and B. Fani, "Robust Control Strategies for Microgrids: A Review," *IEEE Syst. J.*, vol. 16, no. 2, pp. 2401–2412, Jun. 2022. <https://doi.org/10.1109/JSYST.2021.3077213>
- [23] A. Yazdani and R. Iravani, *Voltage-Sourced Converters in Power Systems: Modeling, Control, and Applications*. Hoboken, NJ, USA: Wiley, 2010. <https://ieeexplore.ieee.org/servlet/opac?bknumber=6739364>
- [24] K. Belhouichet, A. Zemmit, L. Ouchen, N. Alti, and A. Zorig, "Experimental investigation on dielectric properties of 1512L insulator using finite element analysis (FEM)," *Diagnostyka*, vol. 25, no. 2, p. 2024209, 2024. <https://doi.org/10.29354/diag/186863>
- [25] R. H. Lasseter et al., "CERTS Microgrid Laboratory Test Bed," *IEEE Trans. Power Del.*, vol. 26, no. 1, pp. 325–332, Jan. 2011. <https://doi.org/10.1109/TPWRD.2010.2051819>

A NEW EXTENDED EXPONENTIAL RAYLEIGH DISTRIBUTION: ESTIMATION AND APPLICATION

DINESH KUMAR^{a,1}, OMPRAKASH PRAJAPATI^{a,2,*}, SULTAN PARVEEN^{b,3},
ABHIJIT SAHOO^{a,4}

•

^aDepartment of Statistics, Banaras Hindu University, Varanasi 221005, India

^bDepartment of Mathematics and Statistics, Aliah University, Kolkata 700160, India

¹ dineshkumarstat@bhu.ac.in ² omprakash.au@bhu.ac.in ³ parveen.bhu@gmail.com

⁴ abhijitsahoo@bhu.ac.in

*Corresponding author

Abstract

In the present paper, we have introduced a new extended exponential Rayleigh distribution with the help of the Transformed-Transformer technique using traditional exponential and Rayleigh distributions, and some important statistical properties of the newly developed distribution have been studied. In addition, the parameters of the distribution have been estimated through the maximum likelihood, least squares, and maximum product of spacing methods of estimation on the basis of complete samples of it. Two real data sets have been considered in order to show the applicability of the newly developed distribution to the real life problems.

Keywords: Transformed-Transformer, Exponential distribution, Rayleigh distribution, Estimation, Simulation Study.

1. INTRODUCTION

In the field of survival and reliability analysis to capture realistic situations, the Rayleigh distribution plays a key role to analyze the uncertainty involved in the data. The Rayleigh distribution was originally introduced by [18]. The probability density function (pdf) and cumulative distribution function (cdf) of the Rayleigh distribution are given by equations (1) and (2) respectively.

$$\psi(t) = 2\theta t e^{-\theta t^2}; t > 0, \theta > 0. \quad (1)$$

$$\Psi(t) = 1 - e^{-\theta t^2}; t > 0, \theta > 0. \quad (2)$$

Its hazard rate function, i.e., $h(t) = 2\theta t$, is an increasing function, linearly with a variate value. It is clearly a rarely attainable situation. A number of authors have motivated to generalize or transform Rayleigh distribution in order to get a new distribution which has different shapes of their hazard rate function such as monotonic decreasing or bath tub or inverted bath tub etc.

For example, [7] have introduced a new Rayleigh-Rayleigh distribution (RRD) and studied its various statistical properties and compared its performance over existing Rayleigh distribution on real data sets. [15] introduced the generalized Rayleigh distribution and various estimation techniques have been used to estimate its unknown parameters. [20] introduced a new generalization of the Rayleigh distribution using the conservability approach. [8] have introduced the power exponentiated Rayleigh distribution, and studied its various statistical properties, also demonstrates its effectiveness through real data sets and simulated data using goodness-of-fit measures. [17] have introduced the transmuted generalized Rayleigh distribution using the

quadratic rank transmutation map (QRTM) and demonstrated its important statistical properties. [13] obtained Bayesian estimators for the Rayleigh distribution parameter using squared error and LINEX loss functions. [2] analyzes the impact of various censored sampling schemes on the estimation of the parameter of the Rayleigh distribution. [1] proposed the unit generalized Rayleigh distribution, establish its statistical and entropy properties, and demonstrates its effectiveness through simulations and also carried out real data analysis. [16] discussed the estimation of parameters of the exponentiated Rayleigh distribution using Type II censored data. [22] compared maximum likelihood and Bayesian estimation methods for the inverted exponentiated Rayleigh distribution under step-stress type-I hybrid censoring.

Motivated by the review of the above literature on the generalizations and transformations of the Rayleigh distribution, we introduce a new extended exponential Rayleigh (NEER) distribution using the Transformed-Transformer technique proposed by [5]. Let $\psi(t)$ be the pdf of a random variable $T \in [a, b]$, for $-\infty < a \leq t \leq b < \infty$. Let X be another random variable with pdf $g(x)$ and cdf $G(x)$. Let $R[G(x)]$ be a function of the cdf $G(x)$ of the random variable X , such that $R[G(x)]$ satisfies the following conditions:

$$\left. \begin{aligned} R[G(x)] &\in [a, b], \\ R[G(x)] &\text{ is differentiable and monotonically non-decreasing,} \\ R[G(x)] &\rightarrow a \text{ as } X \rightarrow -\infty \text{ and } R[G(x)] \rightarrow b \text{ as } X \rightarrow \infty. \end{aligned} \right\} \quad (3)$$

Here, $R[G(x)]$ is called a transformer.

Now, according to the Transformed-Transformer technique or T-X technique, the cdf $F(x)$ of the new family of distributions is defined as follows,

$$F(x) = \int_a^{R[G(x)]} \psi(t) dt = \Psi\{R[G(x)]\}, \quad (4)$$

where, $\Psi(t)$ is the cdf corresponding to the pdf $\psi(t)$ of the random variable T . Also, the pdf $f(x)$, say corresponding to the cdf $F(x)$ is obtained as follows,

$$f(x) = \left\{ \frac{d}{dx} R[G(x)] \right\} \psi[R[G(x)]]. \quad (5)$$

Consider the support of T is $[a, \infty)$ with $a \geq 0$, without loss of generality, we can assume that $a = 0$. $R[G(x)]$ can be defined in various forms, such as $-\ln(1 - G(x))$, $\frac{G(x)}{1-G(x)}$, $-\ln(1 - G^\alpha(x))$ and $\frac{G^\alpha(x)}{1-G^\alpha(x)}$, where $\alpha > 0$.

In particular, if we take the support of random variable T as $[0, \infty)$, then the functional form of $R[G(x)]$ may be defined as

$$R[G(x)] = \frac{G(x)}{1 - G(x)},$$

which satisfies all three conditions of equation (3). Then the cdf $F(x)$ and the pdf $f(x)$ of newly generalized distribution using Transformed-Transformer technique or T-X technique are given by equations (6) and (7) respectively.

$$F(x) = \int_0^{\frac{G(x)}{1-G(x)}} \psi(t; \theta) dt = \Psi\left\{ \frac{G(x)}{1 - G(x)} \right\}, \quad (6)$$

$$f(x) = \frac{g(x)}{[1 - G(x)]^2} \psi\left\{ \frac{G(x)}{1 - G(x)} \right\}, \quad (7)$$

where, $g(x) = \frac{d}{dx} G(x)$ is the pdf corresponding to cdf $G(x)$.

In the literature, numerous generalized family of distributions have been proposed and studied using the Transformed-Transformer technique or T-X technique. Notable examples include exponentiated Weibull-G family by [14], Weibull-G family by [9], T-geometric family of

discrete distribution by [4], Gompertz-G family of distributions by [3], Exponentiated Half-Logistic Family of Distributions by [12], The type I half-logistic family of distributions by [11], T-normal family of distributions by [6] etc. These developments reflect a rising interest in generalizing traditional distributions to better representative of complex data sets.

Motivated by the need for a flexible lifetime distribution, this article proposes a new statistical model and investigates its theoretical properties, estimation methods, and practical applications. The remainder of this article is organized as follows: Section 2 describe the pdf and cdf of the proposed distribution. The statistical properties are studies in Section 3. The parameter estimation procedures are described in Section 4. A Monte Carlo simulation study to evaluate the performance of the estimator is explained in Section 5. Applications of the proposed model to two real data sets are presented in Section 6. Finally, Section 7 summarizes the main findings and concludes the article.

2. NEW EXTENDED EXPONENTIAL RAYLEIGH DISTRIBUTION

The pdf $g(x)$ and the cdf $G(x)$ of the exponential distribution using the rate parameter λ is given by equations (8) and (9) respectively.

$$g(x) = \lambda e^{-\lambda x}; x > 0, \lambda > 0. \quad (8)$$

$$G(x) = 1 - e^{-\lambda x}; x > 0, \lambda > 0. \quad (9)$$

If the random variable T follow the Rayleigh distribution having pdf defined in equation (1), consider, Rayleigh distribution as baseline distribution and $G(x)$ be the cdf of exponential distribution defined in equation (9), then the functional form of $R[G(x)]$ is as

$$R[G(x)] = \frac{G(x)}{1 - G(x)} = \frac{1 - e^{-\lambda x}}{1 - (1 - e^{-\lambda x})} = e^{\lambda x} - 1, \quad (10)$$

and consequently the cdf of New Extended Exponential Rayleigh distribution is given as follows:

$$F(x; \lambda, \theta) = 1 - e^{-\theta(e^{\lambda x} - 1)^2}; x > 0, \lambda > 0, \theta > 0. \quad (11)$$

Also, the pdf $f(x; \lambda, \theta)$ corresponding to the cdf $F(x; \lambda, \theta)$ of our proposed distribution is given as follows:

$$f(x; \lambda, \theta) = 2\theta\lambda e^{\lambda x}(e^{\lambda x} - 1)e^{-\theta(e^{\lambda x} - 1)^2}; x > 0, \lambda > 0, \theta > 0. \quad (12)$$

We will denote new extended exponential Rayleigh distribution having pdf as in (12) by NEER(λ, θ) distribution. The shapes of pdf $f(x; \lambda, \theta)$ and cdf $F(x; \lambda, \theta)$ of NEER(λ, θ) distribution for various combination of values of its parameters λ and θ are shown in Figure 1 and 2 respectively.

3. STATISTICAL PROPERTIES

In this section, we examine the statistical properties of the NEER(λ, θ) distribution, with particular emphasis on the survival function, hazard rate function, quantile function, and order statistics.

3.1. Survival and Hazard Rate Function

The survival function and hazard rate function of the NEER(λ, θ) distribution are given as follows:

$$S(x) = e^{-\theta(e^{\lambda x} - 1)^2}, \quad (13)$$

$$h(x) = 2\theta\lambda e^{\lambda x}(e^{\lambda x} - 1). \quad (14)$$

From Figure 3, it is evident that the hazard rate function is strictly increasing for various combinations of the parameters λ and θ . This monotonic behavior indicates that the proposed

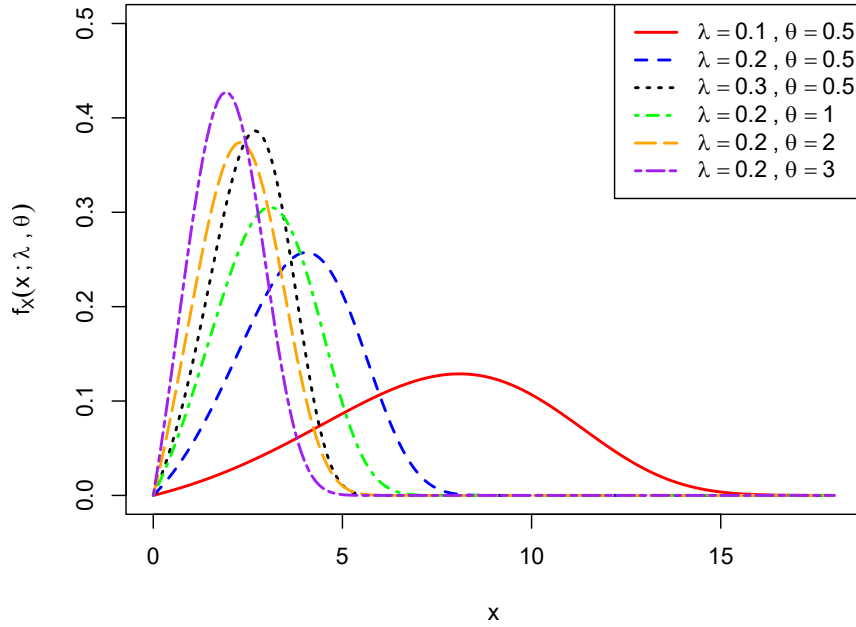


Figure 1: The pdf of NEER(λ, θ) distribution for different combinations values of λ and θ .

distribution is well suited for modeling lifetime data exhibiting progressive aging and wear-out mechanisms. Consequently, it has broad potential applications in reliability engineering, survival analysis, medical research, actuarial science, and preventive maintenance, where the risk of failure increases continuously over time.

3.2. Quantile function

If $Q(u)$ be the quantile function of NEER(λ, θ) distribution, then it can be obtained as the solution of

$$\begin{aligned} F(Q(u)) &= u, \\ \Rightarrow 1 - e^{-\theta(e^{\lambda Q(u)} - 1)^2} &= u, \end{aligned}$$

where $u \sim U[0, 1]$ is a uniformly distributed random variable. On solving the above the equation for $Q(u)$, we obtained

$$Q(u) = \frac{1}{\lambda} \ln \left[1 + \left(-\frac{1}{\theta} \ln(1 - u) \right)^{\frac{1}{2}} \right]. \quad (15)$$

If Q_i be the i^{th} quantile function of NEER(λ, θ) distribution, then the first quantile Q_1 , second quantile Q_2 , and third quantile Q_3 can be obtained by putting $u = \frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$ respectively in equation (15) as follows:

$$\begin{aligned} Q_1 &= \frac{1}{\lambda} \ln \left[1 + \left(\frac{1}{\theta} \ln \left(\frac{4}{3} \right) \right)^{\frac{1}{2}} \right], \\ Q_2 &= \frac{1}{\lambda} \ln \left[1 + \left(\frac{1}{\theta} \ln(2) \right)^{\frac{1}{2}} \right], \end{aligned}$$

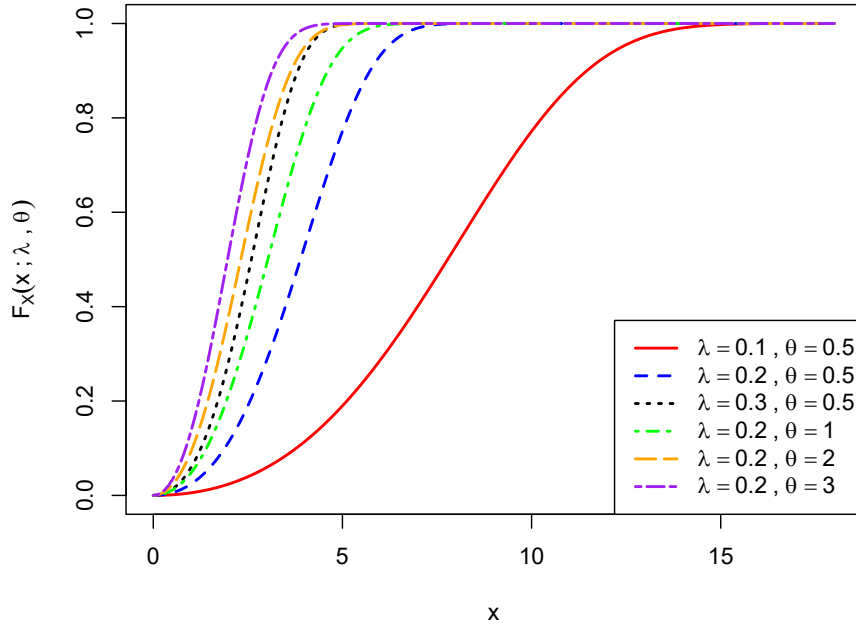


Figure 2: The cdf of NEER(λ, θ) distribution for different combinations values of λ and θ .

$$Q_3 = \frac{1}{\lambda} \ln \left[1 + \left(\frac{1}{\theta} \ln(4) \right)^{\frac{1}{2}} \right].$$

3.3. Order Statistics

Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ be the order statistics of a random sample $\underline{X} = (X_1, X_2, \dots, X_n)$ of size n from NEER(λ, θ) distribution. Then the pdf of r^{th} order statistics is given by

$$f_{X_{(r)}}(x_{(r)}; \lambda, \theta) = \frac{1}{B(r, n-r+1)} [F(x_{(r)}; \lambda, \theta)]^{r-1} f(x_{(r)}; \lambda, \theta) [1 - F(x_{(r)}; \lambda, \theta)]^{n-r}. \quad (16)$$

By using equations (11) and (12) in equation (16), the pdf of the r^{th} order statistics $x_{(r)}$ is obtained as follows:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} 2\lambda\theta e^{\lambda x} (e^{\lambda x} - 1) e^{-(n-r+1)\theta(e^{\lambda x}-1)^2} \left[1 - e^{-\theta(e^{\lambda x}-1)^2} \right]^{r-1}. \quad (17)$$

For $r = 1$ and n , we obtained the pdfs of $x_{(1)} = \min(x_1, x_2, \dots, x_n)$ and $x_{(n)} = \max(x_1, x_2, \dots, x_n)$ respectively as;

$$f_{X_{(1)}}(x; \lambda, \theta) = 2n\lambda\theta e^{\lambda x} (e^{\lambda x} - 1) e^{-n\theta(e^{\lambda x}-1)^2}, \quad (18)$$

$$f_{X_{(n)}}(x; \lambda, \theta) = 2n\lambda\theta e^{\lambda x} (e^{\lambda x} - 1) e^{-\theta(e^{\lambda x}-1)^2} \left[1 - e^{-\theta(e^{\lambda x}-1)^2} \right]^{n-1}. \quad (19)$$

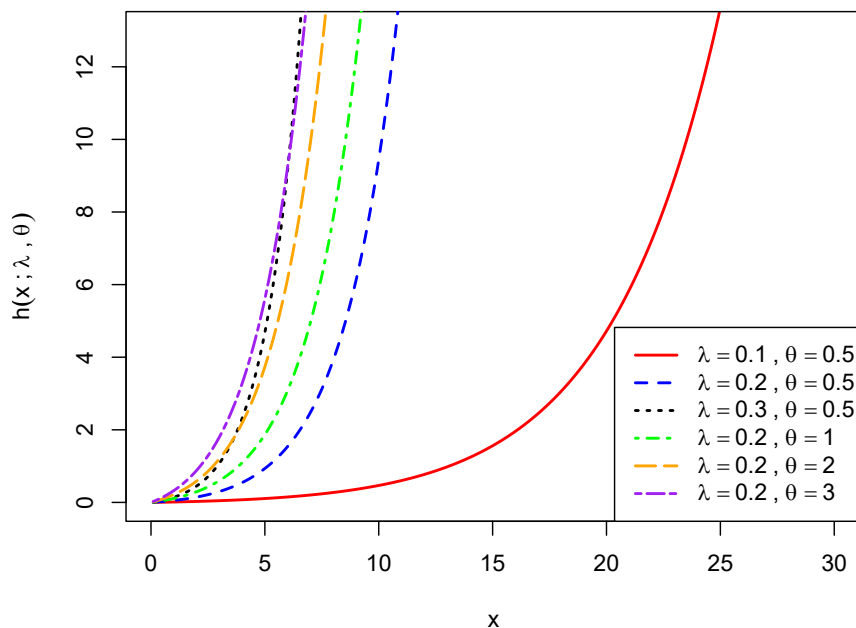


Figure 3: The hazard rate functions of NEER(λ, θ) distribution for different combinations values of λ and θ .

4. ESTIMATION PROCEDURES

4.1. Maximum Likelihood Estimation (MLE)

Let $\underline{X}=(X_1, X_2, \dots, X_n)$ be a random sample of size n and let X_i NEER(λ, θ) distribution $\forall i = 1, 2, \dots, n$. Then the likelihood function of the random sample $\underline{X}$ is given by

$$L = \prod_{i=1}^n f(x_i; \lambda, \theta), \quad (20)$$

and consequently the log-likelihood function of the random sample $\underline{X}$ is given by

$$\ln L = n \ln(2) + n \ln(\lambda) + n \ln(\theta) + \lambda \sum_{i=1}^n x_i - \theta \sum_{i=1}^n (e^{\lambda x_i} - 1)^2 + \sum_{i=1}^n \ln(e^{\lambda x_i} - 1). \quad (21)$$

Now, the log-likelihood equations or normal equations for estimating the parameters λ and θ can be obtained by differentiating the equation (22) with respect to λ and θ , we get

$$\begin{aligned} \frac{\partial \ln L}{\partial \lambda} &= 0, \\ \Rightarrow \frac{n}{\lambda} + \sum_{i=1}^n x_i - 2\theta \sum_{i=1}^n (e^{\lambda x_i} - 1)x_i e^{\lambda x_i} + \sum_{i=1}^n \frac{x_i e^{\lambda x_i}}{e^{\lambda x_i} - 1} &= 0, \end{aligned} \quad (22)$$

and

$$\begin{aligned} \frac{\partial \ln L}{\partial \theta} &= 0, \\ \Rightarrow \frac{n}{\theta} - \sum_{i=1}^n (e^{\lambda x_i} - 1)^2 &= 0. \end{aligned} \quad (23)$$

Since the equations (22) and (23) are two non-linear likelihood equations. Solving these equations provides the maximum likelihood estimates (MLEs) $\hat{\lambda}_{ML}$ and $\hat{\theta}_{ML}$ of the parameters λ and θ . However, since these equations are not in closed form, they cannot be solved analytically. Therefore, some numerical approximation technique can be used for their approximate solutions, particularly the Newton-Raphson method is used to obtained the MLE's of the parameters λ and θ .

4.2. Least Squares Estimation (LSE)

Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ be the order statistics of the chosen random sample $X = (X_1, X_2, \dots, X_n)$. Then the least squares estimators of parameters λ and θ can be obtained by minimizing the following w.r.t. λ and θ .

$$\begin{aligned} z(\lambda, \theta) &= \sum_{i=1}^n \left(F(x; \lambda, \theta) - \frac{i}{n+1} \right)^2, \\ \Rightarrow z(\lambda, \theta) &= \sum_{i=1}^n \left(1 - e^{-\theta(e^{\lambda x(i)} - 1)^2} - \frac{i}{n+1} \right)^2. \end{aligned} \quad (24)$$

To minimize $z(\lambda, \theta)$ as defined in equation (24), we take partial derivatives of equation (24) with respect to λ and θ , and equating them to zero, we get the following normal equations.

$$\begin{aligned} \frac{\partial z(\lambda, \theta)}{\partial \lambda} &= 0, \\ \Rightarrow 4\theta \sum_{i=1}^n \left(1 - e^{-\theta(e^{\lambda x(i)} - 1)^2} - \frac{i}{n+1} \right) e^{-\theta(e^{\lambda x(i)} - 1)^2} (e^{\lambda x(i)} - 1) e^{\lambda x(i)} x(i) &= 0, \end{aligned} \quad (25)$$

and

$$\begin{aligned} \frac{\partial z(\lambda, \theta)}{\partial \theta} &= 0, \\ \Rightarrow 2 \sum_{i=1}^n \left(1 - e^{-\theta(e^{\lambda x(i)} - 1)^2} - \frac{i}{n+1} \right) e^{-\theta(e^{\lambda x(i)} - 1)^2} (e^{\lambda x(i)} - 1)^2 &= 0. \end{aligned} \quad (26)$$

The above two equations (25) and (26) are non-linear, which can not be solved analytically for λ and θ . Therefore, as mentioned above numerical approximation technique namely, Newton-Raphson method is used for solving these two equations for the parameters λ and θ respectively and we will abbreviate these estimates by $\hat{\lambda}_{LSE}$ and $\hat{\theta}_{LSE}$ respectively.

4.3. Maximum Product Spacing (MPS) estimation

Now a days, a popular method of estimation technique is the maximum product of sapcing method of estimation and it is introduced by [10]. Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ be the order statistics of chosen random sample $\underline{X} = (X_1, X_2, \dots, X_n)$. Then the spacing is defined as follow;

$$\begin{aligned} D_i &= \int_{x_{(i-1)}}^{x_{(i)}} f(x; \lambda, \theta) dx, \\ &= F(x_{(i)}; \lambda, \theta) - F(x_{(i-1)}; \lambda, \theta) \quad ; i = 1, 2, \dots, (n+1). \end{aligned}$$

Where $F(x_{(0)}; \lambda, \theta) = 0$ and $F(x_{(n+1)}; \lambda, \theta) = 1$. Therefore,

$$D_i = \begin{cases} F(x_{(1)}; \lambda, \theta) & \text{for } i = 1, \\ F(x_{(i)}; \lambda, \theta) - F(x_{(i-1)}; \lambda, \theta) & \text{for } i = 2, 3, \dots, n, \\ 1 - F(x_{(n)}) & \text{for } i = n + 1. \end{cases}$$

After substituting the values of $F(x_1)$, $F(x_i)$, $F(x_{i-1})$ and $F(x_n)$ in the above equation, we get;

$$D_i = \begin{cases} 1 - e^{-\theta(e^{\lambda x(1)} - 1)^2} & \text{for } i = 1, \\ e^{-\theta(e^{\lambda x(i-1)} - 1)^2} - e^{-\theta(e^{\lambda x(i)} - 1)^2} & \text{for } i = 2, 3, \dots, n, \\ e^{-\theta(e^{\lambda x(n)} - 1)^2} & \text{for } i = n + 1. \end{cases} \quad (27)$$

Let G be the geometric mean of the spacings $D_1, D_2, \dots, D_{n+1}$. Then The MPS estimates $\hat{\lambda}_{MS}$ and $\hat{\theta}_{MS}$ of λ and θ respectively can be obtained by maximizing the geometric mean of spacing as follow:

$$G = \left(\prod_{i=1}^{n+1} D_i \right)^{\frac{1}{n+1}}, \quad (28)$$

and hence the natural log of geometric mean G is obtained as follows

$$\begin{aligned} \ln G &= \frac{1}{n+1} \sum_{i=1}^{n+1} \ln(D_i), \\ &= \frac{1}{n+1} \left[\ln \left(1 - e^{-\theta(e^{\lambda x(1)} - 1)^2} \right) + \sum_{i=2}^n \ln \left(e^{-\theta(e^{\lambda x(i-1)} - 1)^2} - e^{-\theta(e^{\lambda x(i)} - 1)^2} \right) - \theta(e^{\lambda x(n)} - 1)^2 \right]. \end{aligned} \quad (29)$$

The MPS estimates $\hat{\lambda}_{MS}$ and $\hat{\theta}_{MS}$ of λ and θ can be obtained as simultaneous solution of the following equations

$$\begin{aligned} \frac{\partial \ln G}{\partial \lambda} &= 0, \\ \Rightarrow \frac{1}{n+1} &\left[\frac{2\theta e^{-\theta(e^{\lambda x(1)} - 1)^2} (e^{\lambda x(1)} - 1) e^{\lambda x(1)} x(1)}{1 - e^{-\theta(e^{\lambda x(1)} - 1)^2}} - 2\theta(e^{\lambda x(n)} - 1) e^{\lambda x(n)} x(n) \right. \\ &\left. + \sum_{i=2}^n \left(\frac{-2\theta e^{-\theta(e^{\lambda x(i-1)} - 1)^2} (e^{\lambda x(i-1)} - 1) e^{\lambda x(i-1)} x(i-1) + 2\theta e^{-\theta(e^{\lambda x(i)} - 1)^2} (e^{\lambda x(i)} - 1) e^{\lambda x(i)} x(i)}{e^{-\theta(e^{\lambda x(i-1)} - 1)^2} - e^{-\theta(e^{\lambda x(i)} - 1)^2}} \right) \right] = 0, \end{aligned} \quad (30)$$

and,

$$\begin{aligned} \frac{\partial \ln G}{\partial \theta} &= 0, \\ \Rightarrow \frac{1}{n+1} &\left[\frac{e^{-\theta(e^{\lambda x(1)} - 1)^2} (e^{\lambda x(1)} - 1)^2}{1 - e^{-\theta(e^{\lambda x(1)} - 1)^2}} - (e^{\lambda x(n)} - 1)^2 \right. \\ &\left. + \sum_{i=2}^n \left(\frac{-e^{-\theta(e^{\lambda x(i-1)} - 1)^2} (e^{\lambda x(i-1)} - 1)^2 + e^{-\theta(e^{\lambda x(i)} - 1)^2} (e^{\lambda x(i)} - 1)^2}{e^{-\theta(e^{\lambda x(i-1)} - 1)^2} - e^{-\theta(e^{\lambda x(i)} - 1)^2}} \right) \right] = 0, \end{aligned} \quad (31)$$

Since equation 30 and 31 can not be expressed in nice close form, therefor in order to get the solution for λ and θ , use of numerical technique is suggested.

5. SIMULATION STUDY

In the present section, we have carried simulation study in order to know the performance of the various considered estimators of the parameters λ and θ for their long-run behavior. The estimator have been compared in terms of their mean squared errors (MSEs), To calculate MSEs, we have generated 10000 different samples from NEER(λ, θ) distribution, in order to get convergence of

the result. The MSEs of the estimators are the functions of n , λ , and θ . For calculation of MSEs, we have arbitrarily considered the values $\lambda = 0.2, 0.8, 1.2$ and $\theta = 0.2, 0.8, 1.2$ and generated 10000 different samples of sizes $n = 10, 20, 30, 50$, and 120 using R software. We have also calculated average estimates (AEs), absolute bias (AB), confidence interval (CI), and width of the confidence interval (W_CI) of the considered estimators of the parameters λ and θ . Since the parameter space is restricted to non-negative values, so negative lower confidence limits were truncated at zero. The results of simulation are summarised in Tables [1-3].

Table 1: Average estimates, mean square error (MSE), absolute bias (AB), confidence interval (CI), and width of confidence interval (W_CI) for the parameters λ and θ obtained using the Maximum Likelihood Estimation (MLE) method.

λ	θ	n	$\hat{\lambda}$	$\hat{\theta}$	MSE_{λ}	MSE_{θ}	AB_{λ}	AB_{θ}	CI_{λ}		CI_{θ}		W_CI_{λ}	W_CI_{θ}
									Lower	Upper	Lower	Upper		
0.2	0.2	10	0.2115	0.2672	0.0034	0.0762	0.0376	0.1398	0.1333	0.2896	0.0000	0.6018	0.1563	0.6692
		20	0.2047	0.2355	0.0015	0.0453	0.0244	0.0935	0.1677	0.2418	0.0903	0.3807	0.0740	0.2904
		30	0.2021	0.2249	0.0009	0.0166	0.0183	0.0633	0.1786	0.2256	0.1361	0.3138	0.0470	0.1777
		50	0.2018	0.2107	0.0004	0.0072	0.0132	0.0456	0.1869	0.2168	0.1554	0.2660	0.0300	0.1106
		120	0.2002	0.2039	0.0001	0.0013	0.0057	0.0178	0.1948	0.2057	0.1837	0.2240	0.0109	0.0403
0.2	0.8	10	0.2018	1.0313	0.0023	0.4612	0.0302	0.4060	0.1011	0.3025	0.0000	2.3990	0.2013	2.7353
		20	0.1990	0.9383	0.0011	0.2871	0.0215	0.2694	0.1448	0.2532	0.2287	1.6480	0.1084	1.4193
		30	0.2005	0.9219	0.0010	0.2433	0.0206	0.2565	0.1702	0.2307	0.5334	1.3104	0.0605	0.7769
		50	0.1961	0.8823	0.0004	0.1175	0.0120	0.1640	0.1755	0.2167	0.6116	1.1530	0.0412	0.5414
		120	0.1960	0.8055	0.0001	0.0108	0.0074	0.0352	0.1891	0.2028	0.7213	0.8897	0.0137	0.1684
0.2	1.2	10	0.2021	1.5355	0.0021	0.8953	0.0321	0.5553	0.0872	0.3170	0.0000	3.6963	0.2298	4.3215
		20	0.1987	1.3960	0.0010	0.4115	0.0216	0.3873	0.1425	0.2549	0.3635	2.4285	0.1125	2.0650
		30	0.2000	1.3150	0.0008	0.3016	0.0180	0.3029	0.1602	0.2398	0.6186	2.0114	0.0796	1.3929
		50	0.1953	1.3292	0.0004	0.1645	0.0129	0.2320	0.1737	0.2168	0.9275	1.7309	0.0431	0.8033
		120	0.1984	1.2018	0.0002	0.0425	0.0086	0.0953	0.1899	0.2069	1.0550	1.3486	0.0170	0.2936
0.8	0.2	10	0.9897	0.2746	0.1726	0.2495	0.2955	0.2318	0.7048	1.2746	0.0000	0.6055	0.5698	0.6619
		20	0.8632	0.2691	0.0610	0.1236	0.1689	0.1715	0.7133	1.0130	0.0856	0.4526	0.2997	0.3670
		30	0.8338	0.2301	0.0314	0.0355	0.1155	0.1084	0.7312	0.9364	0.1223	0.3379	0.2052	0.2157
		50	0.8202	0.2197	0.0171	0.0259	0.0866	0.0843	0.7627	0.8776	0.1641	0.2753	0.1149	0.1112
		120	0.7960	0.2095	0.0051	0.0063	0.0426	0.0437	0.7709	0.8212	0.1850	0.2339	0.0502	0.0489
0.8	0.8	10	1.0059	0.9169	0.2337	1.1836	0.3239	0.6507	0.5710	1.4409	0.0000	2.2909	0.8699	2.7480
		20	0.8779	1.0680	0.1022	0.9098	0.2071	0.6221	0.6856	1.0702	0.3710	1.7650	0.3846	1.3940
		30	0.8459	0.9582	0.0557	0.5884	0.1577	0.4643	0.7115	0.9804	0.4962	1.4201	0.2688	0.9239
		50	0.8205	0.9579	0.0304	0.4914	0.1235	0.4118	0.7378	0.9033	0.6664	1.2494	0.1655	0.5830
		120	0.8053	0.8780	0.0123	0.1670	0.0760	0.2416	0.7725	0.8381	0.7711	0.9849	0.0656	0.2138
0.8	1.2	10	0.9422	1.4993	0.1581	2.7643	0.2723	0.9789	0.4787	1.4057	0.0000	3.6214	0.9270	4.2441
		20	0.8568	1.6382	0.0823	2.5911	0.2042	0.9449	0.6369	1.0767	0.5038	2.7726	0.4398	2.2687
		30	0.8390	1.4626	0.0554	1.1283	0.1668	0.7118	0.6977	0.9804	0.7693	2.1560	0.2826	1.3867
		50	0.8337	1.3430	0.0355	0.6433	0.1335	0.5522	0.7457	0.9217	0.9156	1.7705	0.1760	0.8549
		120	0.8214	1.2931	0.0186	0.3939	0.0947	0.4047	0.7859	0.8570	1.1296	1.4566	0.0711	0.3270
1.2	0.2	10	1.4841	0.3057	0.3922	0.2820	0.4768	0.2746	1.0592	1.9090	0.0000	0.7121	0.8499	0.8129
		20	1.3813	0.2670	0.1942	0.2157	0.3277	0.2078	1.1687	1.5938	0.0889	0.4450	0.4251	0.3560
		30	1.3068	0.2524	0.1200	0.1049	0.2570	0.1654	1.1692	1.4444	0.1447	0.3601	0.2753	0.2154
		50	1.2608	0.2289	0.0638	0.0404	0.1840	0.1168	1.1710	1.3506	0.1631	0.2946	0.1796	0.1314
		120	1.2243	0.2085	0.0226	0.0107	0.1088	0.0674	1.1906	1.2580	0.1870	0.2299	0.0674	0.0429
1.2	0.8	10	1.3665	1.4520	0.3363	5.4327	0.4257	1.1421	0.7632	1.9699	0.0000	3.3436	1.2066	3.7832
		20	1.2371	1.3634	0.1483	3.9112	0.2851	0.8986	0.9115	1.5626	0.2717	2.4552	0.6510	2.1836
		30	1.2089	1.3269	0.1167	3.4793	0.2373	0.7900	1.0011	1.4166	0.6212	2.0327	0.4155	1.4114
		50	1.2239	1.0734	0.0876	1.3402	0.2213	0.5839	1.1051	1.3428	0.7372	1.4097	0.2376	0.6725
		120	1.1855	0.9252	0.0281	0.2693	0.1207	0.2969	1.1290	1.2421	0.7814	1.0691	0.1131	0.2876
1.2	1.2	10	1.2842	2.0482	0.2234	7.5080	0.3460	1.3692	0.6454	1.9230	0.0000	4.6598	1.2776	5.2231
		20	1.2246	1.7966	0.1117	4.9470	0.2536	1.0271	0.8746	1.5747	0.3931	3.2002	0.7001	2.8071
		30	1.1609	1.7484	0.0603	2.9393	0.1828	0.7988	0.9057	1.4160	0.7612	2.7356	0.5103	1.9745
		50	1.1702	1.6502	0.0542	1.7794	0.1619	0.6833	1.0360	1.3045	1.1335	2.1669	0.2685	1.0334
		120	1.1884	1.3602	0.0243	0.4384	0.1092	0.3790	1.1297	1.2470	1.1693	1.5511	0.1173	0.3818

Table 2: Average estimates, mean square error (MSE), absolute bias (AB), confidence interval (CI), and width of confidence interval (W_CI) for the parameters λ and θ obtained using the Least Square Estimation (LSE) method.

λ	θ	n	$\hat{\lambda}$	$\hat{\theta}$	MSE_{λ}	MSE_{θ}	AB_{λ}	AB_{θ}	CI_{λ}		CI_{θ}		W_CI_{λ}	W_CI_{θ}
									Lower	Upper	Lower	Upper		
0.2	0.2	10	0.1784	0.5717	0.0044	0.5083	0.0510	0.4123	0.0000	0.5032	0.0000	3.3382	0.6497	5.5330
		20	0.1819	0.4971	0.0035	0.4010	0.0456	0.3416	0.0000	0.3771	0.0000	2.0809	0.3904	3.1675
		30	0.1887	0.4105	0.0028	0.2749	0.0408	0.2622	0.0684	0.3089	0.0000	1.2271	0.2405	1.6331
		50	0.1915	0.3366	0.0019	0.1389	0.0336	0.1902	0.1201	0.2629	0.0000	0.7513	0.1428	0.8294
		120	0.1953	0.2649	0.0010	0.0351	0.0239	0.1097	0.1622	0.2284	0.1141	0.4157	0.0662	0.3016
0.2	0.8	10	0.1797	1.4546	0.0036	1.4647	0.0436	0.8015	0.0000	0.6104	0.0000	9.2854	0.8615	15.6614
		20	0.1800	1.4762	0.0029	1.2418	0.0397	0.7660	0.0000	0.4059	0.0000	5.7666	0.4518	8.5809
		30	0.1897	1.4319	0.0026	1.0032	0.0388	0.6808	0.0110	0.3684	0.0000	4.7232	0.3574	6.5826
		50	0.1850	1.2880	0.0020	0.9579	0.0343	0.6232	0.0798	0.2901	0.0000	3.1396	0.2103	3.7033
		120	0.1991	0.9916	0.0016	0.5488	0.0307	0.4397	0.1536	0.2446	0.3037	1.6795	0.0910	1.3758
0.2	1.2	10	0.1905	1.7497	0.0029	1.6851	0.0399	0.9282	0.0000	0.6549	0.0000	11.4134	0.9288	19.3274
		20	0.1806	1.9288	0.0023	1.4207	0.0355	0.7902	0.0000	0.4304	0.0000	7.8464	0.4997	11.8353
		30	0.1892	1.9177	0.0021	1.3853	0.0334	0.7698	0.0170	0.3615	0.0000	5.9334	0.3445	8.0313
		50	0.1848	1.8382	0.0020	1.2674	0.0323	0.7690	0.0681	0.3014	0.0000	4.5854	0.2332	5.4945
		120	0.1951	1.6308	0.0018	0.7459	0.0317	0.5792	0.1457	0.2445	0.5159	2.7457	0.0989	2.2298
0.8	0.2	10	0.7645	0.5904	0.0889	0.6895	0.2347	0.4796	0.0000	2.3034	0.0000	4.0321	3.0778	6.8833
		20	0.7568	0.5075	0.0657	0.5150	0.2041	0.3827	0.0411	1.4726	0.0000	2.0157	1.4315	3.0165
		30	0.7540	0.4692	0.0545	0.4321	0.1844	0.3400	0.2664	1.2417	0.0000	1.4580	0.9753	1.9776
		50	0.7741	0.3407	0.0340	0.1603	0.1449	0.2105	0.4428	1.1055	0.0000	0.8713	0.6628	1.0611
		120	0.7777	0.2742	0.0202	0.0494	0.1120	0.1329	0.6558	0.8997	0.1259	0.4226	0.2439	0.2967
0.8	0.8	10	0.7735	1.3329	0.0756	1.4366	0.1896	0.8229	0.0000	2.7119	0.0000	8.8517	3.8767	15.0375
		20	0.7648	1.4503	0.0748	1.1746	0.1857	0.7017	0.0000	1.7855	0.0000	6.0017	2.0413	9.1028
		30	0.7736	1.2945	0.0569	1.1420	0.1725	0.6947	0.0820	1.4652	0.0000	4.2582	1.3832	5.9274
		50	0.7764	1.2804	0.0495	1.0096	0.1673	0.6909	0.3705	1.1823	0.0000	3.0481	0.8118	3.5356
		120	0.7923	1.0480	0.0271	0.5853	0.1204	0.4491	0.5797	1.0049	0.2705	1.8255	0.4252	1.5550
0.8	1.2	10	0.7653	1.7078	0.0599	1.5002	0.1578	0.8056	0.0000	2.8344	0.0000	12.0770	4.1381	20.7383
		20	0.7811	1.7755	0.0428	1.4833	0.1463	0.7836	0.0000	1.8579	0.0000	7.2130	2.1536	10.8751
		30	0.7518	1.8425	0.0398	1.2026	0.1400	0.7791	0.0000	1.5538	0.0000	6.4397	1.6040	9.1944
		50	0.7758	1.7425	0.0396	1.0350	0.1385	0.6365	0.2991	1.2526	0.0000	4.3349	0.9535	5.1847
		120	0.8072	1.5099	0.0317	0.7403	0.1287	0.5874	0.5542	1.0602	0.2439	2.7758	0.5059	2.5319
1.2	0.2	10	1.0703	0.6763	0.1864	0.8049	0.3363	0.5488	0.0000	3.2678	0.0000	4.3595	4.3950	7.3664
		20	1.1829	0.4522	0.1442	0.4613	0.2910	0.3391	0.0000	2.3939	0.0000	2.0413	2.4220	3.1781
		30	1.1553	0.4237	0.1135	0.3536	0.2630	0.2971	0.4029	1.9078	0.0000	1.3637	1.5049	1.8799
		50	1.1689	0.3583	0.0850	0.2145	0.2272	0.2313	0.7266	1.6112	0.0000	0.8320	0.8846	0.9474
		120	1.1794	0.2588	0.0379	0.0354	0.1526	0.1156	0.9881	1.3707	0.1090	0.4086	0.3827	0.2996
1.2	0.8	10	1.1073	1.4034	0.1369	1.4798	0.2812	0.8380	0.0000	4.0592	0.0000	9.5947	5.9038	16.3826
		20	1.1224	1.4047	0.1241	1.2673	0.2597	0.7701	0.0000	2.6533	0.0000	5.8017	3.0617	8.7940
		30	1.1346	1.4306	0.1163	1.1457	0.2534	0.7083	0.0988	2.1703	0.0000	4.4486	2.0715	6.0360
		50	1.1846	1.2006	0.1132	1.0317	0.2443	0.6648	0.5114	1.8577	0.0000	3.1039	1.3462	3.8066
		120	1.1702	1.1272	0.0734	0.8294	0.2109	0.5658	0.9002	1.4401	0.3901	1.8642	0.5398	1.4741
1.2	1.2	10	1.1241	1.7816	0.1006	1.9270	0.2281	0.9067	0.0000	4.1963	0.0000	12.3563	6.1444	21.1494
		20	1.1265	1.8310	0.0994	1.6022	0.2229	0.8935	0.0000	2.8540	0.0000	8.1551	3.4549	12.6482
		30	1.1187	1.9161	0.0931	1.4531	0.2101	0.7760	0.0430	2.1943	0.0000	5.9817	2.1513	8.1312
		50	1.1408	1.8962	0.0705	1.2250	0.2057	0.7709	0.3883	1.8934	0.0000	4.9097	1.5051	6.0271
		120	1.1673	1.6609	0.0658	0.8569	0.1779	0.6340	0.8741	1.4606	0.5534	2.7684	0.5864	2.2151

Table 3: Average estimates, mean square error (MSE), absolute bias (AB), confidence interval (CI), and width of confidence interval (W_CI) for the parameters λ and θ obtained using the Maximum Product of Spacing (MPS) estimation method

λ	θ	n	$\hat{\lambda}$	$\hat{\theta}$	MSE_{λ}	MSE_{θ}	AB_{λ}	AB_{θ}	CI_{λ}		CI_{θ}		W_CI_{λ}	W_CI_{θ}
									Lower	Upper	Lower	Upper		
0.2	0.2	10	0.1787	0.3537	0.0019	0.1066	0.0316	0.1682	0.0000	0.3700	0.0000	1.4388	0.3827	2.1700
		20	0.1853	0.2808	0.0009	0.0351	0.0217	0.0940	0.0451	0.3255	0.0000	0.9478	0.2803	1.3339
		30	0.1891	0.2495	0.0005	0.0123	0.0163	0.0591	0.0775	0.3007	0.0000	0.7394	0.2232	0.9797
		50	0.1933	0.2235	0.0003	0.0039	0.0104	0.0341	0.1080	0.2786	0.0000	0.5672	0.1706	0.6874
		120	0.1976	0.2067	0.0001	0.0008	0.0048	0.0118	0.1320	0.2633	0.0000	0.4544	0.1312	0.4953
0.2	0.8	10	0.1831	1.0031	0.0010	0.7454	0.0253	0.8033	0.0000	0.3841	0.0000	3.9862	0.4019	5.9662
		20	0.1896	0.8650	0.0005	0.4651	0.0177	0.6651	0.0446	0.3346	0.0000	2.7741	0.2900	3.8182
		30	0.1936	0.8444	0.0003	0.4271	0.0136	0.6445	0.0611	0.3261	0.0000	2.5206	0.2650	3.3523
		50	0.1931	0.8199	0.0002	0.3870	0.0107	0.6199	0.0849	0.3013	0.0000	2.1795	0.2164	2.7191
		120	0.1936	0.8030	0.0001	0.3638	0.0077	0.6030	0.1363	0.2510	0.0921	1.5139	0.1146	1.4218
0.2	1.2	10	0.1898	1.3486	0.0008	1.4581	0.0229	1.1486	0.0000	0.4168	0.0000	5.6518	0.4541	8.6063
		20	0.1926	1.2409	0.0004	1.1040	0.0163	1.0409	0.0199	0.3653	0.0000	4.3408	0.3454	6.2000
		30	0.1930	1.2207	0.0003	1.0474	0.0161	1.0207	0.0469	0.3390	0.0000	3.7942	0.2921	5.1471
		50	0.1924	1.2043	0.0002	1.0090	0.0128	1.0043	0.0941	0.2908	0.0000	2.9100	0.1966	3.4115
		120	0.1930	1.2001	0.0001	1.0003	0.0095	1.0001	0.1381	0.2479	0.2450	2.1552	0.1098	1.9102
0.8	0.2	10	0.7417	0.3294	0.0250	0.3656	0.1094	0.6040	0.0000	1.5560	0.0000	1.4554	1.6286	2.2519
		20	0.7595	0.2503	0.0094	0.3550	0.0621	0.5931	0.2392	1.2798	0.0000	0.8486	1.0406	1.1965
		30	0.7707	0.2239	0.0054	0.3450	0.0475	0.5800	0.3516	1.1899	0.0000	0.6584	0.8383	0.8689
		50	0.7797	0.2074	0.0023	0.3298	0.0311	0.5604	0.4492	1.1102	0.0000	0.5275	0.6610	0.6403
		120	0.7874	0.1960	0.0007	0.3058	0.0171	0.5209	0.5882	0.9866	0.0114	0.3806	0.3984	0.3692
0.8	0.8	10	0.7544	0.9447	0.0170	0.1195	0.0924	0.1813	0.0000	1.9492	0.0000	4.8215	2.3896	7.7536
		20	0.7709	0.8818	0.0074	0.0509	0.0612	0.1018	0.0000	1.5666	0.0000	3.4134	1.5915	5.0631
		30	0.7794	0.8339	0.0034	0.0146	0.0423	0.0433	0.0709	1.4880	0.0000	3.0356	1.4172	4.4034
		50	0.7895	0.8129	0.0015	0.0034	0.0286	0.0188	0.2454	1.3335	0.0000	2.4875	1.0882	3.3493
		120	0.7954	0.8020	0.0005	0.0001	0.0162	0.0061	0.4294	1.1613	0.0000	1.9246	0.7319	2.2452
0.8	1.2	10	0.7656	1.2951	0.0123	0.3115	0.0829	0.4969	0.0000	1.8846	0.0000	6.3602	2.2379	10.1302
		20	0.7782	1.2456	0.0047	0.2190	0.0530	0.4456	0.0000	1.6007	0.0000	4.8956	1.6450	7.2999
		30	0.7836	1.2215	0.0030	0.1839	0.0417	0.4215	0.1001	1.4672	0.0000	4.2308	1.3671	6.0185
		50	0.7899	1.2074	0.0016	0.1666	0.0295	0.4074	0.2120	1.3679	0.0000	3.7448	1.1559	5.0749
		120	0.7977	1.2019	0.0006	0.1616	0.0161	0.4019	0.4144	1.1809	0.0000	2.8747	0.7666	3.3455
1.2	0.2	10	1.0979	0.3681	0.0716	0.9974	0.1833	0.9985	0.0000	2.4706	0.0000	1.6792	2.7454	2.6222
		20	1.1382	0.2737	0.0291	0.9649	0.1045	0.9796	0.1866	2.0898	0.0000	1.0082	1.9032	1.4689
		30	1.1517	0.2456	0.0166	0.9307	0.0736	0.9552	0.3509	1.9525	0.0000	0.8232	1.6016	1.1551
		50	1.1698	0.2205	0.0071	0.9000	0.0445	0.9293	0.5554	1.7841	0.0000	0.6329	1.2287	0.8249
		120	1.1895	0.2015	0.0011	0.8088	0.0165	0.8523	0.7850	1.5939	0.0000	0.4566	0.8089	0.5102
1.2	0.8	10	1.1236	0.9709	0.0345	0.1851	0.1362	0.4074	0.0000	2.9337	0.0000	5.1695	3.6203	8.3973
		20	1.1525	0.8715	0.0131	0.1664	0.0845	0.3997	0.0000	2.3854	0.0000	3.5494	2.4657	5.3558
		30	1.1655	0.8375	0.0078	0.1610	0.0661	0.3867	0.1808	2.1501	0.0000	2.9161	1.9693	4.1573
		50	1.1694	0.8025	0.0039	0.1544	0.0509	0.3858	0.4400	1.8987	0.0000	2.2974	1.4586	2.9898
		120	1.1768	0.7926	0.0019	0.1528	0.0355	0.3825	0.7545	1.5990	0.0000	1.6490	0.8444	1.7129
1.2	1.2	10	1.1429	1.3411	0.0237	0.1563	0.1180	0.1590	0.0000	2.8424	0.0000	6.6513	3.3992	10.6205
		20	1.1703	1.2367	0.0094	0.0278	0.0760	0.0514	0.0000	2.4791	0.0000	5.1150	2.6177	7.7565
		30	1.1749	1.2173	0.0059	0.0080	0.0615	0.0291	0.1814	2.1684	0.0000	4.1433	1.9870	5.8521
		50	1.1889	1.2064	0.0032	0.0007	0.0466	0.0174	0.3466	2.0312	0.0000	3.6562	1.6845	4.8996
		120	1.1866	1.1981	0.0015	0.0002	0.0323	0.0109	0.6141	1.7592	0.0000	2.8606	1.1451	3.3250

The simulation results indicate that the mean squared error (MSE), Absolute bias (AB), and width of confidence interval (W_CI) decreases with increasing sample size for all estimation methods, confirming their consistency. The MPS estimator consistently produces the smallest MSE across most parameter configurations, demonstrating superior efficiency and robustness. Although the MLE exhibits relatively high MSE for small samples, its performance improves substantially with increasing sample size. The LSE shows moderate performance but generally remains less efficient than MPS. Therefore, based on the MSE criterion, the MPS estimator is the preferred estimation method for the considered distribution.

6. REAL DATA APPLICATION

In this section, we have considered two real data sets to illustrate the usefulness of the proposed distribution in terms of fitting to the real characteristic. The first data set (Table 4) refer to the failure times of 84 aircraft windshields. This data set studied by [19] and fitted the Weibull-Lomax distribution. The second data set (Table 5) represent the lifetime (in years) of 40 blood cancer (leukemia) patients. Recently, the data set is used by [21] and fitted a new extended Weibull distribution. The TTT plots of the first and second data sets are shown in Figures 4 & 5 respectively. Here, we have compared the performance of the proposed NEER(λ, θ) distribution over some existing well known distributions such as Weibull distribution (WD), Gamma distribution (GD), and Exponentiated Exponential (EE) distribution. For the purpose, the tool used are -log-likelihood, AIC, BIC, K-S test statistic and its p-value. These criteria are defined as follows.

$$\begin{aligned} AIC &= 2k - 2 \ln \hat{L}, \\ BIC &= k \ln n - 2 \ln \hat{L}, \\ KS \text{ test statistic} &= \sup_x |F(x) - F_0(x)|, \end{aligned}$$

where k is the number of parameters and n is the number of units in the data set, $\hat{L}$ is the maximum likelihood, $F(x)$ is the theoretical cdf and $F_0(x)$ is the empirical cdf.

Table 6 & 7 present the estimated parameters for the NEER(λ, θ) distribution along with those for the alternative models. These estimates are obtained through appropriate fitting techniques, providing a basis for comparing how effectively each distribution represents the observed data and captures its underlying patterns.

Table 4: Real Data Set of failure times of 84 Aircraft Windshield

0.040	1.866	2.385	3.443	0.301	1.876	2.481	3.467	0.309	1.899	2.610	3.478
0.557	1.911	2.625	3.578	0.943	1.912	2.632	3.595	1.070	1.914	2.646	3.699
1.124	1.981	2.661	3.779	1.248	2.010	2.688	3.924	1.281	2.038	2.820	4.035
1.281	2.085	2.890	4.121	1.303	2.089	2.902	4.167	1.432	2.097	2.934	4.240
1.480	2.135	2.962	4.255	1.505	2.154	2.964	4.278	1.506	2.190	3.000	4.305
1.568	2.194	3.103	4.376	1.615	2.223	3.114	4.449	1.619	2.224	3.117	4.485
1.652	2.229	3.166	4.570	1.652	2.300	3.344	4.602	1.757	2.324	3.376	4.663

Table 5: The lifetime (in years) of 40 blood cancer (leukemia) patients.

0.315	0.496	0.616	1.145	1.208	1.263	1.414	2.025	2.036	2.162	2.211	2.370
2.532	2.693	2.805	2.910	2.912	3.192	3.263	3.348	3.348	3.427	3.499	3.534
3.767	3.751	3.858	3.986	4.049	4.244	4.323	4.381	4.392	4.397	4.647	4.753
4.929	4.973	5.074	5.381								

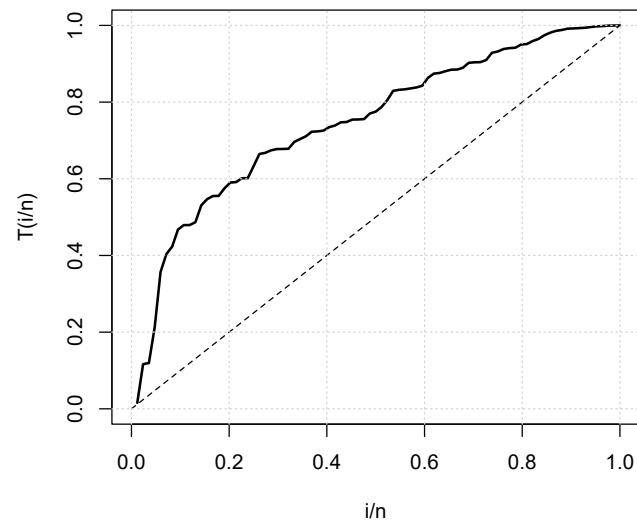


Figure 4: TTT plots for the failure times of 84 Aircraft Windshield

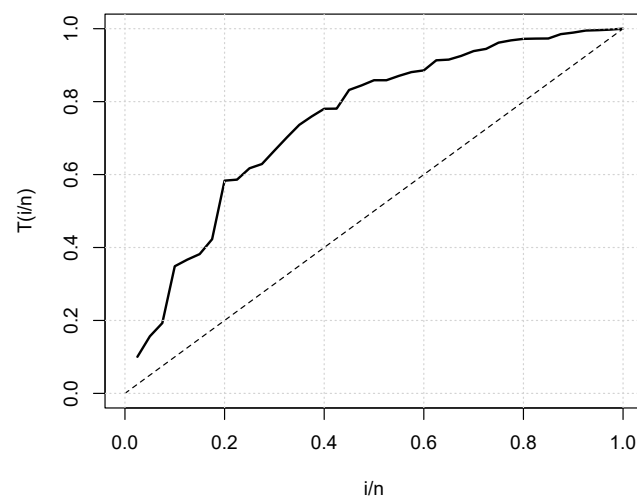


Figure 5: TTT plots for the lifetime (in years) of 40 blood cancer (leukemia) patients.

Table 6: Estimates, $-\log \hat{L}$, AIC, BIC, KS-statistic and p-value for all the considered distributions based on the data of failure times of 84 Aircraft Windshield.

Models	Estimates		$-\log(\hat{L})$	AIC	BIC	KS-statistic	p-value
	$\hat{\lambda}$	$\hat{\theta}$					
NEER(λ, θ)	0.17729	2.13650	128.376	260.751	265.613	0.08159	0.60181
WD(λ, θ)	2.86286	2.37437	130.052	264.105	268.966	0.05368	0.95815
GD(λ, θ)	3.49222	1.36553	136.935	277.871	282.732	0.10351	0.30759
EE(λ, θ)	0.75793	3.56051	139.839	283.678	288.540	0.12097	0.15750

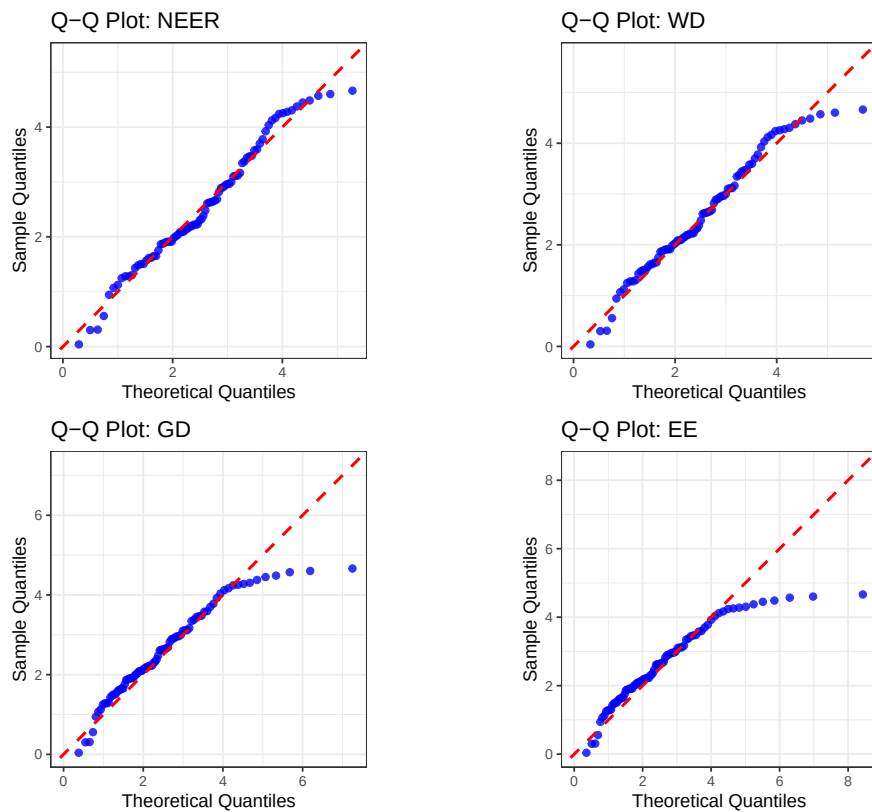


Figure 6: Q-Q plots for the fitted distributions of data set 4.

Table 7: Estimates, $-\log \hat{L}$, AIC, BIC, KS-statistic and p-value for all the considered distributions based on the data of the lifetime (in years) of 40 blood cancer (leukemia) patients.

Models	Estimates		$-\log(\hat{L})$	AIC	BIC	KS-statistic	p-value
	$\hat{\lambda}$	$\hat{\theta}$					
NEER(λ, θ)	0.23253	0.56066	67.5249	139.050	142.428	0.09352	0.84330
WD(λ, θ)	3.51846	2.49949	69.5580	143.116	146.494	0.11834	0.58839
GD(λ, θ)	3.46475	1.10317	73.5487	151.097	154.475	0.15815	0.24283
EE(λ, θ)	0.61415	3.51889	74.9624	153.925	157.303	0.16116	0.22440

From Table 6 & 7, it is clear that the proposed distribution fits better as compared to the other considered distributions in terms of the considered model selections criterion $-\log \hat{L}$, AIC and

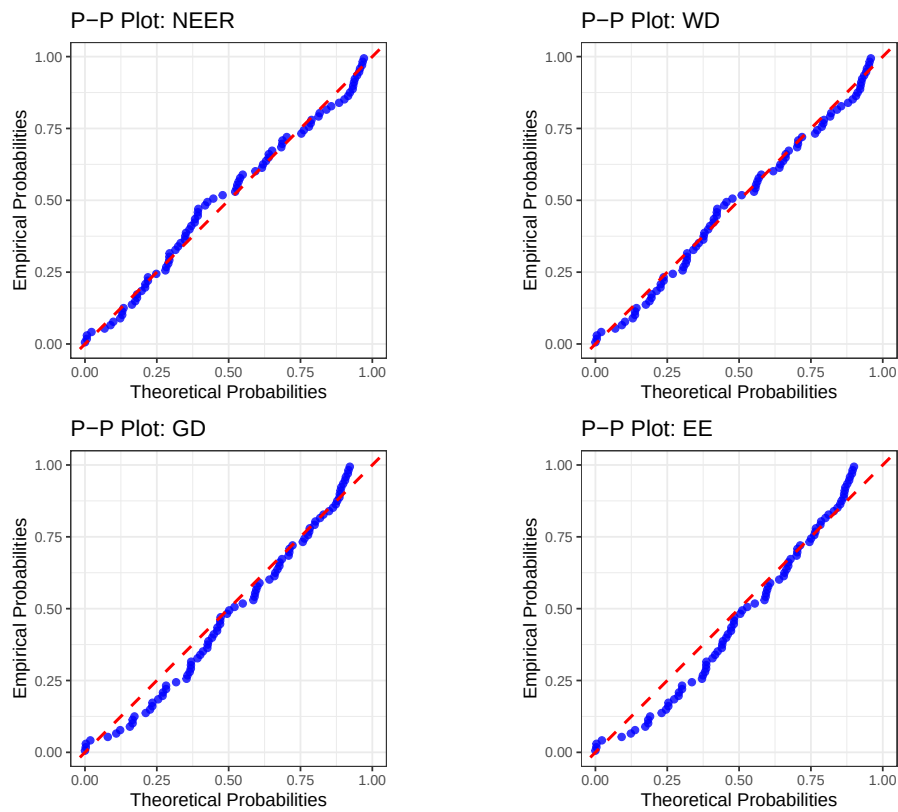


Figure 7: *P–P plots for the fitted distributions of data set 4.*

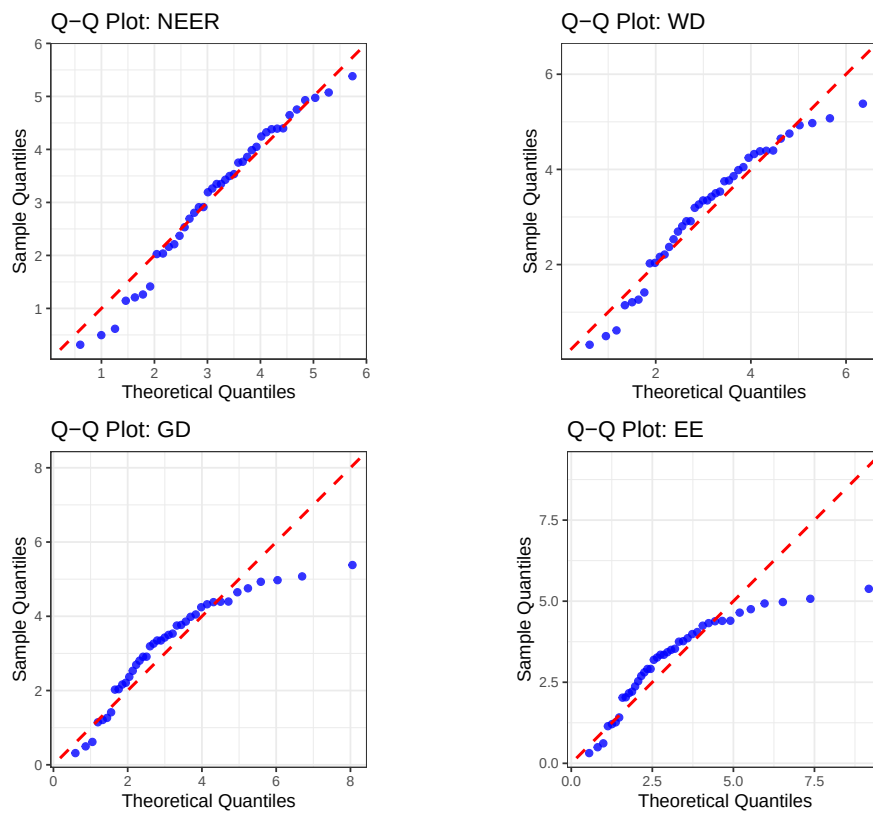


Figure 8: *Q–Q plots for the fitted distributions of data set 5.*

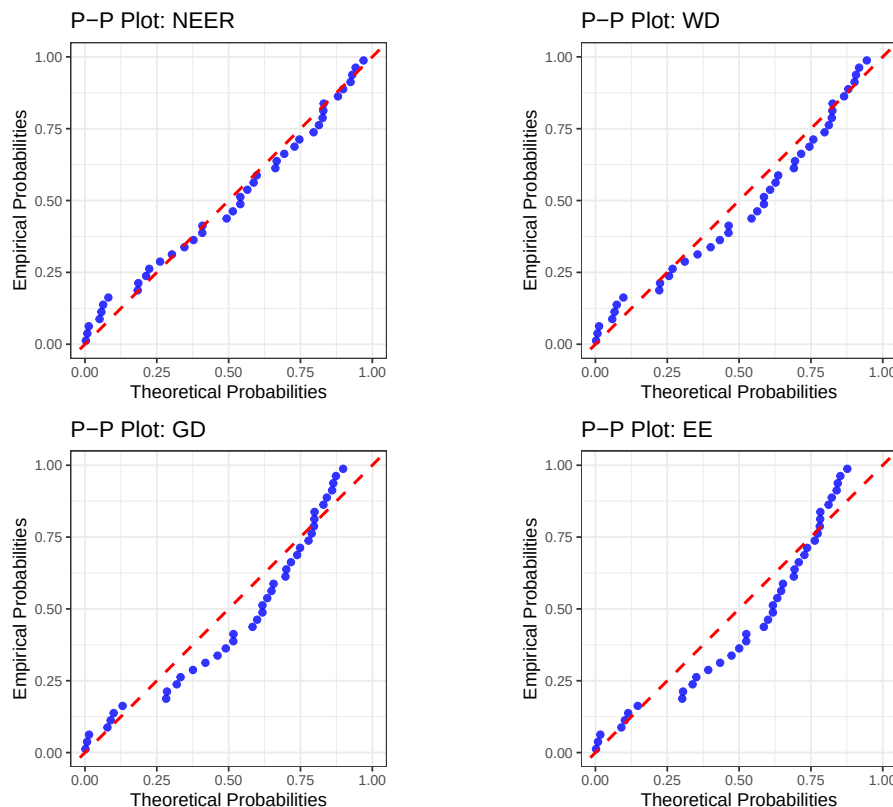


Figure 9: *P-P plots for the fitted distributions of data set 5.*

BIC, as the criterion values are smaller for the proposed $NEER(\lambda, \theta)$ distribution compared to those of for the other considered distributions.

7. CONCLUSIONS

In this study, we have introduced new extended exponential Rayleigh distribution with the parameters λ and θ , i.e., $NEER(\lambda, \theta)$ and we have derived its several key statistical properties such as survival function, hazard rate function, quantile function and order statistics. The distributional behavior is shown by plotting its pdf, cdf and hazard rate function for various combination values of the parameters λ and θ . The parameters λ and θ were estimated using MLE, LSE, and MPS method of estimation on the basis of complete sample from it. The simulation study is carried out to know the performance of these estimations in terms of MSEs. From Tables [1- 3], it is clear that as the sample size n increases the MSEs of these considered estimation methods decreases for all the considered values of the parameters λ and θ . It was noted that the MPS estimates out performs the corresponding MLE and LSE. Finally the two real data applications depicts that $NEER(\lambda, \theta)$ distribution provides a superior fit as compared to the three considered rival probability distributions showing the effectiveness and practical relevance of our proposed $NEER(\lambda, \theta)$ distribution to the real problems.

STATEMENTS AND DECLARATIONS

Data availability

No primary data are used.

Ethics approval

The authors of this article have not conducted any studies with human participants.

Funding

The authors did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that there is no conflicts of interest regarding the publication of this manuscripts.

REFERENCES

- [1] Abd El-Raouf, M. M. and AbaOud, M. (2023). A novel extension of generalized Rayleigh model with engineering applications. *Alexandria Engineering Journal*, 73:269–283. <https://doi.org/10.1016/j.aej.2023.04.063>
- [2] Abd Elfattah, A. M., Hassan, A. S. and Ziedan, D. M. (2006). Efficiency of maximum likelihood estimators under different censored sampling schemes for Rayleigh distribution. *Interstat*, 1:1–16.
- [3] Alizadeh, M., Cordeiro, G. M., Pinho, L. G. B. and Ghosh, I. (2017). The Gompertz-G family of distributions. *Journal of Statistical Theory and Practice*, 11:179–207. <https://doi.org/10.1080/15598608.2016.1267668>
- [4] Alzaatreh, A., Lee, C. and Famoye, F. (2012). On the discrete analogues of continuous distributions. *Statistical Methodology*, 9(6):589–603. <https://doi.org/10.1016/j.stamet.2012.03.003>
- [5] Alzaatreh, A., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1):63–79. <https://doi.org/10.1007/s40300-013-0007-y>
- [6] Alzaatreh, A., Lee, C. and Famoye, F. (2014). T-normal family of distributions: a new approach to generalize the normal distribution. *Journal of Statistical Distributions and Applications*, 1:1–18. <https://doi.org/10.1186/2195-5832-1-16>
- [7] Ateeq, K., Qasim, T. B. and Alvi, A. R. (2019). An extension of Rayleigh distribution and applications. *Cogent Mathematics & Statistics*, 6(1):1622191. <https://doi.org/10.1080/25742558.2019.1622191>
- [8] Bhat, A. A. and Ahmad, S. P. (2023). An extension of exponentiated Rayleigh distribution: properties and applications. *Thailand Statistician*, 21(1):209–227.
- [9] Bourguignon, M., Silva, R. B. and Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12(1):53–68. [https://doi.org/10.6339/JDS.2014.12\(1\).1210](https://doi.org/10.6339/JDS.2014.12(1).1210)
- [10] Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B*, 45(3):394–403. <https://doi.org/10.1111/j.2517-6161.1983.tb01268.x>
- [11] Cordeiro, G. M., Alizadeh, M. and Marinho, P. R. D. (2016). The type I half-logistic family of distributions. *Journal of Statistical Computation and Simulation*, 86(4):707–728. <https://doi.org/10.1080/00949655.2015.1031233>
- [12] Cordeiro, G. M., Alizadeh, M. and Ortega, E. M. M. (2014). The exponentiated half-logistic family of distributions: properties and applications. *Journal of Probability and Statistics*, 2014(1):864396. <https://doi.org/10.1155/2014/864396>
[https://doi.org/10.1155/2014/864396DigitalObjectIdentifier\(DOI\)](https://doi.org/10.1155/2014/864396DigitalObjectIdentifier(DOI))
- [13] Dey, S. (2009). Comparison of Bayes estimators of the parameter and reliability function for Rayleigh distribution under different loss functions. *Malaysian Journal of Mathematical Sciences*, 3:247–264.

- [14] Hassan, A. S. and Elgarhy, M. (2016). A new family of exponentiated Weibull-generated distributions. *rn*, 51:55–71.
- [15] Kundu, D. and Raqab, M. Z. (2005). Generalized Rayleigh distribution: different methods of estimations. *Computational Statistics & Data Analysis*, 49(1):187–200. <https://doi.org/10.1016/j.csda.2004.05.008>
- [16] Mahmoud, M. A. W. and Ghazal, M. G. M. (2017). Estimations from the exponentiated Rayleigh distribution based on generalized type-II hybrid censored data. *Journal of the Egyptian Mathematical Society*, 25(1):71–78. <https://doi.org/10.1016/j.joems.2016.06.008>
- [17] Merovci, F. (2014). Transmuted generalized Rayleigh distribution. *Journal of Statistics Applications & Probability*, 3(1):9. <https://doi.org/10.18576/jsap/030102>
- [18] Rayleigh, L. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *Philosophical Magazine*, 10(60):73–78. <https://doi.org/10.1080/14786448008626893>
- [19] Tahir, M. H., Cordeiro, G. M., Mansoor, M. and Zubair, M. (2015). The Weibull-Lomax distribution: properties and applications. *Hacetatepe Journal of Mathematics and Statistics*, 44(2):455–474. <http://doi.org/10.15672/HJMS.2014147465>
- [20] Voda, V. G. (2007). A new generalization of Rayleigh distribution. *Reliability: Theory & Applications*, 2(2):47–56.
- [21] Klakattawi, H. S. (2022). Survival analysis of cancer patients using a new extended Weibull distribution. *PLOS ONE*, 17(2):e0264229. <https://doi.org/10.1371/journal.pone.0264229>
- [22] Yadav, P., Kumar, D. and Singh, S. (2026). Inference on inverted exponentiated Rayleigh data from accelerated life testing with hybrid censoring. *Quality Technology & Quantitative Management*, 23(1):47–68. <http://doi.org/10.1080/16843703.2025.2464411>

A FLEXIBLE EXTENSION OF THE POWER FUNCTION DISTRIBUTION: PROPERTIES AND APPLICATIONS

UCHENNA U. UWADI¹, GEORGE UCHECHUKWU², ELEBE E. NWEZZA¹,
VICTOR C. OGBUEHI²

¹Department of Mathematics and Statistics,
Alex Ekwueme Federal University Ndufu-Alike, Nigeria

²Department of Statistics,
Michael Okpara University of Agriculture, Umudike, Nigeria

Abstract

This paper introduces a new extension of the Power Function Distribution, referred to as the Exponentiated Gumbel Power Function (EGuPF) distribution derived from the exponentiated Gumbel family of distributions using the power function as the baseline distribution. The EGuPF distribution extends the classical power function distribution by incorporating additional shape parameters, thereby enabling it to capture J-shaped, reverse J-shaped, and bathtub-shaped hazard rate behaviours. The new distribution is expressed as an infinite linear combination of a power function distribution. The quantile function, moments, incomplete moments, Trimmed-L Moments, mode, Moment of Residual Life (MRL), Moment of Reversed Residual Life (MRRL), entropy, and order statistics of EGuPF distribution are derived. Parameter estimation is carried out using the maximum likelihood estimation (MLE) method, and the performance of the estimators is evaluated through Monte Carlo simulation. Simulation results show that the estimators of the parameters of the EGuPF distribution are consistent, efficient, and reliable. The application of the EGuPF distribution is demonstrated using two real datasets: survival times of acute myelogenous leukaemia patients and monthly tax revenue data. In both cases, EGuPF distribution outperforms competing distributions with the same baseline, achieving the lowest AIC, BIC, and goodness-of-fit statistics. These findings establish the EGuPF distribution as a flexible and robust model for analyzing bounded and skewed data.

Keywords: Power Function Distribution, Exponentiated Gumbel-G, Reliability, Estimation, Simulation

1. INTRODUCTION

The power function distribution (PFD) is a bounded continuous probability distribution that has been widely studied due to its simplicity and flexibility in modeling data constrained within a finite range. The PFD is related to the standard Pareto distribution through an inverse transformation [1]. Several authors have investigated its characterization and structural properties; notable contributions include [2],[3], and [4]. The probability density function (pdf) and cumulative distribution function (cdf) of the PFD, as given by [5], are respectively

$$F(x) = \left(\frac{x}{b}\right)^a \quad (1)$$

and

$$f(x) = \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1}, 0 < x < b, a > 0 \quad (2)$$

Statistical literature is very rich in classical probability distributions such as PFD, exponential, lognormal etc for modelling time to event data. There are instances where datasets arising from physical phenomenon do not follow known classical probability distributions. Hence researchers have come up with different methods of transformations otherwise known as generated family of distributions aimed at making baseline distributions more flexible. Some examples of generated family of distribution include beta-G [6], Kumaraswamy-G [7], Transformed-Transformer [8], Weibull-G [9], T-Normal family [10], exponentiated weibull-G [11], Topp-Leone family [12], exponentiated gumbel G [13], T-Kumaraswamy family [14], Marshal Olkin gumbel-G [15].

Several extensions of the PFD have been proposed to enhance its flexibility for bounded, lifetime, reliability, and survival data. [16] introduced the Transmuted PFD using the quadratic rank transmutation map, providing greater flexibility in the shape of the distribution. [17] proposed the Weibull PFD, using the Weibull-G class whose hazard rate can assume increasing, decreasing, and bathtub shapes, with improved performance in application to two different bathtub lifetime data. [18] derived the Kumaraswamy PFD and investigated its TL-Moments and L-Moments, while exponentiated Weibull power function [19] proposed the Exponentiated Weibull PFD, studied its properties, and fitted it to three-real life datasets.[20] subsequently introduced the Exponentiated Kumaraswamy PFD and investigated the effect of the extra shape parameters on the skewness and kurtosis of the distribution.

Furthermore, [21] introduced the Exponentiated PFD and studied the limiting properties of the pdf, cdf, sf, and hrf, and applied it to two failure time data, while [22] proposed the Exponentiated Generalized PFD, offering greater flexibility, particularly for Type-II censored data. [23] derived the Another Generalized Transmuted PFD, which showed improved fit over its sub-models for component failure-time data. More recently, [24] proposed a Cubic Transmuted PFD based on a second-order transmutation map, further enhancing the shape flexibility of the PFD, and estimated the parameters of the proposed model using five different estimation methods.

In this paper, we introduce an extension of PFD given (1) and (2) using a Exponentiated Gumbel family of distribution proposed by [13].The motivations of the study are: (a) To improve the characteristics of the PFD (b) To derive an extension of PFD that can capture bathtub shaped , J-shaped and reverse J-Shaped hazard rate behaviours (c) Empirical evidence form modelling the survival times, in weeks, of 33 patients suffering from Acute Myelogenous Leukaemia and monthly actual taxes revenue in Egypt from January 2006 to November 2010 show that the proposed extension of PFD provides a better fit when compared with other distribution with the same baseline. The rest of the paper is organized as follows. After compiling this paper write the outline ie the sections of the paper.

2. METHODS

2.1. The Exponentiated Gumbel Power Function Distribution

A family of distribution known as exponentiated Gumbel family of distribution was proposed by [13], using the logit of the cdf of a random variable. The cdf of the exponentiated gumbel family of distribution is given by (3)

$$G(x) = 1 - \left[1 - \exp \left\{ -B \left(\frac{F(x)}{1 - F(x)} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha} \quad x > 0, \alpha, \sigma > 0, -\infty < \mu < \infty \quad (3)$$

The pdf corresponding to (3) is

$$g(x) = \frac{\alpha B f(x) (F(x))^{-\frac{1}{\sigma}-1}}{\sigma (1-F(x))^{-\left(\frac{1}{\sigma}-1\right)}} \exp \left\{ -B \left(\frac{F(x)}{1-F(x)} \right)^{-\frac{1}{\sigma}} \right\} \left[1 - \exp \left\{ -B \left(\frac{F(x)}{1-F(x)} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha-1} \quad (4)$$

Where, $B = \exp\left(\frac{\mu}{\sigma}\right)$ and $F(x)$ and $f(x)$ are arbitrary cdf and pdf of a random variable X often referred to as the baseline.

Choosing the baseline distribution to be PFD given in (1) and (2) and substituting same in (3) and (4) birthed a new member of the exponentiated gumbel family called exponentiated gumbel power function distribution denoted as EGuPF distribution. The cdf and density function of EGuPF distribution are respectively given by (5) and (6)

$$G(x) = 1 - \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha} \quad (5)$$

$$g(x) = \frac{\alpha B a b^a x^{-\left(\frac{a}{\sigma}+1\right)}}{\sigma (b^a - x^a)^{-\left(\frac{1}{\sigma}-1\right)}} \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha-1} \quad (6)$$

$0 < x < b, a, \alpha, \sigma > 0, -\infty < \mu < \infty$

Note: if $\alpha = 1$ (5) reduces to Gumbel power function by [25]. The survival function $S(x)$ and hazard rate $h(x)$ function associated with (5) are respectively

$$S(x) = \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha}$$

and

$$h(x) = \frac{\alpha B a b^a x^{-\left(\frac{a}{\sigma}+1\right)}}{\sigma (b^a - x^a)^{-\left(\frac{1}{\sigma}-1\right)}} \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{-1}$$

While the cumulative hazard rate $H(x)$ and reversed hazard rate $r(x)$ function of EGuPF distribution are

$$H(x) = -\alpha \log \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]$$

and

$$r(x) = \frac{\alpha B a b^a x^{-\left(\frac{a}{\sigma}+1\right)} \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha-1}}{\sigma (b^a - x^a)^{-\left(\frac{1}{\sigma}-1\right)} \left[1 - \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\alpha} \right]}$$

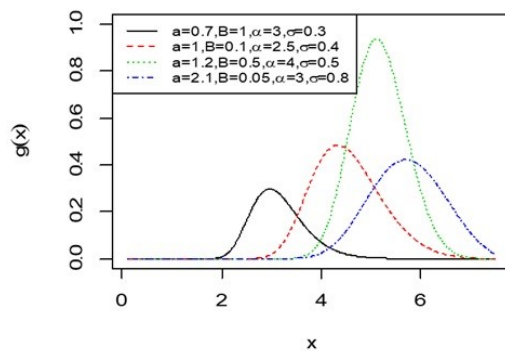


Figure 1: Plot of pdf of EGuPF distribution for selected parameter values.

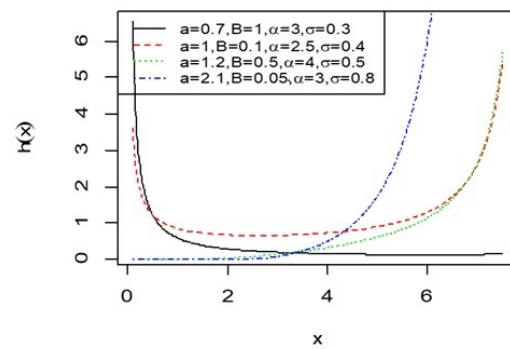


Figure 2: Plot of hrf of EGuPF distribution for selected parameter values.

The plots of the pdf of EGuPF distribution is given Figure 1, which show that EGuPF distribution is unimodal while the plots of its hazard rate function $h(x)$ shown in Figure 2 reveal that it has bathtub shaped, J-shaped and reverse J-shaped. The various shapes assumed by the $h(x)$ makes the proposed distribution fit for modelling data sets whose failure rate assumes the above shapes.

Theorem 1: Given that Y is distributed as exponentiated Gumbel distribution with location parameter μ , scale parameter σ , and shape parameter α , then the random variable $X = b(\exp(-y) + 1)^{-a}$ is distributed as EGuPF distribution with parameters μ, σ, α, a , and b

Proof: Given that

$$r(y) = \frac{\alpha B}{\sigma} \exp\left(-\frac{y}{\sigma}\right) \exp\left\{-B \exp\left(\frac{-y}{\sigma}\right)\right\} \left[1 - \exp\left\{-B \exp\left(\frac{-y}{\sigma}\right)\right\}\right]^{\alpha-1}$$

$$\text{and } x = b(\exp(-y) + 1)^{-a} \text{ implies } y = -\log\left(\left(\frac{b}{x}\right)^a - 1\right)$$

$$\frac{dy}{dx} = \left(\frac{a}{b}\right) \frac{\left(\frac{x}{b}\right)^{-1}}{\left(1 - \left(\frac{x}{b}\right)^a\right)}$$

By transformation,

$$g(x) = r(x) \left| \frac{dy}{dx} \right| \quad (7)$$

Substituting for $\frac{dy}{dx}$ in (7) and simplifying gives

$$g(x) = \frac{\alpha B a \left(\frac{x}{b}\right)^{-\left(\frac{a}{\sigma}+1\right)}}{\sigma b \left(1 - \left(\frac{x}{b}\right)^a\right)^{-\left(\frac{1}{\sigma}-1\right)}} \exp\left\{-B \left(\frac{\left(\frac{x}{b}\right)^a}{1 - \left(\frac{x}{b}\right)^a}\right)^{-\frac{1}{\sigma}}\right\} \left[1 - \exp\left\{-B \left(\frac{\left(\frac{x}{b}\right)^a}{1 - \left(\frac{x}{b}\right)^a}\right)^{-\frac{1}{\sigma}}\right\}\right]^{\alpha-1}$$

This distribution is the same as the one given in (6).

2.2. Quantile Function

The quantile function plays an important role in statistical theory and applications, particularly in simulation studies. Quantile function is defined as

$$u = F(x_u) = P(X \leq x_u), \quad 0 < u < 1.$$

The quantile function $Q(u)$ of the EGuPF distribution is easily obtained by inverting (5) we have

$$Q(u) = b \left[1 + \left\{ -\frac{1}{B} \log \left[1 - (1-u)^{\frac{1}{a}} \right] \right\}^{\sigma} \right]^{-\frac{1}{a}} \quad (8)$$

In particular the median of EGuPF is obtained by substituting $u = \frac{1}{2}$ in (8) as

$$Q(u) = b \left[1 + \left\{ -\frac{1}{B} \log \left[1 - \left(\frac{1}{2} \right)^{\frac{1}{a}} \right] \right\}^{-\sigma} \right]^{-\frac{1}{a}}$$

Random samples from EGuPF distribution can be generated using (8) by assuming that u follows Unifrom distribution

2.3. Important Expansions

Two important representations of the pdf and cdf of EGuPF distribution are derived in this subsection. The general binomial expansion, for $w > 0$ real non integer and $|z| < 1$

$$(1-z)^{w-1} = \sum_{i=1}^{\infty} \binom{w-1}{i} (-1)^i z^i \quad (9)$$

Furthermore, the power series expansion for exponential function $\exp(-z)$ is given by

$$\exp(-z) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} z^j \quad (10)$$

Using (9) and (10) the pdf of EGuPF distribution given in (6) can written as

$$g(x) = \sum_{i,j,k=0}^{\infty} \Psi_{ijk} f_{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}(x) \quad (11)$$

where $\Psi_{ijk} = \frac{(-1)^{i+j+k} \alpha B (B(i+1))^j}{j! \sigma \left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)} \binom{\alpha-1}{i} \binom{\frac{j}{\sigma} + \frac{1}{\sigma} - 1}{k}$ and $f_{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}(x)$ is the pdf of power function distribution.

Hence the EGuPF distribution can be expressed as a linear combination of power function with parameters $a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)$ and b .

Similarly, given that ℓ is a positive integer then using (9) and (10) $G(x)^\ell$ can be expressed as

$$G(x)^\ell = \sum_{m=0}^{\ell} \sum_{p,q,u=0}^{\infty} \omega_{mpqu} F_{a\left(u-\frac{q}{\sigma}\right)} \quad (12)$$

where $\omega_{mpqu} = (-1)^{m+p+q+u} \binom{\ell}{m} \binom{\alpha m}{p} \binom{\frac{q}{\sigma}}{u} (Bq)^q$ and $F_{a\left(u-\frac{q}{\sigma}\right)}$ is the cdf of power function distribution with power parameter $a\left(u-\frac{q}{\sigma}\right)$ and scale parameter b . The expressions for

$g(x)$ and $G(x)^\ell$ given in (11) and (12) respectively are very useful in deriving some properties of EGuPF distribution.

2.4. Properties

In this subsection, we examined some of the statistical and reliability properties of EGuPF distribution.

2.4.1. Moments and Incomplete Moments

The r th moment of a random variable is defined as $\mu_r = \int_{-\infty}^{\infty} x^r g(x) dx$. The r th moment of EGuPF distribution can be obtained using (11) as follows

$$\begin{aligned}\mu_r &= \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \int_0^b x^r f_{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}(x) dx \\ \mu_r &= \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \int_0^b x^r \frac{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}{b} \left(\frac{x}{b}\right)^{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)-1} dx \\ \mu_r &= ab^r \sum_{ijk=0}^{\infty} \Psi_{ijk} \frac{\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}{\left(r+a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)\right)}\end{aligned}\quad (13)$$

The first four central moments can be obtained by setting $r = 1, 2, 3, 4$. The moment generating function (mgf) of EGuPF distribution is given by

$$M_X(t) = E\left(e^{tX}\right) = \sum_{r=0}^{\infty} \frac{t^r}{r!} E(X^r) \quad (14)$$

Substituting (13) in (14)

$$M_X(t) = a \sum_{r=0}^{\infty} \sum_{ijk=0}^{\infty} \Psi_{ijk} \frac{(tb)^r \left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}{r! \left(r+a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)\right)}$$

Second, the incomplete moment, say $\phi_s(z)$ of EGuPF distribution is obtained using (11) as

$$\begin{aligned}\phi_s(z) &= \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \int_0^z x^s f_{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}(x) dx \\ \phi_s(z) &= a \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \frac{\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right) z^{s+a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}}{\left(s+a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)\right) b^{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}}\end{aligned}\quad (15)$$

The upper incomplete moment of the proposed distribution defined using the pdf given in (11) we have

$$\gamma_s(z) = \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \int_z^b x^s f_{a\left(k-\frac{j}{\sigma}-\frac{1}{\sigma}\right)}(x) dx$$

The upper incomplete moment of EGuPF distribution takes the form

$$\phi_s(z) = a \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \frac{\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}{b^{a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}} \left[\frac{b^{s+a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)} - z^{s+a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}{\left(s + a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)\right)} \right] \quad (16)$$

The mean deviation from mean and the median deviation from the mean can be easily obtained using the following relations $\sigma_1(x) = 2\mu G(\mu) - 2J(\mu)$ and $\sigma_2(x) = \mu - 2J(M)$ where $J(c) = \int_0^c xg(x) dx$. The mean deviation from the mean and median deviation from the mean of EGuPF distribution respectively are given by

$$\sigma_1(x) = 2\mu G(\mu) - 2a \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \frac{\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right) \mu^{s+a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}{\left(s + a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)\right) b^{a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}$$

and

$$\sigma_2(x) = \mu - 2a \sum_{i,j,k=0}^{\infty} \Psi_{ijk} \frac{\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right) M^{s+a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}{\left(s + a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)\right) b^{a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}$$

2.4.2. Trimmed-L Moments

The trimmed L-moments was introduced by [26] as an extension of L-moments obtained from linear combinations of order statistics after symmetric or asymmetric trimming. The r th Trimmed L-moment $\lambda_r^{t_1, t_2}$ as defined by [26] is

$$\lambda_r^{(t_1, t_2)} = \frac{1}{r} \sum_{k_1=0}^{r-1} (-1)^{k_1} \binom{r-1}{k_1} E(Y_{r+t_1-k_1, r+t_1+t_2}), r = 1, 2 \quad (17)$$

For a symmetric case $t_1 = t_2 = t$, (17) reduces to

$$\lambda_r^{(t)} = \frac{1}{r} \sum_{k_1=0}^{r-1} (-1)^{k_1} \binom{r-1}{k_1} E(Y_{r+t-k_1, r+2t})$$

then using the well-known representation of the expectation of order statistics, we obtain

$$\lambda_r^{(t)} = \frac{1}{r} \sum_{k_1=0}^{r-1} (-1)^{k_1} \binom{r-1}{k_1} \frac{(r+2t)!}{(r+t-k_1-1)!(t+k_1)!} \int_0^b xg(x) G(x)^{r+t+k_1-1} (1-G(x))^{(t+k_1)} dx$$

Expanding the term $(1-G(x))^{t+k_1}$ using the binomial series yields

$$\lambda_r^{(t)} = \frac{1}{r} \sum_{k_1=0}^{r-1} \sum_{d=0}^{t+k_1} (-1)^{k_1+d} \binom{r-1}{k_1} \binom{t+k_1}{d} \frac{(r+2t)!}{(r+t-k_1-1)!(t+k_1)!} \int_0^b xg(x) G(x)^{r+t+k_1+d-1} dx \quad (18)$$

Substituting (11) and (12) in (18) and replacing ℓ in (12) with $r+t+k_1+d-1$, the r th trimmed-L moment of EGuPF distribution takes the following form

$$\lambda_r^{(t)} = \frac{b}{r} \sum_{k_1=0}^{r-1} \sum_{d=0}^{t+k_1} \sum_{m=0}^{r+t+k_1+d-1} \sum_{i,j,k,p,q,u=0}^{\infty} \Psi_{ijk} \omega_{mpqu} (-1)^{k_1+d} \frac{(r+2t)!}{(r+t-k_1-1)!(t+k_1)!} \binom{r-1}{k_1} \binom{t+k_1}{d} \\ \times \left(a \left(u + k - \frac{q}{\sigma} - \frac{j}{\sigma} - \frac{1}{\sigma} \right) + 1 \right)^{-1}$$

The above expression provides the general form of the trimmed-L moments for the EGuPF distribution in terms of convergent infinite series involving the coefficients Ψ_{ijk} and ω_{mpqu} . The population measures of location, scale, skewness and kurtosis can be obtained for $t = 1, r = 1, 2, 3, 4$ respectively.

2.4.3. Mode

The mode of a distribution is the value of x that maximizes the probability density function. It is especially informative for skewed distributions where the mean may be misleading. The mode of EGuPF distribution can be determined by differentiating the (6) with respect to x and equating to zero and its is given as

$$g'(x) = \frac{a-1}{x} - \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} \left[\left(\frac{1}{\sigma} + 1\right) \left(\frac{x}{b}\right)^{-a} + \left(\frac{1}{\sigma} - 1\right) \left(\frac{b^a - x^a}{b^a}\right)^{-1} \right] \\ + \frac{aB}{\sigma b} \frac{\left(\frac{x}{b}\right)^{-\left(\frac{a}{\sigma} + 1\right)}}{\left(\frac{b^a - x^a}{b^a}\right)^{-\left(\frac{1}{\sigma} - 1\right)}} \left[1 - (\alpha - 1) \exp \left\{ -B \left(\frac{b^a - x^a}{x^a}\right)^{\frac{1}{\sigma}} \right\} \left(1 - \exp \left\{ -B \left(\frac{b^a - x^a}{x^a}\right)^{\frac{1}{\sigma}} \right\} \right)^{-1} \right]$$

2.5. Reliability Measures

Reliability measures provide important insights into the ageing properties and failure dynamics of lifetime distributions. This section presents the derivation of some fundamental reliability measures of the EGuPF distribution.

2.5.1. Moment of Residual Life (MRL)

The n th moment of the residual life function denoted by $m_n(t) = E[(x - t)^n / X > t], n = 1, 2, \dots$, completely determines $G(x)$ [27]. The n th moment of residual life of X is given by

$$m_n(t) = \frac{1}{1 - G(t)} \int_t^\infty (x - t)^n dG(x) = E[(X - t)^n / X > t] \quad (19)$$

Using the general binomial expansion, (19) can be written as

$$m_n(t) = \frac{1}{1 - G(t)} \sum_{r=0}^n (-1)^{n-r} \binom{n}{r} t^{n-r} \int_t^\infty x^r dG(x)$$

Hence the MRL of EGuPF takes the following form

$$m_n(t) = \frac{a}{1 - G(t)} \sum_{r=0}^n \sum_{ijk=0}^\infty (-1)^{n-r} \Psi_{ijk} \binom{n}{r} t^{n-r} \frac{\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}{b^a \left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)} \left[\frac{b^{s+a \left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)} - t^{s+a \left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}{\left(s + a \left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)\right)} \right]$$

2.5.2. Moment of Reversed Residual Life (MRRL)

The n th moment of the reversed residual life, defined as $m_n(t) = E[(t - X)^n | X \leq t]$ for $t > 0$ and $n = 1, 2, \dots$, completely determines $G(x)$.

$$M_n(t) = \frac{1}{G(t)} \int_0^t (t - x)^n dG(x)$$

Using (9) MRRL can be expressed as

$$M_n(t) = \frac{1}{G(t)} \sum_{r=0}^n (-1)^r \binom{n}{r} t^{n-r} \int_0^t x^r dG(x)$$

Therefore the MRRL for EGuPF distribution is

$$M_n(t) = \frac{a}{G(t)} \sum_{r=0}^n \sum_{i,j,k=0}^{\infty} (-1)^r \Psi_{ijk} \binom{n}{r} t^{n-r} \frac{\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right) t^{s+a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)}}{b^{a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)} \left(s + a\left(k - \frac{j}{\sigma} - \frac{1}{\sigma}\right)\right)}$$

2.6. Entropy

In this subsection, we derive the Rényi entropy of order $\gamma > 0, \gamma \neq 1$ for the proposed distribution. Rényi entropy generalizes Shannon entropy and is particularly useful in reliability, information theory, and uncertainty quantification. The Renyi entropy is defined by [28] as

$$I_{R(\gamma)} = \frac{1}{1-\gamma} \log \int_0^{\infty} g^{\gamma}(x) dx, \gamma > 0, \gamma \neq 1 \quad (20)$$

Theorem 2 : The Renyi entropy of a random variable that is distributed as EGuPF distribution is given by,

$$I_{R(\gamma)} = \frac{1}{1-\gamma} \log \left[b \sum_{ijk=0}^{\infty} \eta_{ijk} \left(ak - \gamma \left[\left(\frac{a}{\sigma} + 1 \right) + \frac{j}{\sigma} \right] \right)^{-1} \right]$$

Proof: From (20) we have that

$$g^{\gamma}(x) = \left(\frac{\alpha B a}{\sigma b} \right)^{\gamma} \frac{\left(\frac{x}{b} \right)^{-\gamma \left(\frac{a}{\sigma} + 1 \right)}}{\left(1 - \left(\frac{x}{b} \right)^a \right)^{-\gamma \left(\frac{1}{\sigma} - 1 \right)}} \exp \left\{ -B \gamma \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \left[1 - \exp \left\{ -B \left(\frac{x^a}{b^a - x^a} \right)^{-\frac{1}{\sigma}} \right\} \right]^{\gamma(\alpha-1)}$$

Applying the general binomial expansion and power series expansion for exponential function to $g^{\gamma}(x)$ we obtain

$$g^{\gamma}(x) = \sum_{ijk=0}^{\infty} \eta_{ijk} \left(\frac{x}{b} \right)^{ak - \left[\gamma \left(\frac{a}{\sigma} + 1 \right) + \frac{j}{\sigma} \right]} \quad (21)$$

$$\text{where } \eta_{ijk} = \left(\frac{\alpha B a}{\sigma b} \right)^{\gamma} \frac{(-1)^{i+j+k}}{j!} \binom{\gamma(\alpha-1)}{i} \binom{\gamma \left(\frac{1}{\sigma} - 1 \right) + \frac{j}{\sigma}}{k} (B(\gamma+i))^j$$

substituting (21) in (20) and integrating we have

$$I_{R(\gamma)} = \frac{1}{1-\gamma} \log \left[b \sum_{ijk=0}^{\infty} \eta_{ijk} \left(ak - \gamma \left[\left(\frac{a}{\sigma} + 1 \right) + \frac{j}{\sigma} \right] \right)^{-1} \right]$$

Furthermore, the q-entropy is defined by

$$H_q(X) = \frac{1}{1-q} \log \left(1 - \int_0^{\infty} g^q(x) dx \right); q > 0, q \neq 1,$$

Hence, the q-entropy of EGuPF distribution is given as

$$H_q(X) = \frac{1}{1-q} \log \left[1 - b \sum_{ijk=0}^{\infty} \eta_{ijk} \left(ak - q \left[\left(\frac{a}{\sigma} + 1 \right) + \frac{j}{\sigma} \right] \right)^{-1} \right]$$

2.7. Order Statistics

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from the EGuPF distribution with pdf $g(x)$ and cdf $G(x)$, and let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denote the corresponding order statistics. The r th order statistics is given by

$$g_{r;n}(x) = \frac{1}{B(r, n-r+1)} \sum_{v=0}^{n-r} (-1)^v \binom{n-r}{v} g(x) G(x)^{v+r-1} \quad (22)$$

where $B(.,.)$ is Beta function, substituting (11) and (12) into (22) and replacing l with $v+r-1$, and simplify, the order statistics of EGuPF distribution takes the form

$$g_{r;n}(x) = \frac{1}{B(r, n-r+1)} \sum_{v=0}^{n-r} \sum_{m=0}^{v+r-1} \sum_{ijk=0}^{\infty} \sum_{pqu=0}^{\infty} \Phi^* \left(\frac{a \left(k - \frac{j}{\sigma} - \frac{1}{\sigma} \right)}{b} \right) \left(\frac{x}{b} \right)^{a \left(k + u - \frac{q}{\sigma} - \frac{j}{\sigma} - \frac{1}{\sigma} \right) - 1}$$

$$\text{where } \Phi^* = (-1)^v \binom{n-r}{v} \Psi_{ijk} \omega_{mpqu}$$

2.8. Maximum Likelihood Estimation

Let $X_1, X_2, \dots, X_n$ be a random sample of size n drawn from the proposed EGuPF distribution with parameter vector $\Lambda = (\mu, \alpha, a, b, \sigma)$ and pdf given in (6); the unknown parameters are estimated using the maximum likelihood method. The log-likelihood function l for the complete sample, corresponding to the parameter vector Λ is given by

$$l = n \log \alpha + n \log B + n \log a - n \log \sigma - n \log b - \left(\frac{a}{\sigma} + 1 \right) \sum_{i=1}^n \log \left(\frac{x_i}{b} \right) + \left(\frac{1}{\sigma} - 1 \right) \times \sum_{i=1}^n \log \left(1 - \left(\frac{x_i}{b} \right)^a \right) - B \sum_{i=1}^n Z_i + (\alpha - 1) \sum_{i=1}^n \log [1 - \exp \{-BZ_i\}] \quad (23)$$

where

$$Z_i = \left(\frac{\left(\frac{x_i}{b} \right)^a}{1 - \left(\frac{x_i}{b} \right)^a} \right)^{-\frac{1}{\sigma}}, \quad \text{The sample maxima is the estimator of } b, \text{ i.e., } \hat{b} = X_{(n)}.$$

The associated score function for the likelihood function in (23) are:

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \log (1 - \exp \{-BZ_i\}) \\ \frac{\partial l}{\partial \mu} &= \frac{n}{\sigma} - \frac{1}{\sigma} B \sum_{i=1}^n Z_i + (\alpha - 1) \sum_{i=1}^n \left[\frac{Z_i B \exp(-BZ_i)}{\sigma [1 - \exp(-BZ_i)]} \right] \\ \frac{\partial l}{\partial a} &= \frac{n}{a} - \frac{1}{\sigma} \sum_{i=1}^n \log \left(\frac{x_i}{b} \right) Z_i - \left(\frac{1}{\sigma} - 1 \right) \sum_{i=1}^n \frac{x_i^a \log \left(\frac{x_i}{b} \right)}{(b^a - x_i^a)} + \frac{b^a B}{\sigma} \sum_{i=1}^n \frac{Z_i \log \left(\frac{x_i}{b} \right)}{(b^a - x_i^a)} \\ &\quad + \frac{b^a B}{\sigma} (\alpha - 1) \sum_{i=1}^n \left\{ \frac{Z_i \log \left(\frac{x_i}{b} \right) \exp \{-BZ_i\}}{(b^a - x_i^a)(1 - \exp \{-BZ_i\})} \right\} \\ \frac{\partial l}{\partial \sigma} &= -\frac{n}{a} + \frac{a}{\sigma^2} \sum_{i=1}^n \log \left(\frac{x_i}{b} \right) - \frac{1}{\sigma^2} \sum_{i=1}^n \log \left(1 - \left(\frac{x_i}{b} \right)^a \right) - \frac{B}{\sigma^2} \sum_{i=1}^n Z_i \log \left(\frac{x_i^a}{b^a - x_i^a} \right) \\ &\quad + \frac{B}{\sigma^2} (\alpha - 1) \sum_{i=1}^n \left\{ \frac{Z_i \exp \{-BZ_i\}}{(1 - \exp \{-BZ_i\})} \log \left(\frac{x_i^a}{b^a - x_i^a} \right) \right\} \end{aligned}$$

The system of equations above obtained from the partial derivatives of (23) with respect to the individual parameters does not admit explicit closed-form solutions. Therefore, the system of nonlinear equations is solved numerically using iterative optimisation algorithms like the BFGS algorithm or the Newton–Raphson method to produce the maximum likelihood estimates (MLEs) of the parameters.

3. RESULTS

In this section, we present results from simulation studies alongside the application of the EGuPF distribution to two real dataset.

3.1. Simulation Studies

A simulation experiment is conducted to ascertain the consistency and stability of the maximum likelihood estimators (MLEs) of the parameters in the proposed distribution. The simulation is repeated $N = 5000$ times, each having sample sizes $n = 200, 500, 800, 1200$ and 1500 for parameter values $\mu = 1.5, a = 1.2, \alpha = 1.2$ and $\sigma = 2$. For the purpose of evaluation, the following five measures were computed: mean estimates (ME), average bias (AVB), root mean squared error (RMSE), coverage probability (CP) and average confidence width (ACW). The estimated MEs, AVBs and RMSEs are given by

$$\begin{aligned} \text{ME} &= \frac{1}{N} \sum_{i=1}^N \Lambda_i, \\ \text{AVB} &= \frac{1}{N} \sum_{i=1}^N (\Lambda_i - \Lambda), \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (\Lambda_i - \Lambda)^2}, \end{aligned}$$

for $\Lambda = (\mu, \alpha, a, \sigma)$. Coverage probability (CP) is the percentage of intervals that contain the true value of Λ , while average confidence width (ACW) is the mean of the absolute value of the 95% confidence interval for $\hat{\Lambda}$. The Table 1 show the results of simulation studies.

Table 1: Simulation results for parameter values, $\mu = 1.5$, $a = 1.5$, $\alpha = 1.2$ and $\sigma = 2$.

Parameter	Sample size	ME	AVB	RMSE	ACW	CP
μ	200	2.7628	1.2628	3.0306	18.2354	0.9855
	500	1.9946	0.4946	1.8543	7.6529	0.9730
	800	1.7054	0.2054	1.1482	4.5767	0.9510
	1200	1.5120	0.0120	0.6582	2.7127	0.9305
	1500	1.4687	-0.0312	0.6353	2.2169	0.8950
α	200	2.6483	1.4483	2.4793	10.7388	1
	500	2.0852	0.8852	1.6453	5.2920	1
	800	1.8633	0.6632	1.2034	3.5617	0.9995
	1200	1.7302	0.5301	0.8749	2.5347	0.999
	1500	1.6882	0.4882	0.8134	2.1968	0.996
a	200	7.3815	5.8815	17.701	99.6582	0.9935
	500	3.9075	2.4075	9.8804	30.1218	0.985
	800	2.6579	1.1579	5.0372	12.0799	0.975
	1200	2.0629	0.5629	2.3151	5.5437	0.9595
	1500	1.9852	0.4851	2.5876	4.8543	0.9585
σ	200	3.4542	1.4542	2.9918	13.6400	0.9985
	500	2.7634	0.7634	2.0373	8.15045	0.9955
	800	2.4869	0.4869	1.6086	6.1649	0.9900
	1200	2.3228	0.3228	1.2344	4.7983	0.9845
	1500	2.2601	0.2601	1.1479	4.2350	0.977

Table 1 reports the simulation outcomes for the parameters μ , α , a , and σ under increasing sample sizes. The results show a consistent pattern: as sample size increases, the mean estimates approach their true values, the absolute bias and RMSE decrease, and the confidence intervals become narrower, reflecting improved accuracy and efficiency of the estimators.

3.2. Applications

The proposed EGuPF distribution is compared with the following probability distributions that have the same baseline namely, Transmuted PFD [16], Another generalized Transmuted PFD [23], PFD, Weibull PFD [17], and exponentiated weibull PFD [19].

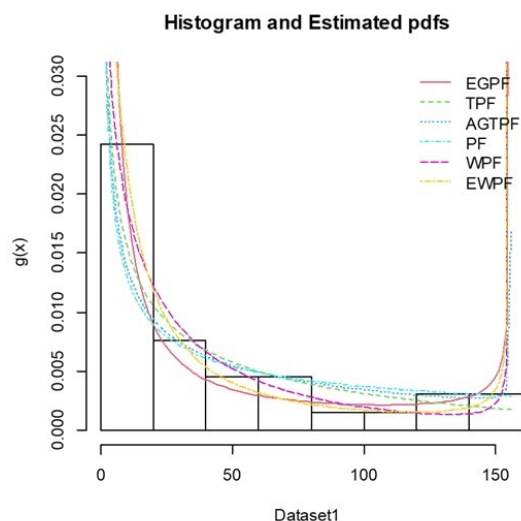
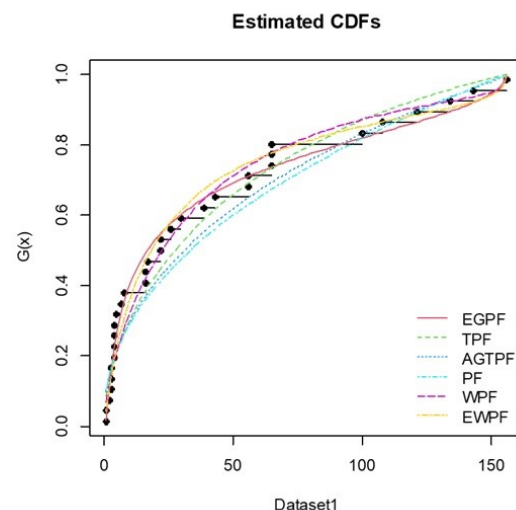
The first data (Dataset 1) were first analyzed by [29]. The data represent the survival times, in weeks, of 33 patients suffering from Acute Myelogenous Leukaemia. The data are: 65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65, 56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, and 43. The skewness and Kurtosis of dataset 1 are 1.221 and 3.349 respectively. These values indicates that Dataset 1 is positively skewed and more peaked than the normal distribution. Table 2 gives the ML estimates of parameters of EGuPF distribution and other competing models with the same baseline while Table 3 presents some Statistics measures that facilitates comparions among models. The plots of the estimated pdfs and cdfs for dataset 1 are given in Figures 3 and 4

Table 2: MLEs of EGuPF parameters and other competing models for dataset 1 (standard errors in parenthesis).

Distributions	Estimates				
EGUPF($\mu, \sigma, \alpha, a, b$)	-2.472 (0.625)	0.462 (1.04)	0.214 (0.506)	0.586 (0.345)	156
TPF(λ, a, b)	0.523 (0.287)	0.562 (0.100)	156		
AGTPF(α, λ, a, b)	0.652 (0.18)	0.775 (0.24)	0.484 (0.10)	156	
PF(a, b)	0.450 (0.078)	156			
WPF(α, β, a, b)	10.540 (7.25)	0.068 (0.043)	2.833 (0.618)	156	
EWPF($\theta, \alpha, \beta, a, b$)	11.680 (14.45)	4.708 (1.19)	0.0600 (0.026)	3.898 (2.07)	156

Table 3: Statistics Measures for Dataset1 .

Distributions	AIC	BIC	CVM	AD	KS	-loglik
EGUPF	297.28	303.27	0.043	0.319	0.096	144.6449
TPF	307.34	310.33	0.109	0.892	0.125	151.6718
AGTPF	306.91	311.40	0.155	0.984	0.127	150.4578
PF	307.40	308.89	0.196	1.238	0.144	152.7016
WPF	300.13	304.62	0.074	0.496	0.125	147.0697
EWPF	299.76	305.75	0.065	0.414	0.121	145.8845

**Figure 3:** Plots of Estimated pdf of all competing models for Dataset1.**Figure 4:** Plots of Estimated cdf of all competing models for Dataset1.

The second data (Dataset 2) have been used by [30]. The data represent the monthly actual taxes revenue in Egypt from January 2006 to November 2010. The distribution is highly skewed to the right. The actual taxes revenue data (in 1000 million Egyptian pounds) are: 5.9, 20.4, 14.9, 16.2, 17.2, 7.8, 6.1, 9.2, 10.2, 9.6, 13.3, 8.5, 21.6, 18.5, 5.1, 6.7, 17, 8.6, 9.7, 9.2, 35.7, 15.7, 9.7, 10, 4.1, 36, 8.5, 8, 9.2, 26.2, 21.9, 16.7, 21.3, 35.4, 14.3, 8.5, 10.6, 19.1, 20.5, 7.1, 7.7, 18.1, 16.5, 11.9, 7, 8.6, 12.5, 10.3, 11.2, 6.1, 8.4, 11, 11.6, 11.9, 5.2, 6.8, 8.9, 7.1, 10.8. The values of Skewness and

kurtosis for dataset 2 are 1.08 and 4.25 respectively. The dataset 2 is positively skewed and leptokurtic. Table 4 gives the ML estimates of fitted distributions while Table 5 provides the different statistics measures. Figures 5 and 6 shows the plots of estimated pdfs and cdfs of fitted distributions.

Table 4: MLEs of EGuPF parameters and other competing models for dataset2 (standard errors in parenthesis)

Distributions	Estimates				
EGUPF($\mu, \sigma, \alpha, a, b$)	-10.782 (7.04)	4.299 (4.09)	1.32 (0.93)	8.112 (5.94)	36
TPF(λ, a, b)	0.894 (0.09)	1.168 (0.12)	36		
AGTPF(α, λ, a, b)	1.471 (0.32)	0.812 (0.13)	1.372 (0.17)	36	
PF(a, b)	0.827 (0.11)	36			
WPF(α, β, a, b)	5.405 (0.95)	0.041 (0.03)	44.049 (36.35)	36	
EWPF($\theta, \alpha, \beta, a, b$)	72.392 (81.12)	8.370 (1.11)	0.032 (0.02)	14.713 (10.35)	36

Table 5: Statistics Measures for Dataset2 .

Distributions	AIC	BIC	CVM	AD	KS	-loglik
EGUPF	373.67	381.98	0.033	0.205	0.056	182.8351
TPF	415.48	419.64	0.887	4.793	0.195	205.7399
AGTPF	414.59	420.83	0.602	3.599	0.176	204.2999
PF	432.66	434.74	1.731	8.443	0.298	215.3321
WPF	393.39	399.62	0.291	1.893	0.143	193.6944
EWPF	375.74	384.05	0.071	0.399	0.081	183.8702

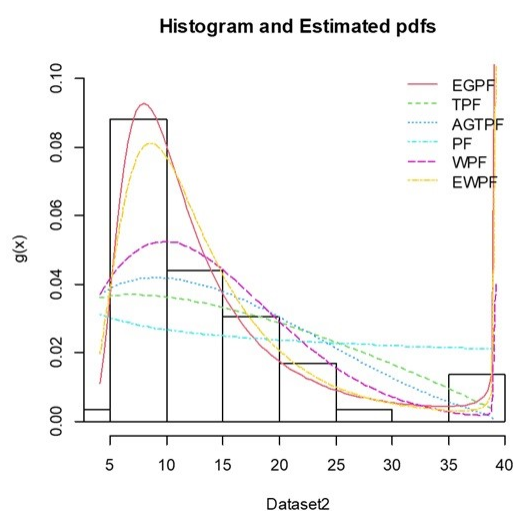


Figure 5: Plots of Estimated pdf of all competing models for Dataset2.

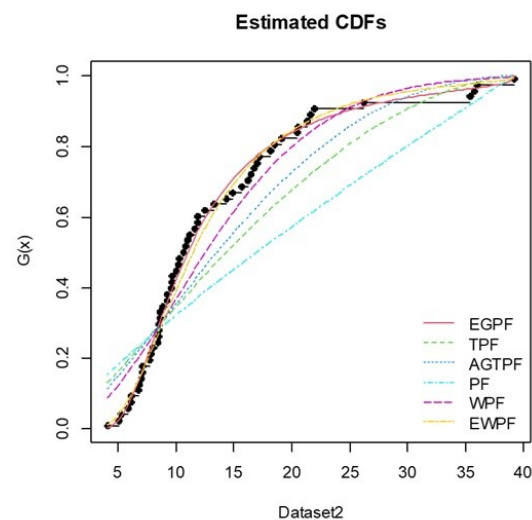


Figure 6: Plots of Estimated pdf of all competing models for Dataset2.

3.3. Discussion

The findings from the simulation analysis provide strong empirical support for the asymptotic behaviour of the maximum likelihood estimators (MLEs) of the parameters of the proposed EGuPF distribution. As the sample size increases from $n = 200$ to $n = 1500$, a clear and systematic pattern is observed across all parameters (μ, α, a, σ). In particular, the mean estimates progressively approach their true parameter values, demonstrating the consistency of the estimators. This convergence behaviour is evident for all parameters, but is especially pronounced for μ , where the estimate decreases from 2.7628 at $n = 200$ to 1.4687 at $n = 1500$, closely approaching the true value of 1.5.

Furthermore, the average bias (AVB) and root mean square error (RMSE) decrease monotonically as the sample size increases, thereby confirming the efficiency and stability of the estimators. The reduction in bias indicates that the estimators become increasingly unbiased in large samples, while the decline in RMSE across all parameters of the proposed model reflects improved estimation precision. In addition, the average confidence width (ACW) decreases consistently with increasing sample size, suggesting enhanced precision of the interval estimates, whereas the coverage probabilities (CP) remain close to the nominal 95% level across different sample sizes. Collectively, these results demonstrate that the estimators of the parameters of the EGuPF distribution are consistent, efficient, and reliable.

The practical applicability and flexibility of the proposed EGuPF distribution are further demonstrated using two real datasets. Dataset 1, which consists of survival times of patients with acute myelogenous leukaemia, exhibits positive skewness and leptokurtic behaviour. The maximum likelihood estimates of the parameters for the EGuPF distribution and competing models are presented in Table 2, with the corresponding standard errors given in parentheses. The estimates of the parameters of EGuPF distribution given in the first row of Table 1 reflect the characteristics of the observed survival pattern, which is concentrated toward shorter survival times with a positively skewed right tail. The negative estimate of $\hat{\mu}$ is consistent with the location of the fitted distribution toward the lower survival-time region, while the relatively small $\hat{\sigma}$ contributes to the concentration and shape of the observation. The estimate $\hat{a} = 0.586$ provides the power-function component needed to describe the shape of the survival times, whereas the relatively small $\hat{\alpha} = 0.214$ indicates a substantial exponentiated-Gumbel modification of the baseline power-function distribution, thus providing flexibility for the observed right-skewed pattern. The estimated upper endpoint $\hat{b} = 156$ coincides with the longest observed survival time, indicating that the fitted distribution accommodates the observed range of survival times.

Model comparison statistics for assessing the performance of the EGuPF distribution and competing models are reported in Table 3. The results indicate that the EGuPF model outperforms all competing distributions, as evidenced by the lowest values of the Akaike Information Criterion (AIC = 297.28), Bayesian Information Criterion (BIC = 303.27), Cramér-von Mises (CVM = 0.043), Anderson-Darling (AD = 0.319), and Kolmogorov-Smirnov (KS = 0.096) statistics, in addition to the highest log-likelihood value. This improved performance can be attributed to the additional flexibility introduced by the extra parameters, which enable the model to better capture skewness and tail behaviour.

A similar pattern is observed for Dataset 2, which represents highly right-skewed tax revenue data. For the Egyptian monthly tax-revenue data, the estimates of parameters of EGuPF model reflect a distribution that is more dispersed and positively skewed, with occasional relatively high revenue values. The negative value of $\hat{\mu}$ reflects the location of the fitted transformed distribution. The large estimate of $\hat{\sigma}$ is consistent with the greater spread of the tax-revenue observations and shape of the distribution. Also, the large value of $\hat{a}$ indicates a substantial power-function shape required to represent the data. The estimate $\hat{\alpha} = 1.32 > 1$ suggests

a stronger exponentiated-Gumbel modification of the power-function baseline, providing additional flexibility for the observed right-skewed and leptokurtic behaviour. The estimated upper endpoint $\hat{b} = 36$ coincides with the largest observed monthly tax revenue, indicating that the fitted EGuPF model captures the observed range of the revenue data.

Once again, the EGuPF distribution provides the best fit among all considered models, yielding the smallest values of AIC (373.6703), BIC (381.9804), and goodness-of-fit statistics (CVM, AD, and KS).

The graphical assessments, including plots of the estimated probability density functions (pdfs) and cumulative distribution functions (cdfs) for both datasets, as presented in Figures 3–6, further corroborate the numerical findings reported in Tables 3 and 5. These figures show that the EGuPF distribution closely follows the empirical pdf and cdf more accurately than the competing models in both cases, thereby reinforcing its adequacy and superior fitting capability.

4. CONCLUSION

This study proposes a new extension of the power function distribution, referred to as the EGuPF distribution, based on the exponentiated Gumbel-G family of distributions introduced by [13]. The probability density function of the EGuPF distribution is unimodal, while its hazard rate function exhibits bathtub-shaped, J-shaped, and reverse J-shaped behaviours, demonstrating its flexibility in modelling diverse lifetime data. The density function is expressed as an infinite linear combination of power function distributions, which enhances its analytical tractability. Several important statistical and reliability properties of the proposed distribution are investigated, including the quantile function, moments, trimmed L-moments, mode, mean residual life (MRL), and mean reversed residual life (MRRL). The maximum likelihood estimators of the model parameters do not admit closed-form expressions; therefore, they are obtained numerically using suitable optimization routines. Results from the simulation study indicate that the estimators are consistent, efficient, and stable across different sample sizes. Applications to two real datasets demonstrate the practical usefulness of the proposed model, as the EGuPF distribution outperforms competing distributions with the same baseline in terms of goodness-of-fit measures.

REFERENCES

- [1] Dallas, A. C. (1976). Characterizing the Pareto and Power Distributions. *Annals of the Institute of Statistical Mathematics*, 28(1), 491–497. <https://doi.org/10.1007/BF02504764>
- [2] Chang, S. K. (2007). Characterizations of the Power Function Distribution by the Independence of Record Values. *Journal of the Chungcheong Mathematical Society*, 20, 139–146.
- [3] Tavangar, M. (2011). Power Function Distribution Characterized by Dual Generalized Order Statistics. *Journal of the Iranian Statistical Society*, 10, 13–27.
- [4] Ahsanullah, M., Shakil, M., and Golam-Kibria, B. M. G. (2013). A Characterization of the Power Function Distribution Based on Lower Records. *ProbStat Forum*, 6, 68–72.
- [5] Meniconi, M., and Barry, D. M. (1996). The Power Function Distribution: A Useful and Simple Distribution to Assess Electrical Component Reliability. *Microelectronics Reliability*, 36(9), 1207–1212. [https://doi.org/10.1016/0026-2714\(95\)00053-4](https://doi.org/10.1016/0026-2714(95)00053-4)
- [6] Eugene, N., Lee, C., and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics–Theory and Methods*, 31, 497–512. <https://doi.org/10.1081/STA-120003130>

- [7] Cordeiro, G. M., and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81, 883–893. <https://doi.org/10.1080/00949650903530745>
- [8] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63–79. <https://doi.org/10.1007/s40300-013-0007-y>
- [9] Bourguignon, M., Silva, R. B., and Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12, 53–68. [https://doi.org/10.6339/JDS.2014.12\(1\).1210](https://doi.org/10.6339/JDS.2014.12(1).1210)
- [10] Alzaatreh, A., Lee, C., and Famoye, F. (2014). T-normal family of distributions: A new approach to generalize the normal distribution. *Journal of Statistical Distributions and Applications*, 1, 1–18. <https://doi.org/10.1186/2195-5832-1-16>
- [11] Hassan, A. S., and Elgarhy, M. (2016). A new family of exponentiated Weibull-generated distributions. *International Journal of Mathematics and its Applications*, 4, 135–148.
- [12] Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S., and Shahbaz, M. Q. (2016). Topp-Leone family of distributions: Some properties and application. *Pakistan Journal of Statistics and Operation Research*, 12, 443–451. <https://doi.org/10.18187/pjsor.v12i3.1458>
- [13] Uwadi, U. U., Okereke, E. W., and Omekara, C. O. (2019). Exponentiated Gumbel Family of Distributions: Properties and Applications. *International Journal of Basic Science and Technology*, 5, 100–117.
- [14] Osatohanmwen, P., Oyegue, F. O., and Ogbonmwan, S. M. (2020). A new family of distributions on the unit interval: The T-Kumaraswamy family. *Journal of Data Science*, 18(2), 218–237. [https://doi.org/10.6339/JDS.202004_18\(2\).0001](https://doi.org/10.6339/JDS.202004_18(2).0001)
- [15] Nwezza, E. E., Ogbuehi, V. C., Uwadi, U. U., and Omekara, C. O. (2020). A new Gumbel-generated family of distributions: Properties, bivariate distribution, and application. *American Journal of Applied Mathematics and Statistics*, 8(1), 9–20.
- [16] Haq, M. A., Butt, N. S., Usman, R. M., and Fattah, A. A. (2016). Transmuted Power Function Distribution. *GU Journal of Science*, 29(1), 177–185.
- [17] Tahir, M. H., Alizadeh, M., Mansoor, M., Cordeiro, G. M., and Zubair, M. (2016). The Weibull-power function distribution with applications. *Hacettepe Journal of Mathematics and Statistics*, 45(1), 245–265. <https://doi.org/10.15672/HJMS.2014428212>
- [18] Abdul-Moniem, I. B. (2017). The Kumaraswamy Power Function Distribution. *Journal of Statistics Applications and Probability*, 6(1), 81–90. <https://doi.org/10.18576/jsap/060107>
- [19] Hassan, A. S., and Nassr, S. D. (2017). The exponentiated Weibull-power function distribution. *Journal of Data Science*, 16, 589–614. [https://doi.org/10.6339/JDS.201710_15\(4\).00002](https://doi.org/10.6339/JDS.201710_15(4).00002)
- [20] Bursa, N., and Özel, G. (2017). The exponentiated Kumaraswamy-power function distribution. *Hacettepe Journal of Mathematics and Statistics*, 46(2), 277–292. <https://doi.org/10.15672/HJMS.2017.411>
- [21] Arshad, M. Z., Iqbal, M. Z., and Ahmad, M. (2020). Exponentiated power function distribution: Properties and applications. *Journal of Statistical Theory and Applications*, 19(2), 297–313. <https://doi.org/10.2991/jsta.d.200514.001>
- [22] Hassan, A. S., and Nassr, S. D. (2020). A new generalization of power function distribution: Properties and estimation based on censored samples. *Thailand Statistician*, 18(2), 215–234.

- [23] Nwezza, E. E., and Uwadi, U. U. (2021). Another generalized transmuted power function distribution for modelling lifetime data. *Journal of Statistics Applications and Probability*, 10(2), 215–425. <https://doi.org/10.18576/jsap/100212>
- [24] Haq, M. A., Aldahan, M. A., Zafar, J., Gomez, H. W., Affify, A. Z., and Mahran, H. A. (2023). A new cubic transmuted power function distribution: Properties, inference, and applications. *PLOS ONE*, 18(2), 1–17. <https://doi.org/10.1371/journal.pone.0281419>
- [25] Uwadi, U. U., Nwezza, E. E., and Okonkwo, C. O. (2024). Properties and Potentials of Gumbel Power Function Distribution to Rainfall and Wind Speed Datasets. *Journal of Applied Sciences and Environmental Management*, 28(5), 1451–1459. <https://doi.org/10.4314/jasem.v28i5.16>
- [26] Elamir, E. A., and Seheult, A. H. (2003). Trimmed L-Moments. *Computational Statistics and Data Analysis*, 43(3), 299–314. [https://doi.org/10.1016/S0167-9473\(02\)00250-5](https://doi.org/10.1016/S0167-9473(02)00250-5)
- [27] Altun, E., Yousof, H. M., Chakraborty, S., and Handique, L. (2018). Zografos–Balakrishnan Burr XII Distribution: Regression Modelling and Applications. *Journal of Mathematics and Statistics*, 19(18), 47–70. <https://doi.org/10.15672/HJMS.2017.410>
- [28] Rényi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 547–561.
- [29] Feigl, P., and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, 21(4), 826–838. <https://doi.org/10.2307/2528247>
- [30] Nassar, M. M., and Nada, N. K. (2011). The beta generalized Pareto distribution. *Journal of Statistics: Advances in Theory and Applications*, 6, 1–17.

FUNDAMENTAL OF MODELING FOR THE ASSESSMENT OF ORGANIZATIONAL RISKS IN THE INTEGRATED SAFETY SYSTEM AT EXPLOSION AND FIRE-HAZARDOUS ENTERPRISES

EVGENIY V. GVOZDEV¹, Alexander I. OVSEYANIK^{2,3}

¹Associate Professor, PhD (Engineering), NRU MGSU, Moscow, Russia
evgvozdev@mail.ru

²Doctor of Technical Sciences, Professor, Department of Civil Defense,
Protection of Population and Territories, Academy of the State Fire Service
(EMERCOM of Russia), Moscow, Russia

³Professor, Department of Life Safety, Financial University under the
Government of the Russian Federation, Moscow, Russia
ovseyanik58@gmail.com

Abstract

The article examines the problem of quantitatively assessing organizational risks within the integrated safety system (ISS) of explosion and fire-hazardous enterprises. An approach is proposed for representing the ISS as an organizational management system, in which organizational risks are considered to be causal factors influencing the probability of hazardous events such as accidents and fires. For a normative evaluation of the contribution of organizational risks, it is proposed to use a conditional information measure, the "conditional bit." It reflects the contribution of a specific risk to a realized hazardous event. The possibility of applying an additive weighted model and an interval scale for ranking organizational risks is substantiated. The practical significance of the results lies in their potential use by the ISS management body while developing measures aimed at reducing organizational risks and the probability of accidents and fires.

Keywords: organizational risks; integrated safety system; conditional bit; rank order centroid; indicator of quantitative evaluation.

The list of abbreviations:

deficiencies	Ds
explosion and fire-hazardous (enterprises)	EFH
fire safety	FS
hazardous event(s)	HE(s)
insufficient actions	IA(s)
insufficient personnel actions	IPA(s)
industrial safety	IS
integrated safety system	ISS
management body of production structural units	MB of PSU
normative documents	NDs
occupational health and safety	OHS
organizational risk(s)	OR(s)
production structural units	PSU
regulatory legal acts	RLAs

I. INTRODUCTION

In the field of organizational risk (OR) assessment, particularly in specific areas (emergency, fire, and occupational risks) where risks are converted from qualitative to quantitative terms, it is proposed to use an ordinal scale as the output data (assessed ORs are ranked in a list in descending order from highest to lowest) [1]. It will allow the identification of OR groups:

- those requiring an immediate response for their prevention;
- those for which there is a time buffer for a preventive response;
- those that can be disregarded in the process of managing the integrated safety system (ISS).

For a reliable assessment of ORs, it is proposed to develop approaches using well-founded (confidence) values on an interval scale to obtain parametric values for the assessment of the negative impact of individual ORs on the outcome in the form of hazardous events (HEs) (an accident and fire) [2]. The use of a ratio scale is justified by the priority of the ranked series for all assessed ORs. It makes it possible to digress from results with point (scalar) values that contain numerous errors, leading to variability in assessment results and reduced comparability of the obtained values when using different initial assumptions. Well-founded values possessing reliability (correctness, fairness) for ORs will be obtained. All of it at explosion and fire-hazardous enterprises (EFH enterprises) will allow the ISS management body to make informed decisions on OR prevention and thus, when using a risk-based approach, achieve new qualitative results [3].

The problematic situation that requires the need to manage ORs which manifest as HEs (accidents and fires) is denoted. A model-based approach is required for the quantitative assessment of ORs within the ISS at EFH enterprises, in order to support management decision-making aimed at preventing such risks.

The object of the study is the integrated safety system which occurs at explosion and fire-hazardous enterprises and is viewed as an organizational management system.

The subject of the study is the models, criteria, and procedures for the quantitative assessment of organizational risks that are taken into account when making management decisions in the ISS.

The purpose of the article is to develop a model-based approach for the quantitative assessment of organizational risks in the ISS of explosion and fire-hazardous enterprises, with the aim of supporting management decisions to reduce these risks.

The scientific novelty of the study is characterized by the development of an approach to the quantitative evaluation of organizational risks within the integrated safety system of explosion and fire-hazardous enterprises, based on a normalized conditional information measure. This approach enables the allocation of the contribution of multiple organizational causes within a single actual hazardous event and the creation of a ranked list of risks to support managerial decision-making.

II. FORMULATING THE OBJECTIVE OF THE ASSESSMENT OF ORGANIZATIONAL RISKS IN THE INTEGRATED SAFETY SYSTEM AT EXPLOSION AND FIRE-HAZARDOUS ENTERPRISES

Organizational risks will be considered as a distinct category of risks associated with insufficient management effectiveness in the ISS or with violations of rules and procedures by the staff responsible for its operations. Within the category of organizational risks R_{OR} , the following new groups are identified:

1. the group considered as insufficient personnel actions (IPAs) in all safety areas (industrial safety (IS), fire safety (FS), occupational health and safety (OHS), and the management body of production structural units (MB of PSU)) while exercising control over personnel of production structural units (PSUs), denoted by R_{IPAs} ;
2. the group considered as deficiencies (Ds) in the implementation of the requirements of regulatory legal acts (RLAs) and normative documents (NDs) by personnel in all areas of safety (IS, FS, OHS, and PSU), denoted by R_{Ds} [2].

There is no universally accepted metric that allows organizational risk to be represented as a quantitative measure of its negative impact on an outcome, such as an accident or fire, in classical information theory and risk theory. However, if information with a qualitative description is generally associated with a bit (the amount of information represented by a binary digit that takes one of two values, “0” or “1”), then it becomes possible to link an information bit to a conditional bit (a measure of the negative impact of organizational risk on an outcome such as an accident or fire). Unlike a classical bit (a value indicating one of two equally probable outcomes), a conditional bit will be represented by a quantitative indicator linked to a qualitative description for each assessed organizational risk. This will reduce the uncertainty of risk groups R_{OR} for given dependency conditions (“Risk” $\rightarrow$ “Accident and Fire”). Such a dependency justifies the use of a conditional bit when describing the connection between organizational risks R_{IPAs} and R_{Ds} and the hazardous events occurring after the realization of these risks (Table 1).

Table 1: Justification for using a conditional bit as a measure of the negative impact of organizational risks on a hazardous event

The connection between organizational risks R_{IPAs} and R_{Ds}	Mathematical interpretation	Benefits
- knowledge of the outcomes of one group of risks with a measure of the uncertainty of its impact on another group of risks; - the ability to consider specific characteristics at EFH enterprises which affect the probability and consequences of risks appearing in HEs; - the ability to measure the severity of negative risk outcomes in HEs.	The ability to relate a conditional bit to entropy in information theory: $1 \text{ cond. bit} = H(R_{OR} C),$ where $H(R_{OR} C)$ is the conditional entropy of risk R_{OR} under condition C; R_{OR} is a qualitative description of the cause of HEs that requires conversion into a quantitative value; C represents conditions taking into account the connections throughout the occurrence and manifestation of R_{OR} in HEs. $H(R_{OR} C) = - \sum_{c \in C} \sum_{r \in R} p(r, c) \log_2 p(r c),$ where $p(r, c)$ is the joint probability of obtaining value r for R_{OR} and value c for condition C; $p(r c)$ is the conditional probability of obtaining value r for R_{OR}, given $C = c$.	- the ability to assess R_{OR} taking into account conditions as they appear and manifest in HEs; - the ability of R_{OR} to obtain a quantitative value that provides a numerical representation of residual uncertainty; - the ability to use methods from information theory to assess R_{OR}, since their compatibility is established.

A mathematical formulation of a scientific problem for converting groups of organizational risks R_{OR} , described qualitatively, into an interval measure (weight), with the possibility of creating a list of risks ranked by level of negative impact, is presented below.

Given:

1. The list of conditions $C = \{c_1, c_2, \dots, c_n\}$, where $n \in \{3, 4\}$ (here and below denotes the set of possible ranks), presented in the accident and fire investigation reports on the Rostekhnadzor website.¹
2. The organizational risk rank $R_{OR} \in \{1, 2, \dots, n\}$ that corresponds to c_n in the report containing the accident and fire investigation materials.

Required:

1. To explain the rationale for using a new metric called the conditional bit to assess organizational risks $R_{OR} \in \{R_{IPAs}, R_{Ds}\}$.
2. To justify the possibility of performing calculations while using organizational risk assessment models in the ISS with $f : R_{OR} \rightarrow w_{R_{OR}}$ that satisfies the condition

$$\sum_{R_{OR}=1}^n w_{R_{OR}} = 1, \quad w_1 > w_2 > \dots > w_n.$$

3. To provide an example that demonstrates the possibility of defining the boundaries of intervals for assessing organizational risks.

¹Open data available at <https://www.gosnadzor.ru/industrial/oil/lessons/index.php>. Information current as of June 8, 2026.

All the results will be used by the management body of the ISS to develop a set of preventive measures.

III. EXPLANATION OF THE EXPEDIENCY OF USING A TIERED MEASURE CALLED CONDITIONAL BIT TO ASSESS ORGANIZATIONAL RISKS

It is proposed to use a level-based measure U_R , referred to as a conditional bit (cond. bit), for all groups of organizational risks as a standardized conditional measure of the contribution of organizational risk to a realized HE. This dimensionless quantity reflects the proportion of the informational-organizational contribution of a specific cause to the resulting negative outcome in the form of cooperative HEs (accidents and fires). The total contribution of all organizational risks listed in the report for a single hazardous event is taken to be equal to 1 cond. bit. Suppose there is a simple event with probabilities for two outcomes P and $q = 1 - P$. It is possible to construct the graph (Figure 1) characterizing the level of uncertainty for emerging ORs represented by U_R , based on the assumption that the maximum value will be directly determined by the weighting factor of the negative impact of the OR itself, as well as its connection with group factors of the OR. Then, the rationale for selecting the level of manifestation of the OR in the HE will be represented by the formula

$$U_R = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) = 1 \text{ cond. bit.} \quad (1)$$

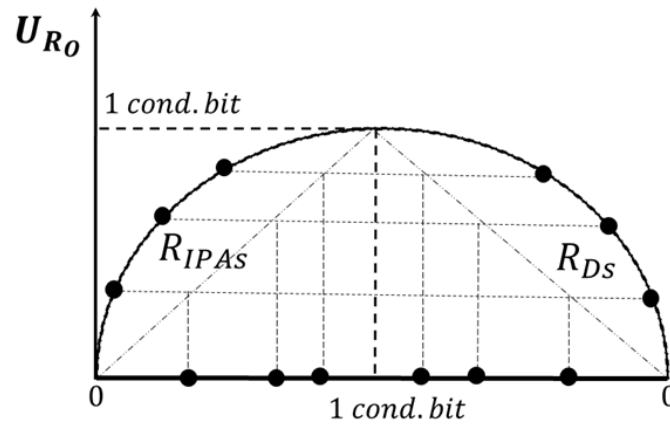


Figure 1: Explanation of the measure in the case of the negative impact of ORs on the outcome expressed as an HE

The graph illustrates a case where, for an event with n discrete outcomes, the value U_R will be maximized when $P_i = 1/n$ and $U_{R_{OR}} = \log_2 n$. Here, it becomes possible to link the level of uncertainty regarding HE information to the presence of organizational risks described in accident reports with qualitative descriptions. In this case, we are interested in information from accident reports, obtained during investigations involving expert consultation.² The level of uncertainty U_x for data taken from accident reports at the Russian oil and gas complex can be calculated using Formula (1), where U_{x1} is the initial information about the OR with a qualitative description and U_{x2} is the final information about the actual HE (accident and fire) [4, 5]. Formally, the sequence of equations justifying the expediency of using a new measure to assess the OR will be as follows:

$$U_R = -1 \log_2(1) - 0 \log_2(0) = 0 \text{ cond. bit,}$$

$$U_{x1} = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{1}{2} \right) = 1 \text{ cond. bit.}$$

²The source is the same as in the preceding footnote.

New information about the sequence of ORs in HEs (accident and fire) is

$$U_{x1} - U_R = 1 - 0 = 1 \text{ cond. bit},$$

while missing information about the HE that has already occurred (accident and fire) is

$$U_{x2} - U_R = 0 - 0 = 0 \text{ cond. bit}.$$

However, when using accident reports at enterprises of the Russian oil and gas complex, the “organizational causes” (organizational risks) section may present not a single cause but several $(2, 3, \dots, n)$ values for ORs. Then there emerges a need to develop models for determining an interval value for each OR listed in the accident report [6]. An example illustrates the possibility of using the new measure (cond. bit) to assess ORs. If P_i is the probability that one of several ORs is the sole cause of an outcome in the form of an HE (accident or fire), then the contribution of the ORs to the HE can be expressed as

$$\sum_{i=1}^n P_i \log_2 \frac{1}{P_i}.$$

Therefore, in the presence of mutually exclusive outcomes, the information associated with classifying the outcome variable for the final HE as risk R_{IPAs} will be $\log_2(1/0.9) = 0.15$ cond. bit, and as risk R_{Ds} will be $\log_2(1/0.1) = 3.32$ cond. bit. The average value confirming the reliability of the results under uncertainty will be $0.9 \cdot 0.15 + 0.1 \cdot 3.32 = 0.47$ cond. bit. Taking permissible errors into account, this indicator is approximated to 0.5 cond. bit, used to assess the two risk components. This result is a reliable weighted average value.

IV. JUSTIFICATION OF A MODEL FOR ASSESSING ORGANIZATIONAL RISKS BASED ON WEIGHTED VALUES ON AN INTERVAL SCALE

In the model for assessing ORs at EFH enterprises in the ISS for the period τ (year), a general, systematic, and comprehensive criterion for the assessment of OR R_{OR}^τ is established. It is an indicator in the form of the HE (accident or fire) which has already occurred, and it is assigned a quantitative value equal to 1 cond. bit. Accordingly, the task consists of determining indicators of the negative impact of ORs (R_{IPAs} and R_{Ds}) presented in open data in the form of accident reports. At the beginning of the study, it was necessary to test various models. The multiplicative model [7] was tested; it is represented by the equation

$$R_{OR}^\tau = \frac{\prod_{i=1}^p (f_i)^{w_i}}{\prod_{j=1}^{n-p} (f_j)^{w_j}} = 1 \text{ cond. bit}, \quad (2)$$

where n is the number of causal factors leading to the occurrence of OR (units); p is the number of ranks connected with the causal factors leading to the occurrence of OR (units); w_i and w_j are the weighted estimates of the causal factors leading to the occurrence of individual ORs (conditional bits); f_i and f_j are value measures of the assessments of causal factors leading to the occurrence of OR (cond. bits); $(f_i)^{w_i}$ belongs to the group of OR assessments with indicators of OR causes equal to 0 cond. bit; and $(f_j)^{w_j}$ belongs to the group of OR assessments with indicators of OR causes equal to 1 cond. bit. However, when using a comprehensive OR assessment criterion of this type, the values of causal factors may exceed one (> 1 cond. bit) because zero values may be considered within the group with acceptable indicators of OR causes. Then R_{OR}^τ will tend toward either infinity or zero, which prevents the condition in Formula (2) from being satisfied. Such a model is not suitable for obtaining individual OR indicators (R_{IPAs} and R_{Ds}).

Next, mixed models [8–10] were selected for testing, for example,

$$R_{OR}^{\tau} = \sum_{i=1}^p w_i f_i - \sum_{j=1}^{n-p} w_j f_j = 1 \text{ cond. bit}, \quad (3)$$

and also

$$R_{OR}^{\tau} = \sum_{i=1}^p w_i f_i + \sum_{j=1}^{n-p} w_j f_j = 1 \text{ cond. bit}. \quad (4)$$

If any assessed causal risk factor (OR) has a value of zero when using these models, it will not be included in the sample since it does not affect the estimates of the remaining ORs. Such models became preferable to the multiplicative model in Formula (2). They made it possible to obtain weighted coefficients reflecting the negative manifestations of ORs in HEs (accidents and fires). However, these models required refinement for the assessment of ORs, since the overall systemic comprehensive OR assessment criterion R_{OR}^{τ} took negative values in some cases, which prevents the condition in Formula (4) from being satisfied. It was considered appropriate to use the model presented as an additive weighted assessment [11]:

$$R_{OR}^{\tau} = \sum_{i=1}^n w_i f_i, \quad \sum_{i=1}^n w_i = 1 \text{ cond. bit}. \quad (5)$$

Here f_j is adjusted according to f_i by normalization of values for the assessed ORs when selecting the maximum value for each of them. If a single OR was specified in an accident report, a measure of 1 cond. bit was assigned to it. Different accident reports typically specify the number of ORs (2, 3, or 4). It was necessary to use specific values to aggregate ORs so that the total additive sample index would be 1 cond. bit (Formula (5)). For accident and fire reports involving two, three, or four organizational risks, it is necessary to determine the order in which weights are assigned, as these weights must differ based on the sequence in which the risks are presented in the general list. The following explanations apply to the first organizational risks:

- based on their contribution to the occurrence of an accident or fire, the first risk in the list is considered the primary risk, and the subsequent risks are considered secondary;
- based on probability, the first risk in the list is the most likely;
- based on the severity of the consequences, the first risk in the list is considered to be the one that caused the most severe consequences in terms of damage;
- based on the time when it occurs, the first risk in the list is the triggering risk, which sets off a chain of subsequent events;
- based on manageability, the first risk in the list is the least controllable.

To justify the choice of an organizational risk assessment model based on weight values on an interval scale with the number of ranks (2–4 units), let us consider risk groups $G_{R_{IPAs}}$ and $G_{R_{Ds}}$ under the condition $R_{OR} \in \{G_{R_{IPAs}}, G_{R_{Ds}}\}$, described by a set of influencing causal risk factors $F = \{f_1, f_2, \dots, f_n\}$, where $n \in \{2, 3, 4\}$. Each causal risk factor f_i has a qualitative description and requires conversion to a quantitative form using an indicator x_i belonging to an interval scale $X \subset \mathbb{R}$ with fixed boundaries of the space $[L, U]$, with selected values $[0, 1]$. To determine the threshold values $\{T_0, T_1, \dots, T_k\}$ that will divide the space $[L, U]$ into non-overlapping intervals, it is proposed to use an aggregated measure for the hazardous-event report $B_{R_{OR}}$. It was necessary to determine weight values $\mathbf{w} = (\omega_1, \omega_2, \dots, \omega_n)$ that will represent the level of negative impact for each causal risk factor. Here, it is necessary to assume the condition $\omega_i > 0$, $\sum_{i=1}^n \omega_i = 1$, where ω_i represents the negative impact indicator for each causal risk factor f_i in the HE report (accidents and fires). Then the aggregated measure will be expressed as

$$B_{R_{OR}} = \sum_{i=1}^n \omega_i x_i, \quad (6)$$

which ensures that the properties of an interval scale are preserved (invariance under linear shifts and additivity of increments).

The selection of interval threshold ranges $\{T_0, T_1, \dots, T_k\}$ under the conditions $L = T_0 < T_1 < \dots < T_k = U$ will correspond to the risk negative-impact level (e.g., “requiring immediate action,”

“manageable over a period of time,” or “negligible”). A methodological approach for justifying the organizational risk assessment model is presented in Table 2.

Table 2: Comparison of models for assessing organizational risks on an interval scale

Approach to risk assessment	Mathematical justification	Expediency of use
Entropic (informational)	$H_j = -\sum_{i=1}^m P_{ij} \log P_{ij}$, where $P_{ij} = x_{ij} / \sum_i x_{ij}$ and $\omega_j = (1 - H_j) / \sum_k (1 - H_k)$	To assess risk factors whose qualitative characteristics are not homogeneous
Pairwise comparisons (expert)	The matrix $A = [a_{ij}]$, $a_{ij} \in [1/9, 9]$; the weight vector $\mathbf{w}$ is the normalized eigenvector of matrix A ; consistency ratio < 0.1	To assess the dominant risk factors that require calibration of indicators to interval values
Regression identification	$\min_{\mathbf{w}} \sum_{k=1}^m (y_k - \sum_{i=1}^n \omega_i x_{ki})^2$, subject to $\sum_i \omega_i = 1$, $\omega_i > 0$; y_k is an integral event-realization coefficient	To assess risk factors based on empirically derived weights
Bayesian belief networks	The initial weights $\mathbf{w}^{(0)}$ are received from experts and then updated according to $\mathbf{w}^{(t)} \propto \mathbf{w}^{(t-1)} L(X_t \mathbf{w})$	To adjust a ranked list of heterogeneous risk factors that have identical weight values

V. AN EXAMPLE ILLUSTRATING THE PRACTICAL APPLICABILITY OF A NEW MODEL FOR ASSESSING ORGANIZATIONAL RISKS

Academician N.N. Moiseev proposed the following definition of a model as a means of cognition [12, p. 166]. He suggested a new model be understood as simplified or “packaged” knowledge that conveyed quite specific, limited information about an object (phenomenon) and reflected certain properties of that object. It is proposed that a new model be regarded as a special form of information encoding. Unlike conventional encoding, where all the source information is known and is translated into another language using established methods, a new model, regardless of the language it employs, encodes information that people have not previously known. In this case, the new model contains potential knowledge that a researcher can acquire, make tangible, and apply in practice. To achieve the objective of this article—to develop a model-based approach to the quantitative assessment of organizational risks in ISS departments at EFH enterprises—it is advisable to use an example to demonstrate the possibility of gaining new knowledge through the results of a risk assessment.

Input data: an expert team consisting of six members had to process the reported results, including an analysis of accident and fire statistics, which were used to assess ORs using a new model. For each accident and fire report, a questionnaire was provided in which respondents were required to answer questions by selecting a specific safety category (IS, FS, OHS, or MB of PSU) that:

1. is directly related to the occurrence and manifestation of ORs in HEs (accidents and fires), and damage was consequently caused to the ISS;
2. is immediately related;
3. is intermediately related;
4. is indirectly related.

Each expert assigned the selected area of safety to the appropriate rank $r \in \{1, 2, 3, 4, 5\}$ according to the established scale. The results of the analysis of individual questionnaire data are presented in a group sample (Table 3).

Table 3: Results of expert data analysis with statistics

Group reports	IS	FS	OHS	MB of PSU
No. 1	5	5	5	1
No. 2	3	4	5	2
No. 3	3	4	6	2
No. 4	5	5	5	1
No. 5	3	5	4	1
No. 6	3	4	5	2
No. 7	3	5	5	1
Total	25	32	33	10
Overall priority	2	4	3	1

Expert concurrence was assessed using Kendall's coefficient of concordance [13, 14], W , by the formula

$$W = \frac{12S}{m^2(n^3 - n)}, \quad (7)$$

where $m = 6$ is the number of experts, $n = 5$ is the number of ranks, and S is the sum of the squares of rank deviations, calculated using the formula

$$S = \sum_{k=1}^j (r_k - \bar{r})^2, \quad (8)$$

where r_k is the sum of the ranks assigned to safety category k , and $\bar{r}$ is the normalized rank index calculated using $\bar{r} = \sum_{k=1}^j r_k / 4$. The calculated data are presented in Table 4.

Table 4: Rank correlation of OR assessment indicators

Safety category	Sum of ranks r_k	$(r_k - \bar{r})^2$
IS	25	$(25 - 25)^2 = 0$
FS	32	$(32 - 25)^2 = 49$
OHS	33	$(33 - 25)^2 = 64$
MB of PSU	10	$(10 - 25)^2 = 225$
Total	100	$S = 338$

Kendall's coefficient of concordance was calculated using Formula (7):

$$W = \frac{12 \cdot 338}{6^2(125 - 5)} = \frac{4056}{4320} \approx 0.93.$$

The resulting figure is considered acceptable, as $W \approx 0.93 > 0.7$.

Since the experts' task focused on assessing the effectiveness of the safety contribution (IS, FS, OHS, and MB of PSU) to the overall condition of the ISS, it was necessary to take into account the following condition: "The higher the risk on the Harrington desirability scale, the smaller the safety function's contribution to the effective functioning of the ISS." A template table was used to assess the effectiveness of the safety contribution (Table 5).

Table 5: *Assessment of the safety contribution to the overall condition of the ISS*

Range d_z	Qualitative assessment of the safety contribution
[0, 2]	unacceptable (critical) contribution
[3, 4]	unsatisfactory contribution
[5, 6]	satisfactory contribution
[7, 8]	good contribution
[9, 10]	excellent contribution

The resulting metrics are inversely proportional to the ranks of organizational risks (r_k). The invertible normalization mapping is

$$\begin{cases} z_k = 0 \text{ (negative)} & \longrightarrow z_k^* = +3 \text{ (maximum negative),} \\ z_k = 10 \text{ (positive)} & \longrightarrow z_k^* = -3 \text{ (minimum negative).} \end{cases} \quad (9)$$

To determine z_k^* , it was necessary to assess the effectiveness of each safety contribution (IS, FS, OHS, and MB of PSU), i.e., to rate them on a 10-point scale $z \in [0, 10]$, where [0, 2] indicates an unacceptable (critical) contribution; [3, 4] indicates an unsatisfactory contribution; [5, 6] indicates a satisfactory contribution; [7, 8] indicates a good contribution; and [9, 10] indicates an excellent contribution.

To normalize the data with an inversion of the level attribute z_k^* , a calculation was performed using the formula

$$z_k^* = 3 - \frac{z_k}{5} \cdot 3. \quad (10)$$

The Harrington function was used to determine the indicators of the negative impact of each safety contribution (IS, FS, OHS, and MB of PSU):

$$x_k = \exp[-\exp(-z_k^*)]. \quad (11)$$

Calculations were performed, and contribution-effectiveness indicators were presented for each safety contribution (IS, FS, OHS, and MB of PSU) (Table 6).

Table 6: *Results of the effectiveness of the safety contribution in the overall condition of the ISS*

Group reports / indicator	IS	FS	OHS	MB of PSU
No. 1	7	9	9	1
No. 2	5	7	9	3
No. 3	3	9	7	1
No. 4	7	9	9	1
No. 5	3	9	7	1
No. 6	2	7	9	3
No. 7	3	9	9	3
Weighted average z_k	4.29	8.43	8.43	1.86
Normalized z_k^*	0.43	-2.06	-2.06	1.89
Safety contribution indicator x_k	0.22	0.88	0.88	0.00
Negative-impact indicator $1 - x_k$	0.78	0.12	0.12	1.00

A mechanism for assessing expert uncertainty through a derivative of the desirability function is presented. It allows the correct use of a level-based criterion z_k while ensuring compatibility with the derivation of indicators x_k (the effectiveness of the safety contribution to the quality of the ISS) and indicators of the negative impact of the safety contribution on the ISS condition, calculated as $(1 - x_k)$.

VI. CONCLUSION

This article proposes a model-based approach to the quantitative assessment of organizational risks within the integrated safety system of explosion and fire-hazardous enterprises. It is shown that organizational risks associated with inadequate personnel actions and deficiencies in compliance with regulatory documents can be considered as factors influencing the occurrence of hazardous events. To assess these risks, a normalized conditional measure (conditional bit) is proposed, which allows for the allocation of the contribution of multiple organizational causes within a single actual hazardous event. Various assessment model options are considered, and the expediency of using an additive weighted model is justified, as it ensures the receipt of comparable indicators of the negative impact of organizational risks. The practical significance of the results lies in the ability to create a ranked list of organizational risks and to determine priorities for management actions within the integrated safety system. The results can be used by the management bodies of integrated safety systems at explosion and fire-hazardous enterprises when developing measures aimed at reducing organizational risks and the probability of accidents and fires.

REFERENCES

- [1] Gvozdev, E.V. Analysis of criteria and methods for assessing organizational risks at explosion and fire-hazardous enterprises. *Occupational Safety in Industry*, 2025, no. 7, pp. 48–55. DOI: [10.24000/0409-2961-2025-7-48-55](https://doi.org/10.24000/0409-2961-2025-7-48-55).
- [2] Gvozdev, E.V. Assessment of organizational risks in the ISS at explosion and fire-hazardous enterprises using the expert method. *Technosphere Safety Technologies*, 2025, no. 2(108), pp. 46–61. DOI: [10.25257/TTS.2025.2.108.46-61](https://doi.org/10.25257/TTS.2025.2.108.46-61).
- [3] Chebotarev, S.S., Ovseyanik, A.I., et al. Risk-oriented approach to substantiating investment decisions for emergency prevention. *Problems of Safety and Emergency Situations*, 2022, no. 3, pp. 13–17. DOI: [10.36535/0869-4176-2022-03-2](https://doi.org/10.36535/0869-4176-2022-03-2).
- [4] Gvozdev, E.V. Development of a method for assessing occupational risks for occupational safety management at electric power enterprises. *Occupational Safety in Industry*, 2026, no. 3, pp. 27–35. DOI: [10.24000/0409-2961-2026-3-27-35](https://doi.org/10.24000/0409-2961-2026-3-27-35).
- [5] Serebrennikov, D.S., Stavratskiy, K.M., Markeev, A.V. Methodology for audit and risk assessment in the field of industrial safety at third-party oil loading terminals. *Oil Industry*, 2017, no. 11, pp. 68–69. DOI: [10.24887/0028-2448-2017-11-68-69](https://doi.org/10.24887/0028-2448-2017-11-68-69).
- [6] Ovseyanik, A.I., Vishnyakova, S.P., Grigoriev, S.M., Lopukhov, A.A. Development of the theory and practice of civil protection economics in the interests of ensuring life safety and safe sustainable industrial development. *Occupational Safety in Industry*, 2024, no. 4, pp. 93–95.
- [7] Orlov, A.I., Tsisarskiy, A.D. Method for risk assessment in the creation of rocket and space technology. *Bulletin of Bauman Moscow State Technical University, Series Mechanical Engineering*, 2017, no. 2(113), pp. 99–107. DOI: [10.18698/0236-3941-2017-2-99-107](https://doi.org/10.18698/0236-3941-2017-2-99-107).
- [8] Zhou, X.H., Zhou, A., Shen, S.L. Novel model for risk assessment of shield tunnelling in soil-rock mixed strata. *Acta Geotechnica*, 2024. DOI: [10.1007/s11440-023-02110-0](https://doi.org/10.1007/s11440-023-02110-0).
- [9] Li, Y., Wang, H., Bai, K., Chen, S. Dynamic intelligent risk assessment of hazardous chemical warehouse fire based on electrostatic discharge method and improved support vector machine. *Process Safety and Environmental Protection*, 2021, vol. 145, pp. 425–434. DOI: [10.1016/j.psep.2020.11.012](https://doi.org/10.1016/j.psep.2020.11.012).

- [10] Luo, Q., Zhang, M. Research on credit risk assessment of listed companies in science and technology sector by introducing industry research report information. *Procedia Computer Science*, 2022, vol. 214, pp. 1317–1324. DOI: [10.1016/j.procs.2022.11.311](https://doi.org/10.1016/j.procs.2022.11.311).
- [11] Ovseyanik, A.I., Kosenok, Yu.N. Modeling the impact of hazardous factors on the “critical facility–environment” system. *Technosphere Safety Technologies*, 2022, no. 3(97), pp. 171–189. DOI: [10.25257/TTS.2022.3.97.171-189](https://doi.org/10.25257/TTS.2022.3.97.171-189).
- [12] Moiseev, N.N. Mathematics in the social sciences. In: *Mathematical Methods in Sociological Research*. Moscow, 1981.
- [13] Shershneva, E.G., Nosova, O.A. Diagnostics of the effectiveness of enterprise digitalization based on the Harrington desirability function. *Digital Models and Solutions*, 2025, vol. 4, no. 1, pp. 5–16. DOI: [10.29141/2949-477X-2025-4-1-1](https://doi.org/10.29141/2949-477X-2025-4-1-1).
- [14] Tarasenko, B.F., Tsybulevsky, V.V., Angelos, N., et al. Comprehensive assessment of small-scale mechanization tools for tillage using the desirability function (Harrington curve). *Polythematic Online Electronic Scientific Journal of the Kuban State Agrarian University*, 2025, no. 207, pp. 79–88. DOI: [10.21515/1990-4665-207-007](https://doi.org/10.21515/1990-4665-207-007).

A SURVIVAL WEIGHTED RESILIENCE FRAMEWORK FOR THE MARSHALL-OLKIN DISTRIBUTION WITH APPLICATIONS IN ENVIRONMENTAL AND BIOMEDICAL MODELING

SHAMEERA T.¹, BINDU P.P.²

•

¹Department of Statistics, MES Mampad College, Malappuram, Kerala, India

shameeranaseerok@gmail.com

² Department of Statistics, Govt. Arts & Science College, Kozhikode, Kerala, India

ppbindukannan@gmail.com

Abstract

This paper develops a survival weighted hazard interpretation of the Marshall–Olkin extension for lifetime distributions. The resulting formulation, termed the Resilience Distribution (RD), introduces a parameter $\theta > 0$ that modulates the baseline hazard through the survival probability, providing a representation of resilience and hazard damping in reliability systems. The framework allows flexible hazard patterns, including increasing, decreasing, and non-monotonic shapes. Key statistical properties, including the density, survival and hazard functions, quantile function, moments, and order statistics, are derived. Theoretical results on stochastic ordering, tail behavior, and hazard elasticity are established. Parameter estimation is performed using maximum likelihood, supported by a simulation study. A special case based on the Lindley distribution is applied to real datasets, demonstrating the flexibility and interpretability of the proposed framework for modeling lifetime data.

Keywords: Resilience modeling, Marshall–Olkin extension, Hazard rate function, Lindley distribution

1. INTRODUCTION

Survival analysis and reliability modeling rely heavily on classical lifetime distributions—such as the exponential, Weibull, Lindley, and log normal models—to describe time to event behavior. Although these distributions provide mathematically tractable foundations, their hazard functions are often structurally limited and may fail to capture the complex failure patterns observed in modern reliability applications. To improve flexibility, numerous generator families have been developed in the literature, including the exponentiated family ([1]), the Alpha-Power Transformation ([2]), and the Marshall–Olkin extension ([3]). These approaches typically enhance flexibility by modifying the scale, shape, or tail behavior of a baseline distribution.

Among these, the Marshall–Olkin framework ([3]) has become one of the most widely used techniques for generating generalized lifetime models. By introducing an additional parameter through a transformation of the baseline distribution function, it produces a rich class of models capable of exhibiting diverse hazard rate behaviors and has been successfully applied in reliability and survival studies.

Despite these advances, an important class of real-world phenomena remains challenging to model adequately: systems that exhibit *resilience*. In such systems, the instantaneous risk of

failure may depend not only on time but also on the accumulated survival experience of the system itself. This behavior arises in engineering components operating under varying stress conditions, biomedical relapse processes, and environmental systems that exhibit temporary recovery before failure. In these situations, the hazard rate evolves dynamically, responding to the current survival state rather than following a purely monotonic trend.

Motivated by this perspective, this study develops a survival-weighted hazard formulation that provides an alternative interpretation of the Marshall–Olkin transformation. Specifically, a resilience parameter $\theta > 0$ is introduced to modulate the baseline hazard through the survival probability. Under an appropriate parameter mapping, the resulting formulation can be expressed as a reparameterized form of the Marshall–Olkin transformation, while offering a distinct interpretation in which the additional parameter directly represents resilience behavior. In this framework, the hazard rate becomes a survival-dependent function, allowing the model to capture adaptive failure dynamics.

From a reliability engineering perspective, this formulation is particularly relevant for modeling systems whose failure behavior depends on accumulated survival experience. Such systems require models in which hazard rates adapt to the system’s operational state, making resilience-based representations essential for realistic analysis.

The main contribution of this work is to provide a survival-weighted hazard interpretation of the Marshall–Olkin framework. Rather than introducing a new distributional form, the proposed approach offers a conceptually distinct perspective in which the additional parameter quantifies resilience and governs how the hazard rate adapts to the survival level of the system. This interpretation establishes a meaningful link between classical lifetime models and resilience-based mechanisms observed in practice.

The proposed formulation retains the analytical tractability of the Marshall–Olkin family, allowing closed-form expressions for the density, distribution, survival function, quantile function, and hazard rate. It also supports a wide range of hazard shapes, including increasing, decreasing, bathtub, and non-monotonic patterns, while preserving the tail behavior of the baseline distribution.

To illustrate its usefulness, a specialization based on the Lindley distribution, referred to as the Resilience Lindley model, is considered. Its statistical properties are derived, and parameter estimation is performed using the maximum likelihood method. A simulation study is conducted to evaluate estimator performance, and the model is applied to real datasets to demonstrate its practical relevance in modeling lifetime data.

Overall, the proposed formulation provides a useful bridge between classical distributional modeling and resilience-based interpretations of hazard dynamics, with potential applications in reliability engineering, environmental studies, and biomedical analysis.

2. THE RESILIENCE DISTRIBUTION (RD)

Classical generator families modify a baseline distribution through algebraic, power, shock, or rank-based transformations of $F(x)$ or $f(x)$. Consequently, the resulting hazard function $h(x)$ changes only indirectly and does not explicitly incorporate the *resilience effect*—the idea that systems with a high conditional survival probability may experience a reduced instantaneous risk of failure.

The *Resilience-based formulation* introduces a mechanism in which the baseline hazard is modulated according to the magnitude of the survival probability. Systems that have survived well up to time x retain resilience and may exhibit a reduced instantaneous hazard, whereas systems with lower survival probability experience weaker damping and relatively higher hazard. This interpretation embeds resilience directly into the hazard structure and allows the model to capture adaptive failure dynamics observed in many practical survival and reliability contexts.

It is important to note that the resulting cumulative distribution function can be expressed as a reparameterized form of the Marshall–Olkin extension ([3]). Under an appropriate parameter mapping, the functional form of the proposed distribution coincides with the classical

Marshall–Olkin transformation. However, the present formulation provides a conceptually distinct interpretation in which the additional parameter represents a resilience effect through survival-weighted hazard modulation. This perspective explicitly links the hazard rate to the survival state of the system, offering an interpretation that is not directly emphasized in the classical Marshall–Olkin formulation.

2.1. Definition

Let $F(x)$ be a baseline cumulative distribution function with probability density function $f(x)$ and survival function $S(x) = 1 - F(x)$. The *Resilience Distribution (RD)* is defined through the transformation

$$F_{RD}(x; \theta) = \frac{F(x)}{1 + \theta S(x)}, \quad \theta > 0, \quad (1)$$

where θ is the *resilience parameter* controlling the degree of hazard modulation induced by the survival probability.

The corresponding probability density function is obtained by differentiation:

$$f_{RD}(x; \theta) = \frac{(1 + \theta)f(x)}{[1 + \theta S(x)]^2}, \quad (2)$$

where $f(x)$ denotes the baseline density.

A random variable X having density (2) will be denoted by

$$X \sim RD(\theta, F).$$

The survival function of the RD takes the form

$$S_{RD}(x; \theta) = \frac{(1 + \theta)S(x)}{1 + \theta S(x)}, \quad (3)$$

which shows how the resilience parameter modifies the baseline survival probability in a nonlinear manner.

The hazard function is given by

$$h_{RD}(x; \theta) = \frac{h(x)}{1 + \theta S(x)}, \quad (4)$$

where $h(x) = f(x)/S(x)$ is the baseline hazard function. This expression reveals that the parameter θ acts as a damping factor on the baseline hazard, with the magnitude of the effect depending on the survival probability.

Remark 1. Equation (1) can be written as

$$F_{RD}(x; \theta) = \frac{F(x)}{1 + \theta(1 - F(x))}.$$

Let $\alpha = 1 + \theta$. Then

$$F_{RD}(x; \theta) = \frac{F(x)}{\alpha + (1 - \alpha)F(x)},$$

which coincides with the classical Marshall–Olkin transformation ([3]). Hence the RD can be viewed as a reparameterized Marshall–Olkin family with a resilience-based interpretation.

3. SOME STATISTICAL PROPERTIES

This section provides some statistical properties of RD.

3.1. Quantile function

For $X \sim RD(\theta, F)$ with baseline quantile function $Q(p) = F^{-1}(p)$, the quantile function of X is:

$$Q_{RD}(p; \theta) = Q\left(\frac{p(1+\theta)}{1+p\theta}\right), \quad p \in (0, 1), \quad (5)$$

where $Q(\cdot)$ denotes the baseline quantile function. In particular, the first quartile, median, and third quartile are obtained by putting $p=0.25, 0.5$ and 0.75 , respectively, in (5).

3.2. Order statistics

Let $X_1, X_2, \dots, X_n \sim RD(\theta, F)$ be an i.i.d. sample from the resilience distribution with baseline distribution $F(x)$ and resilience parameter $\theta > 0$.

For the k -th order statistic $X_{(k)}$, its CDF is:

$$F_{X_{(k)}}(x) = \sum_{i=k}^n \binom{n}{i} [F_{RD}(x; \theta)]^i [S_{RD}(x; \theta)]^{n-i},$$

simplifying, we get,

$$F_{X_{(k)}}(x) = \sum_{i=k}^n \frac{\binom{n}{i} (\theta + 1)^{n-i} [F(x)]^i [S(x)]^{n-i}}{(1 + \theta S(x))^n}. \quad (6)$$

The cdf of first order statistic is:

$$F_{X_{(1)}}(x) = 1 - [S_{RD}(x; \theta)]^n = 1 - \left(\frac{(1 + \theta)S(x)}{1 + \theta S(x)}\right)^n. \quad (7)$$

The cdf of n^{th} order statistic is:

$$F_{X_{(n)}}(x) = [F_{RD}(x; \theta)]^n = \left(\frac{F(x)}{1 + \theta S(x)}\right)^n. \quad (8)$$

The pdf of k^{th} order statistic is:

$$f_{X_{(k)}}(x) = \binom{n}{k} [F_{RD}(x; \theta)]^{k-1} [S_{RD}(x; \theta)]^{n-k} f_{RD}(x; \theta),$$

simplifying, we get,

$$f_{X_{(k)}}(x) = \frac{\binom{n}{k} (\theta + 1)^{n-k+1} [F(x)]^{k-1} [S(x)]^{n-k} f(x)}{(1 + \theta S(x))^{n+1}}. \quad (9)$$

3.3. Moments

For $X \sim RD(\theta, F)$ with baseline distribution $F(x)$, the r -th raw moment is:

$$\mathbb{E}[X^r] = (1 + \theta) \int_{-\infty}^{\infty} \frac{x^r f(x)}{[1 + \theta S(x)]^2} dx,$$

using binomial expansion of $[1 + \theta S(x)]^{-2} = \sum_{k=0}^{\infty} (k+1)(-\theta)^k S(x)^k$, we get,

$$\mathbb{E}[X^r] = \sum_{k=0}^{\infty} (k+1)(-\theta)^k (1 + \theta) \mathbb{E}_F[X^r S(X)^k]. \quad (10)$$

Where, $\mathbb{E}_F[X^r S(X)^k] = \int_{-\infty}^{\infty} X^r S(X)^k f(x) dx$.

The variance exhibits non-monotonic behavior in θ :

$$\text{Var}(X) = (1 + \theta) \mathbb{E}_F \left[\frac{X^2}{(1 + \theta S(X))^2} \right] - \left((1 + \theta) \mathbb{E}_F \left[\frac{X}{(1 + \theta S(X))^2} \right] \right)^2. \quad (11)$$

3.4. Moment Generating Function

The moment generating function $M_X(t) = \mathbb{E}[e^{tX}]$ satisfies:

$$M_X(t) = (1 + \theta) \int_{-\infty}^{\infty} \frac{e^{tx} f(x)}{[1 + \theta S(x)]^2} dx,$$

when the baseline MGF $M_F(t)$ exists, the RD MGF admits the series representation:

$$M_X(t) = \sum_{k=0}^{\infty} (k+1)(-\theta)^k (1+\theta) \mathbb{E}_F[e^{tX} S(X)^k]. \quad (12)$$

4. SOME IMPORTANT THEOREMS

Theorem 1. Strict Survival-Weighted Hazard Damping:

Let $X \sim \text{RD}(\theta, F)$ with baseline hazard rate $h(x)$ and survival function $S(x)$. For all resilience parameters *strictly positive* ($\theta > 0$):

$$h_{\text{RD}}(x; \theta) < h(x) \quad \forall x \text{ where } S(x) > 0, \quad (13)$$

with the damping factor $(1 + \theta S(x))^{-1}$.

Proof. We have,

$$h_{\text{RD}}(x; \theta) = \frac{h(x)}{1 + \theta S(x)},$$

where $h(x)$ is the baseline hazard rate.

Since $S(x) > 0$ and $\theta > 0$ by definition:

$$1 + \theta S(x) > 1.$$

Thus for all $\theta > 0$ and $S(x) > 0$:

$$h_{\text{RD}}(x; \theta) = \frac{h(x)}{1 + \theta S(x)} < h(x).$$

■

Theorem 2. Resilience-Induced Stochastic Ordering:

For two RD variants $X_1 \sim \text{RD}(\theta_1, F)$ and $X_2 \sim \text{RD}(\theta_2, F)$ with $\theta_1 < \theta_2$:

$$X_1 \leq_{hr} X_2, \quad (14)$$

where $\leq_{hr}$ denotes hazard rate ordering. Higher θ increases resilience by uniformly reducing the hazard rate.

Proof. For $\theta_1 < \theta_2$, compare their hazard rates:

$$\begin{aligned} \frac{h_{\text{RD}}(x; \theta_1)}{h_{\text{RD}}(x; \theta_2)} &= \frac{h(x)/(1 + \theta_1 S(x))}{h(x)/(1 + \theta_2 S(x))} \\ &= \frac{1 + \theta_2 S(x)}{1 + \theta_1 S(x)} \\ &\geq 1 \quad \forall x \text{ since } \theta_2 > \theta_1 \text{ and } S(x) > 0. \end{aligned}$$

Thus $h_{\text{RD}}(x; \theta_1) \geq h_{\text{RD}}(x; \theta_2)$ for all x , which by definition means $X_1 \leq_{hr} X_2$ in hazard rate ordering. ■

Theorem 3. Implications of Hazard Rate Ordering: For $X_1 \sim \text{RD}(\theta_1, F)$ and $X_2 \sim \text{RD}(\theta_2, F)$ with $\theta_1 < \theta_2$, the following hold:

1. Stochastic Dominance: $X_2 \leq_{st} X_1$.
2. Mean Residual Life Ordering: $m_2(x) \geq m_1(x) \forall x$.
3. Quantile Dominance: $F_{RD}^{-1}(p; \theta_2) \geq F_{RD}^{-1}(p; \theta_1) \forall p \in (0, 1)$.

Where $\leq_{st}$ denotes usual stochastic ordering and $m_i(x)$ is the mean residual life function.

Proof.

1. Hazard rate ordering implies stochastic ordering: $X_1 \leq_{hr} X_2 \Rightarrow X_2 \leq_{st} X_1$.
2. The mean residual life $m(x) = \mathbb{E}[X - x | X > x]$ satisfies:

$$m_i(x) = \frac{1}{S_{RD}(x; \theta_i)} \int_x^\infty S_{RD}(t; \theta_i) dt,$$

since $S_{RD}(x; \theta_2) \geq S_{RD}(x; \theta_1)$, the integral dominates. Thus, $m_2(x) \geq m_1(x) \forall x$.

3. For quantiles, note that:

$$F_{RD}^{-1}(p; \theta) = F^{-1} \left(\frac{p(1 + \theta)}{1 + p\theta} \right)$$

■

Theorem 4. Tail Resilience Preservation: The RD asymptotically preserves the baseline's tail behavior while modulating mid-range hazards. For $x \rightarrow \infty$:

$$h_{RD}(x; \theta) \sim h(x),$$

where $\sim$ denotes asymptotic equivalence. Resilience effects are concentrated in regions where $S(x)$ is non-negligible.

Proof. As $x \rightarrow \infty$, $S(x) \rightarrow 0$ by definition of survival functions. Therefore:
For the hazard rate:

$$\begin{aligned} h_{RD}(x; \theta) &= \frac{h(x)}{1 + \theta S(x)}, \\ &\sim h(x) \quad \text{since } S(x) \rightarrow 0. \end{aligned}$$

■

Theorem 5. Adaptive Hazard Elasticity and Its Bounds

The elasticity of $h_{RD}(x; \theta)$ with respect to θ is state-dependent and is given by

$$\eta_{RD}(x; \theta) = -\frac{\theta S(x)}{1 + \theta S(x)}. \quad (15)$$

Furthermore, the following properties hold:

1. **Elasticity Bounds:**

$$-\theta \leq \eta_{RD}(x; \theta) \leq 0. \quad (16)$$

2. **Maximal Sensitivity:** The elasticity attains its minimum value when $S(x) = 1$:

$$\min_x \eta_{RD}(x; \theta) = -\frac{\theta}{1 + \theta}. \quad (17)$$

3. **Risk-Regime Insensitivity:**

$$\lim_{S(x) \rightarrow 0} \eta_{RD}(x; \theta) = 0. \quad (18)$$

4. State-dependent Sensitivity: The magnitude of the elasticity is greater in high-survival regions ($S(x) \approx 1$) and approaches zero in high-risk regions ($S(x) \approx 0$).

Proof. The elasticity of the RD hazard rate with respect to θ is defined as

$$\eta_{RD}(x; \theta) = \frac{\partial h_{RD}(x; \theta)}{\partial \theta} \cdot \frac{\theta}{h_{RD}(x; \theta)}.$$

Since

$$h_{RD}(x; \theta) = \frac{h(x)}{1 + \theta S(x)},$$

we have

$$\begin{aligned} \frac{\partial}{\partial \theta} h_{RD}(x; \theta) &= \frac{\partial}{\partial \theta} \left(\frac{h(x)}{1 + \theta S(x)} \right) \\ &= -\frac{h(x)S(x)}{(1 + \theta S(x))^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \eta_{RD}(x; \theta) &= \left(-\frac{h(x)S(x)}{(1 + \theta S(x))^2} \right) \frac{\theta}{\frac{h(x)}{1 + \theta S(x)}} \\ &= -\frac{\theta S(x)}{1 + \theta S(x)}. \end{aligned}$$

This proves Equation (15).

For the bounds, since $0 \leq S(x) \leq 1$ and $\theta > 0$,

$$0 \leq \frac{\theta S(x)}{1 + \theta S(x)} \leq \theta.$$

Consequently,

$$-\theta \leq \eta_{RD}(x; \theta) \leq 0.$$

For maximal sensitivity, consider

$$g(S) = -\frac{\theta S}{1 + \theta S}.$$

Its derivative with respect to S is

$$g'(S) = -\frac{\theta}{(1 + \theta S)^2} < 0.$$

Hence, $g(S)$ is decreasing in S , and therefore its minimum occurs at $S = 1$. Thus,

$$\min_x \eta_{RD}(x; \theta) = -\frac{\theta}{1 + \theta}.$$

Finally, as $S(x) \rightarrow 0$,

$$\eta_{RD}(x; \theta) = -\frac{\theta S(x)}{1 + \theta S(x)} \rightarrow 0.$$

Thus, the elasticity approaches zero in the high-risk region where the survival probability becomes negligible. This completes the proof. ■

5. SOME SPECIAL MODELS

5.1. Resilience Exponential distribution (RED)

Consider $X \sim \exp(\lambda)$ with baseline functions:

$$F(x) = 1 - e^{-\lambda x}, f(x) = \lambda e^{-\lambda x},$$

and

$$S(x) = e^{-\lambda x}, h(x) = \lambda.$$

Then the pdf, cdf, survival and hazard rate functions of RED are:

$$F_{RED}(x; \theta) = \frac{1 - e^{-\lambda x}}{1 + \theta e^{-\lambda x}},$$

$$f_{RED}(x; \theta) = \frac{(1 + \theta)\lambda e^{-\lambda x}}{[1 + \theta e^{-\lambda x}]^2},$$

$$S_{RED}(x; \theta) = 1 - \frac{1 - e^{-\lambda x}}{1 + \theta e^{-\lambda x}},$$

and

$$h_{RED}(x; \theta) = \frac{\lambda}{1 + \theta e^{-\lambda x}}.$$

5.2. Resilience Weibull distribution (RWD)

Consider $X \sim Weibull(k, \lambda)$ with baseline functions:

$$F(x) = 1 - e^{-(x/\lambda)^k}, f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k},$$

and

$$S(x) = e^{-(x/\lambda)^k}, h(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1}.$$

For $X \sim RD(\theta, Weibull(k, \lambda))$, the pdf, cdf, survival and hazard rate functions are:

- Cumulative Distribution Function (CDF):

$$F_{RWD}(x; \theta) = \frac{1 - e^{-(x/\lambda)^k}}{1 + \theta e^{-(x/\lambda)^k}}.$$

- Probability Density Function (PDF):

$$f_{RWD}(x; \theta) = \frac{(1 + \theta) \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}}{\left[1 + \theta e^{-(x/\lambda)^k}\right]^2}.$$

- Survival Function:

$$S_{RWD}(x; \theta) = \frac{(1 + \theta)e^{-(x/\lambda)^k}}{1 + \theta e^{-(x/\lambda)^k}}.$$

- Hazard Rate Function:

$$h_{RWD}(x; \theta) = \frac{\frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1}}{1 + \theta e^{-(x/\lambda)^k}}.$$

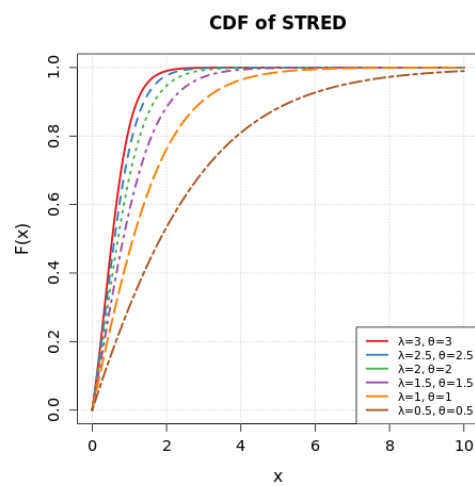


Figure 1: CDF of RED with different values of λ and θ .

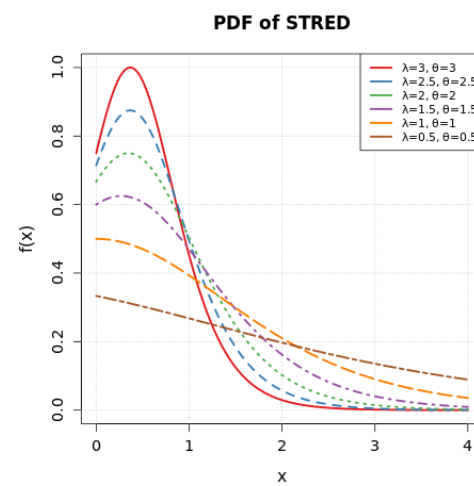


Figure 2: PDF of RED with different values of λ and θ .

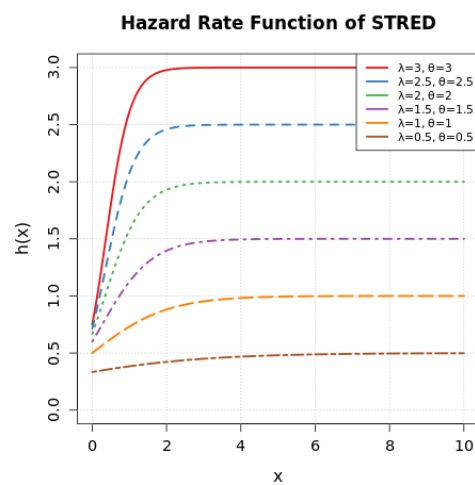


Figure 3: Hazard rate of RED with different values of λ and θ .

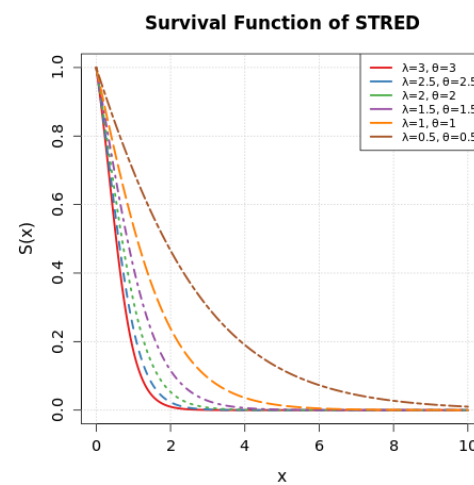


Figure 4: Survival rate of RED with different values of λ and θ .

Remark. When the Weibull shape parameter $k = 1$, the Weibull distribution reduces to the exponential distribution with scale parameter λ . Consequently, the Resilience Weibull Distribution reduces to the Resilience Exponential Distribution for $k = 1$.

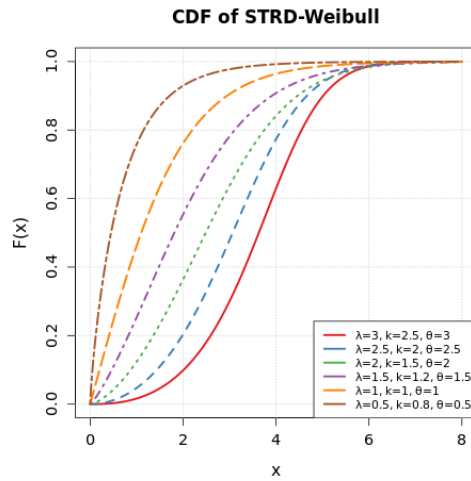


Figure 5: CDF of RWD with different values of λ , k and θ .

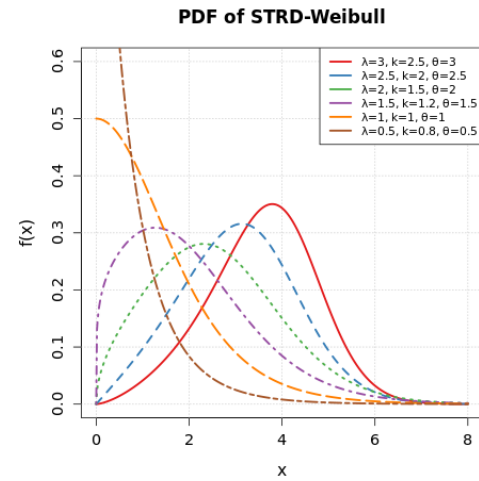


Figure 6: PDF of RWD with different values of λ , k and θ .

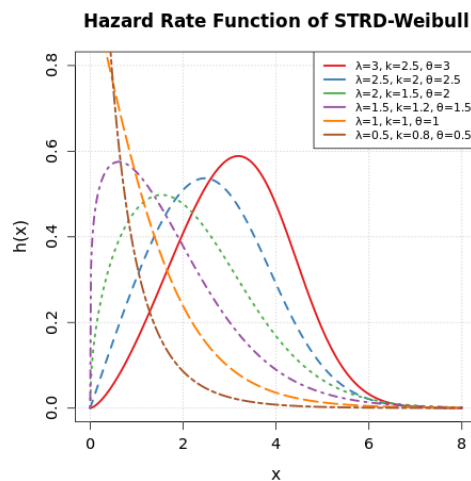


Figure 7: Hazard rate of RWD with different values of λ , k and θ .

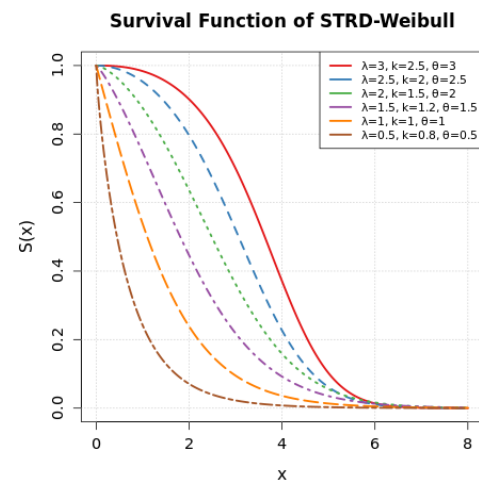


Figure 8: Survival rate of RWD with different values of λ , k and θ .

5.3. Resilience Lindley Distribution(RLD)

For $X \sim \text{Lindley}(\beta)$ with shape parameter $\beta > 0$, the pdf, cdf, survival and hazard rate functions are:

$$f(x; \beta) = \frac{\beta^2}{\beta + 1} (1 + x) e^{-\beta x}, \quad x > 0, F(x; \beta) = 1 - \left(1 + \frac{\beta x}{\beta + 1}\right) e^{-\beta x},$$

and

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

Then, for $X \sim \text{RD}(\theta, \text{Lindley}(\beta))$ where $\beta > 0$ is the Lindley shape parameter, then the pdf, cdf, survival and hazard rate functions are:

- Cumulative Distribution Function (CDF):

$$F_{RLD}(x; \theta) = \frac{1 - \left(1 + \frac{\beta x}{\beta+1}\right) e^{-\beta x}}{1 + \theta \left(1 + \frac{\beta x}{\beta+1}\right) e^{-\beta x}}.$$

- Probability Density Function (PDF):

$$f_{RLD}(x; \theta) = \frac{(1 + \theta) \frac{\beta^2}{\beta+1} (1 + x) e^{-\beta x}}{\left[1 + \theta \left(1 + \frac{\beta x}{\beta+1}\right) e^{-\beta x}\right]^2}.$$

- Survival Function:

$$S_{RLD}(x; \theta) = \frac{(1 + \theta) \left(1 + \frac{\beta x}{\beta+1}\right) e^{-\beta x}}{1 + \theta \left(1 + \frac{\beta x}{\beta+1}\right) e^{-\beta x}}.$$

- Hazard Rate Function:

$$h_{RLD}(x; \theta) = \frac{\frac{\beta^2}{\beta+1} (1 + x)}{1 + \theta \left(1 + \frac{\beta x}{\beta+1}\right) e^{-\beta x}}.$$

6. MAXIMUM LIKELIHOOD METHOD

Let $X_1, X_2, \dots, X_n$ be a random sample from the RD.

The likelihood function for the parameter θ based on the observed sample is;

$$L(\theta) = \prod_{i=1}^n \frac{(1 + \theta) f(x_i)}{[1 + \theta S(x_i)]^2}.$$

Taking the natural logarithm of the likelihood function, the log-likelihood function becomes,

$$\ell(\theta) = \sum_{i=1}^n [\log(1 + \theta) + \log f(x_i) - 2 \log(1 + \theta S(x_i))].$$

This simplifies to

$$\ell(\theta) = n \log(1 + \theta) + \sum_{i=1}^n \log f(x_i) - 2 \sum_{i=1}^n \log(1 + \theta S(x_i)).$$

Differentiating $\ell(\theta)$ with respect to θ yields the score function:

$$\frac{d\ell}{d\theta} = \frac{n}{1 + \theta} - 2 \sum_{i=1}^n \frac{S(x_i)}{1 + \theta S(x_i)}.$$

The maximum likelihood estimator $\hat{\theta}$ is the value of θ that satisfies the nonlinear equation

$$\frac{n}{1 + \theta} = 2 \sum_{i=1}^n \frac{S(x_i)}{1 + \theta S(x_i)}.$$

Since this equation does not admit a closed-form solution, numerical methods such as the Newton-Raphson algorithm or other iterative procedures must be employed to obtain $\hat{\theta}$. The required calculations depend on the choice of the baseline distribution and its corresponding survival function $S(x)$.

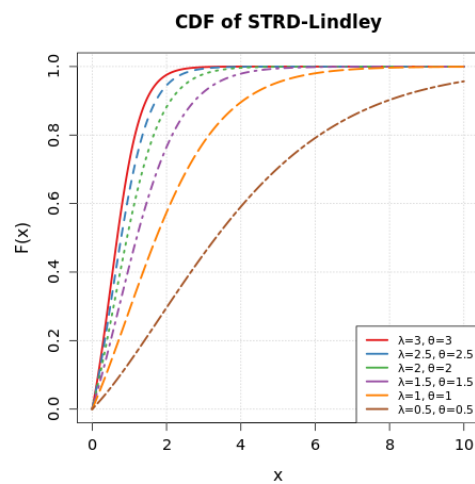


Figure 9: CDF of RLD with different values of β and θ .

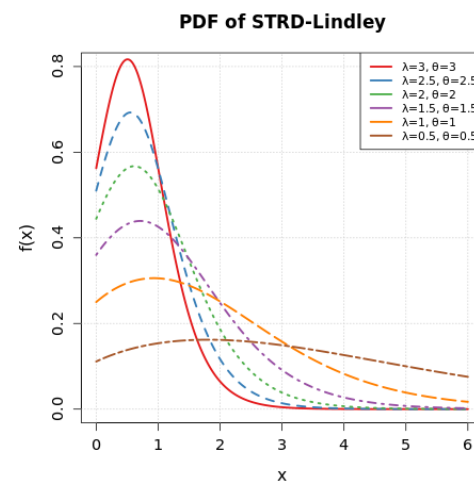


Figure 10: PDF of RLD with different values of β and θ .

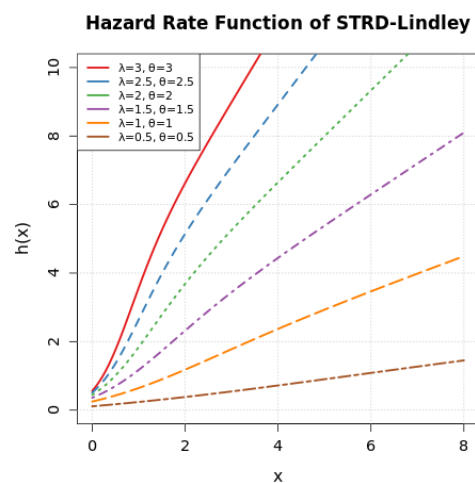


Figure 11: Hazard rate of RLD with different values of β and θ .

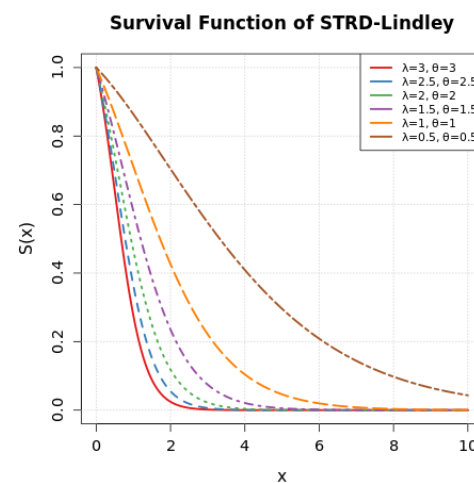


Figure 12: Survival rate of RLD with different values of β and θ .

7. SIMULATION STUDY FOR THE RLD

To investigate the finite-sample performance of the proposed estimation procedure, a Monte Carlo simulation study was conducted for the Resilience Lindley Distribution (RLD) introduced in Section 5.3. The parameters $\beta > 0$ and $\theta > 0$ denote the Lindley parameter and the resilience parameter, respectively.

We considered the following true parameter combinations:

$$(\beta, \theta) \in \{(2, 0.5), (1, 2.5), (0.5, 0.75), (3, 3)\},$$

and the sample sizes

$$n \in \{30, 50, 100, 150, 250, 400, 600, 750, 1000\}.$$

For each pair (β, θ) and each sample size n , the following steps were repeated for $R = 1000$ Monte Carlo replications:

1. Generate a random sample $X_1, \dots, X_n$ from the Resilience Lindley Distribution with true parameter values $(\beta_{\text{true}}, \theta_{\text{true}})$.
2. For each generated sample, compute the maximum likelihood estimators $\hat{\beta}$ and $\hat{\theta}$.
3. Record $(\hat{\beta}_r, \hat{\theta}_r)$ for replication $r = 1, \dots, R$.

After all replications, the following summary measures were computed.

Absolute Error (AE).

$$AE_{\beta} = \frac{1}{R} \sum_{r=1}^R |\hat{\beta}_r - \beta_{\text{true}}|, \quad AE_{\theta} = \frac{1}{R} \sum_{r=1}^R |\hat{\theta}_r - \theta_{\text{true}}|.$$

Absolute Bias (AB).

$$AB_{\beta} = \left| \frac{1}{R} \sum_{r=1}^R \hat{\beta}_r - \beta_{\text{true}} \right|, \quad AB_{\theta} = \left| \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r - \theta_{\text{true}} \right|.$$

Average Variance (AV). Define

$$\bar{\hat{\beta}} = \frac{1}{R} \sum_{r=1}^R \hat{\beta}_r, \quad \bar{\hat{\theta}} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r.$$

Then

$$AV_{\beta} = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_r - \bar{\hat{\beta}})^2, \quad AV_{\theta} = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \bar{\hat{\theta}})^2.$$

Mean Squared Error (MSE).

$$MSE_{\beta} = \frac{1}{R} \sum_{r=1}^R (\hat{\beta}_r - \beta_{\text{true}})^2, \quad MSE_{\theta} = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta_{\text{true}})^2.$$

Across all four parameter settings, a clear and consistent pattern emerges. As the sample size n increases from 30 to 1000:

- AE_{β} and AE_{θ} decrease steadily;
- AB_{β} and AB_{θ} move closer to zero;
- AV_{β} and AV_{θ} shrink, indicating reduced variability;

Table 1: Simulation results for RLD model

Sample Size	AE		AB		AV		MSE	
	β	θ	β	θ	β	θ	β	θ
$\beta = 2, \theta = 0.5$								
30	0.4177	1.0272	0.2207	0.6939	0.2584	3.2090	0.3068	3.6873
50	0.3247	0.6918	0.1364	0.3765	0.1636	1.1102	0.1820	1.2508
100	0.2375	0.4514	0.0514	0.1558	0.0887	0.3509	0.0913	0.3748
150	0.2049	0.3784	0.0421	0.1242	0.0650	0.2659	0.0667	0.2811
250	0.1547	0.2728	0.0048	0.0396	0.0364	0.1177	0.0364	0.1191
400	0.1204	0.2234	0.0118	0.0397	0.0228	0.0782	0.0229	0.0797
600	0.1032	0.1835	0.0088	0.0273	0.0170	0.0553	0.0170	0.0560
750	0.0890	0.1615	0.0038	0.0164	0.0123	0.0403	0.0123	0.0405
1000	0.0748	0.1354	0.0011	0.0101	0.0089	0.0292	0.0088	0.0293
$\beta = 1, \theta = 2.5$								
30	0.1885	2.6886	0.0596	1.5726	0.0542	20.4704	0.0577	22.9231
50	0.1460	1.8233	0.0367	0.8712	0.0332	7.2886	0.0345	8.0403
100	0.1026	1.1921	0.0196	0.4480	0.0166	2.5271	0.0170	2.7253
150	0.0807	0.9285	0.0110	0.2702	0.0102	1.4780	0.0103	1.5496
250	0.0616	0.6692	0.0064	0.1539	0.0059	0.7287	0.0060	0.7516
400	0.0498	0.5574	0.0014	0.0697	0.0039	0.5063	0.0039	0.5106
600	0.0392	0.4139	0.0014	0.0504	0.0024	0.2725	0.0024	0.2748
750	0.0367	0.3967	0.0035	0.0692	0.0021	0.2509	0.0021	0.2554
1000	0.0304	0.3203	0.0010	0.0345	0.0014	0.1665	0.0014	0.1676
$\beta = 0.5, \theta = 0.75$								
30	0.1038	1.4087	0.0465	0.9523	0.0170	6.3881	0.0192	7.2886
50	0.0836	1.0000	0.0275	0.5702	0.0109	2.0141	0.0116	2.3372
100	0.0592	0.5901	0.0139	0.2207	0.0052	0.6021	0.0053	0.6502
150	0.0468	0.4557	0.0039	0.1124	0.0033	0.3595	0.0033	0.3718
250	0.0365	0.3529	0.0048	0.0830	0.0020	0.2012	0.0021	0.2079
400	0.0274	0.2626	0.0030	0.0527	0.0012	0.1090	0.0012	0.1117
600	0.0243	0.2336	0.0016	0.0295	0.0009	0.0852	0.0009	0.0860
750	0.0219	0.2101	0.0001	0.0251	0.0008	0.0688	0.0008	0.0693
1000	0.0185	0.1791	0.0013	0.0224	0.0005	0.0497	0.0005	0.0501
$\beta = 3, \theta = 3$								
30	0.5617	2.6886	0.1520	1.4058	0.5068	17.7024	0.5294	19.6608
50	0.4242	1.9312	0.1238	1.0081	0.2863	7.9257	0.3013	8.9342
100	0.3029	1.2760	0.0560	0.4432	0.1404	2.8954	0.1434	3.0889
150	0.2353	0.9476	0.0287	0.2591	0.0873	1.5324	0.0880	1.5980
250	0.1925	0.7455	0.0169	0.1392	0.0548	0.8792	0.0550	0.8977
400	0.1404	0.5568	0.0107	0.0867	0.0307	0.4947	0.0308	0.5017
600	0.1206	0.4890	0.0098	0.0695	0.0224	0.3775	0.0224	0.3819
750	0.1025	0.4065	0.0029	0.0521	0.0164	0.2600	0.0164	0.2624
1000	0.0884	0.3605	0.0017	0.0351	0.0121	0.2041	0.0121	0.2051

AE = Absolute Error, AB = Absolute Bias, AV = Average Variance, MSE = Mean Squared Error.

- MSE_{β} and MSE_{θ} also decrease, often quite sharply.

This behavior is characteristic of consistent estimators whose precision improves as the sample size grows. For large n , all entries are small, demonstrating that both $\hat{\beta}$ and $\hat{\theta}$ estimate the true parameters accurately.

In almost all scenarios, θ is somewhat harder to estimate than β : AV_θ and MSE_θ are typically larger than their counterparts for β , especially for small to moderate n . However, this difference becomes much less pronounced as n increases.

8. APPLICATION

In this section, we evaluate the performance of the proposed Resilience Lindley Distribution (RLD) using two real-world datasets. The model comparison includes six established distributions: Lindley, Exponential, Gamma-Lindley ([4]), Generalized Lindley ([5]), DUS Lindley([6]), and DUS Exponential ([7]) distributions.

Model performance is evaluated using standard statistical criteria: the log-likelihood (Log-Lik), Akaike Information Criterion (AIC), corrected AIC (AICc), Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), and the Kolmogorov-Smirnov (KS) test with corresponding p-values.

Data set 1: The following data set, studied by [8], represents the successive daily precipitation (in inches) recorded over thirty days in March.

0.77, 1.74, 0.81, 1.20, 1.95, 1.20, 0.47, 1.43, 3.37, 2.20, 3.00, 3.09, 1.51, 2.10, 0.52, 1.62, 1.31, 0.32, 0.59, 0.81, 2.81, 1.87, 1.18, 1.35, 4.75, 2.48, 0.96, 1.89, 0.90, 2.05.

Data set 2: This data set, provided by [9], consists of blood cancer (leukaemia) measurements from 40 patients treated at Ministry of Health hospitals in Saudi Arabia. The observations are as follows:

0.315, 0.496, 0.616, 1.145, 1.208, 1.263, 1.414, 2.025, 2.036, 2.162, 2.211, 2.37, 2.532, 2.693, 2.805, 2.91, 2.912, 3.192, 3.263, 3.348, 3.348, 3.427, 3.499, 3.534, 3.767, 3.751, 3.858, 3.986, 4.049, 4.244, 4.323, 4.381, 4.392, 4.397, 4.647, 4.753, 4.929, 4.973, 5.074, 5.381.

Model Comparison and Evaluation

The fitting results for the two datasets are summarized in Table(2). The RLD achieves the highest log-likelihood values and the lowest AIC, AICc, BIC, and HQIC scores across both datasets, indicating a consistently better fit compared to alternative models. Additionally, the Kolmogorov-Smirnov (KS) test p-values for the RLD are higher than those of competing distributions, reflecting stronger agreement with the empirical data. The parameter estimates for the RLD are $\beta = 1.791$ and $\theta = 5.995$ for Dataset 1, and $\beta = 1.142$ and $\theta = 10$ for Dataset 2, demonstrating the model's adaptability to varying data structures. These findings confirm the RLD's flexibility and reliability in modeling real-world survival and reliability data, reinforcing its potential as a robust tool for applied statistical analysis in environmental and medical research contexts.

The visual diagnostics in Figures 13–18 confirm these quantitative results. The density plots show that the RLD accurately captures the shape characteristics of both datasets. The CDF plots demonstrate close agreement between theoretical and empirical distributions, while the P-P plots exhibit nearly perfect diagonal alignment.

13. CONCLUSION

This paper develops a resilience-based interpretation of the Marshall-Olkin transformation through a survival-weighted hazard formulation. The resulting framework, referred to as the Resilience Distribution (RD), introduces a parameter that modulates the hazard function via the survival probability, providing a meaningful representation of resilience in survival and reliability systems. Rather than proposing a new distributional form, the study offers an alternative perspective in which the additional parameter governs the adaptive behavior of the hazard rate.

A range of statistical properties of the RD has been derived, including the quantile function, moments, moment generating function, and order statistics. Theoretical results such as hazard damping, stochastic ordering, tail behavior, and elasticity properties establish a solid foundation

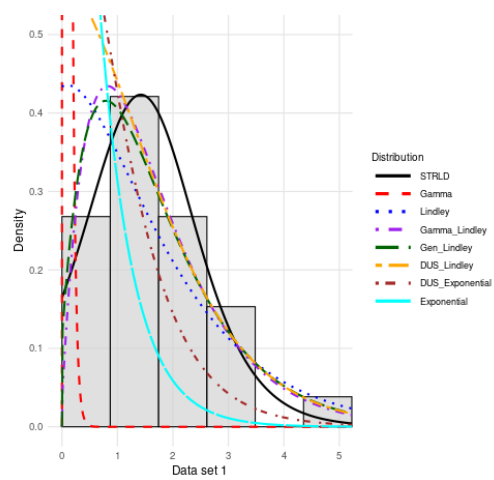


Figure 13: Fitted density curves of data set 1.

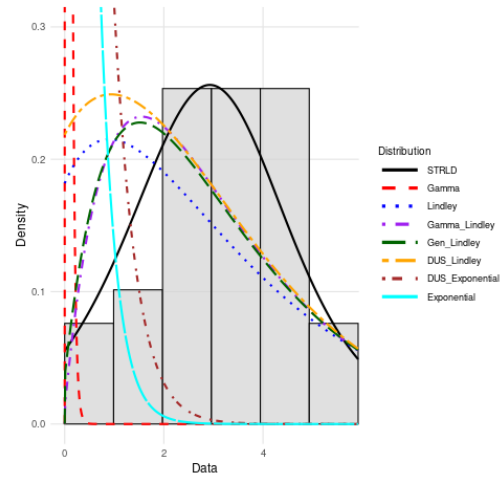


Figure 14: Fitted density curves of data set 2.

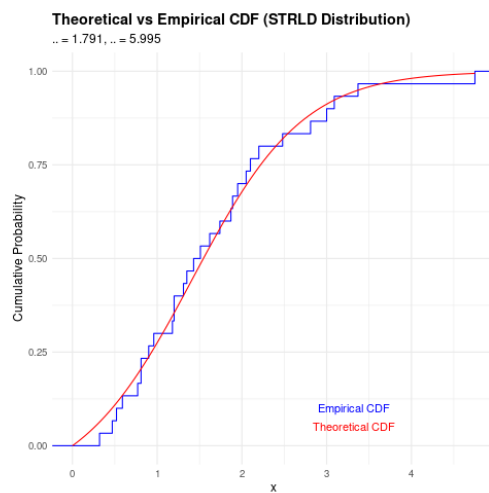


Figure 15: Theoretical vs empirical CDF of data set 1.

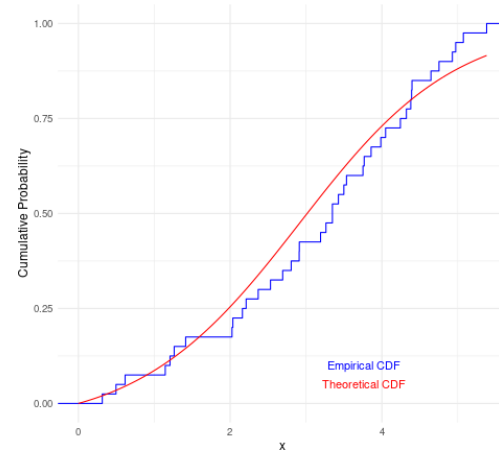


Figure 16: Theoretical vs empirical CDF of data set 2.

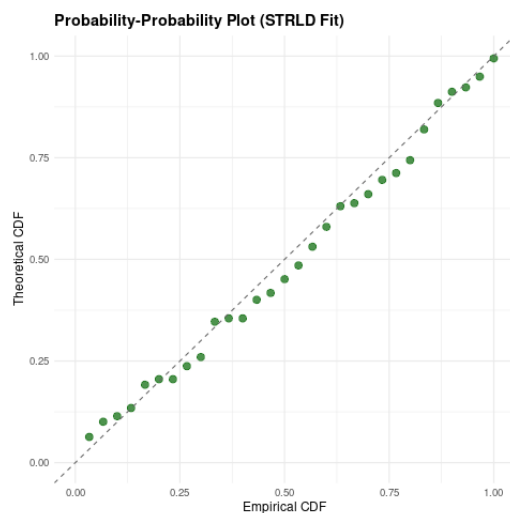


Figure 17: P-P Plot for RLD of data set 1.

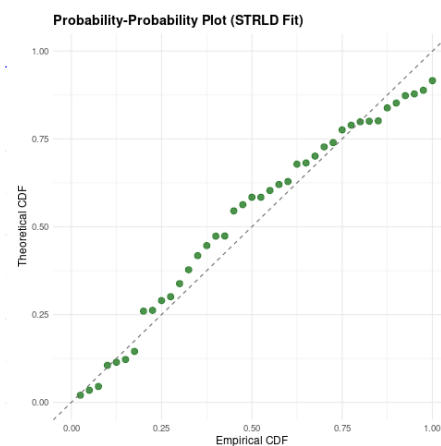


Figure 18: P-P Plot for RLD of data set 2.

Table 2: Comparative Fit Statistics for Lifetime Data Models

Distribution	Parameters	LogL	AIC	AICc	BIC	HQIC	KS Statistic	KSp-value
Dataset 1 (n=60)								
RLD	$\beta = 1.791, \theta = 5.995$	-39.4628	82.9255	83.3700	85.7279	83.8221	0.0673	0.9992
Gamma-Lindley	$\theta = 1.180, \beta = 20$	-39.5559	82.9518	83.3962	85.7442	83.8483	0.1149	0.8229
Generalized Lindley	$\theta = 1.133, \alpha = 10$	-40.0676	84.1353	84.5797	86.9376	85.0318	0.1221	0.7628
Lindley	$\theta = 0.910$	-43.1438	88.2875	88.4303	89.6886	88.7357	0.1882	0.2383
DUS Lindley	$\theta = 1.119$	-41.4765	84.9530	85.0959	86.3542	85.4013	0.1578	0.4436
DUS Exponential	$\lambda = 0.780$	-43.1016	88.2031	88.3459	89.6043	88.6513	0.1988	0.1868
Exponential	$\lambda = 0.597$	-45.4744	92.9488	93.0916	94.3450	93.3971	0.2352	0.0724
Dataset 2 (n=40)								
RLD	$\beta = 1.142, \theta = 10$	-69.5721	143.1443	143.4686	146.5220	144.3656	0.1198	0.6137
Gamma-Lindley	$\theta = 0.631, \beta = 20$	-76.5481	157.0962	157.4205	160.4740	158.3175	0.1973	0.0889
Generalized Lindley	$\theta = 0.618, \alpha = 10$	-76.9203	157.8405	158.1649	161.2183	159.0618	0.2019	0.0767
Lindley	$\theta = 0.527$	-80.2506	162.5012	162.6065	164.1900	163.1118	0.2406	0.0195
DUS Lindley	$\theta = 0.646$	-77.2856	156.5713	156.6766	158.2602	157.1819	0.2168	0.0515
DUS Exponential	$\lambda = 0.417$	-81.7586	165.5171	165.6224	167.2060	166.1277	0.2726	0.0052
Exponential	$\lambda = 0.318$	-85.7782	173.5563	173.6616	175.2452	174.1669	0.3002	0.0015

for the formulation. Parameter estimation based on the maximum likelihood method shows stable and consistent performance, supported by simulation results.

The practical relevance of the approach is demonstrated through a specialization based on the Lindley distribution, applied to real datasets. The results illustrate the flexibility and interpretability of the resilience-based formulation in capturing diverse hazard rate patterns observed in practice.

From a reliability engineering perspective, the proposed framework is particularly useful for modeling systems whose failure behavior depends on accumulated survival experience. By linking hazard dynamics to the survival state, the formulation provides a realistic representation of adaptive failure mechanisms in engineering, environmental, and biomedical contexts.

Overall, the proposed approach provides a useful bridge between classical lifetime models and resilience-based hazard modeling while preserving analytical tractability. Future work may consider extensions incorporating covariate-dependent resilience parameters and regression structures to enhance applicability in complex reliability and risk analysis settings.

ACKNOWLEDGEMENT

The authors acknowledge the use of ChatGPT (OpenAI) for language editing, improving clarity and readability, and assisting with the presentation of the manuscript. The authors reviewed and verified all content and take full responsibility for the accuracy and integrity of the work.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to the conception, methodology development, data analysis, and manuscript preparation. Both authors read and approved the final manuscript.

DATA AVAILABILITY

The datasets used in this study are publicly available and were obtained from previously published sources cited in the article.

FUNDING

The authors received no financial support for the research, authorship, and/or publication of this article.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- [1] Mudholkar, G. S., and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*. 42(2):299–302. <https://doi.org/10.1109/24.229504>
- [2] Mahdavi, A., and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics – Theory and Methods*. 46(10):6543–6557. <https://doi.org/10.1080/03610926.2015.1130839>
- [3] Marshall, A. W., and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*. 84(3):641–652. <https://doi.org/10.1093/biomet/84.3.641>
- [4] Nedjar, S., and Zeghdoudi, H. (2016). On gamma Lindley distribution: Properties and simulations. *Journal of Computational and Applied Mathematics*. 298:167–174. <https://doi.org/10.1016/j.cam.2015.11.047>
- [5] Nadarajah, S., et al. (2011). A generalized Lindley distribution. *Sankhya: The Indian Journal of Statistics, Series B*. 73(2):331–359. <https://doi.org/10.1007/s13571-011-0025-9>
- [6] Qayoom, D., et al. (2024). A new class of Lindley distribution: System reliability, simulation and applications. *Heliyon*. 10(19):e38335. <https://doi.org/10.1016/j.heliyon.2024.e38335>
- [7] Tripathi, A., et al. (2019). Inferences for the DUS-exponential distribution based on upper record values. *Annals of Data Science*. 8:387–403. <https://doi.org/10.1007/s40745-019-00231-6>
- [8] Hinkley, D. (1977). On quick choice of power transformation. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 26(1):67–69. <https://doi.org/10.2307/2346869>
- [9] Abouammoh, A. M., et al. (2000). On new renewal better than used classes of life distribution. *Statistics & Probability Letters*. 48(2):189–194. [https://doi.org/10.1016/S0167-7152\(99\)00204-7](https://doi.org/10.1016/S0167-7152(99)00204-7)
- [10] Henry, D., and Ramirez-Marquez, J. E. (2012). Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering & System Safety*. 99:114–122. <https://doi.org/10.1016/j.ress.2011.09.002>

RELIABILITY MODELING USING AN EQUILIBRIUM-BASED XGAMMA DISTRIBUTION: PROPERTIES, ESTIMATION, AND APPLICATIONS

MOLAY KUMAR RUIDAS¹, SUKANTA PRAMANIK² AND SUDHANSU S. MAITI³

•

¹Department of Statistics, Presidency University, Kolkata, India 700 073,
molayruidas@gmail.com

² Department of Statistics, Siliguri College, Siliguri, India 734 001,
skantapramanik@gmail.com

³Department of Statistics, Visva Bharati University, Santiniketan, India 731 235,
dssm1@rediffmail.com

Abstract

The equilibrium distribution approach is applied to the existing Xgamma model in this work to develop a new lifetime model known as the Induced Xgamma (i-Xg) distribution. Because of its adaptable shape properties, the suggested distribution can be used to model heterogeneous and positively skewed lifetime datasets. Probability density and distribution functions, moments, the moment generating function, reliability measures (mean residual life, hazard rate, and reversed hazard), stochastic dominance, entropy measures, inequality curves, and order statistics, including extremes, are among the important analytical properties that we derive. The model's identifiability is rigorously established. The model parameter is estimated using seven traditional estimation methods: Maximum Likelihood (ML), Least Squares (LS), Weighted LSE (WLS), Cramer-von-Mises (CvM), Maximum Product of Spacings (MPS), Anderson-Darling (AD), and Right-tail AD (RAD). Their performance is assessed in terms of bias and MSE for varying sample sizes using a Monte-Carlo simulation study. The suggested i-Xg model provides a better fit than competing distributions, as shown by applications to a variety of real-life datasets, confirming its value in statistical modelling of survival-type and reliability data.

Keywords: Lorenz curve, Entropy, Generating function, Hazard function, MLE, MRLF, Order Statistics

1. INTRODUCTION

The field of data analytics is significant in the age of data technology. The data set's level of uncertainty is determined using the probability distribution. A class of probability distributions that fits all the data sets is improbable. When a class of probability distributions adequately matches a data collection, one can look for another class that does so. Despite the abundance of probability distributions described in the literature, there is always room to investigate new ones that might fit a data set more effectively than those previously used. No probability distribution can claim to be the best match for a specific data set, and what is the best fit now may not be the best fit tomorrow.

The primary motivation behind this research stems from the mathematical elegance and practical utility of induced or equilibrium distributions in modelling lifetime phenomena. For any

continuous, non-negative random variable X representing a baseline lifetime model with survival function $1 - F(x)$ and a finite mean μ , the integral of its survival function equals its mean:

$$\mu = \int_0^{\infty} (1 - F(x)) dx \quad (1)$$

This mathematical identity guarantees that $g(x) = \frac{1-F(x)}{\mu}$ constitutes a valid probability density function (PDF). By incorporating the failure rate function, $r(x) = \frac{f(x)}{1-F(x)}$, this equilibrium transformation can also be structured as a weighted distribution:

$$g(x) = \frac{w(x)f(x)}{E[w(X)]}, \quad \text{where } w(x) = \frac{1}{r(x)}, \quad (2)$$

which provides a weighted interpretation of the baseline model studied by Patil et al. (1977) and Gupta and Kirmani (1990) [1, 2]. This weighted mechanism offers an intuitive, physics- and actuarial-based framework to analyse waiting times and ageing mechanisms, making it highly valuable across queueing systems, transportation engineering, actuarial modelling, and survival analysis.

Historically, generating new families of continuous distributions relied on classical methodologies, such as the transformation-based systems proposed by Johnson (1949) [3] or the differential-equation frameworks introduced by Pearson (1894) [4]. Modern advancements have yielded highly flexible generators, including the beta-generated framework by Eugene et al. (2002) [5], the Kumaraswamy-generalisation by Cordeiro and de Castro (2011)[6], and the Exponentiated T-X family by Alzaghal et al. (2013) [7]. Recently, equilibrium theory has emerged as an alternate frontier for model generation. Notably, Singh and Das (2023) [9] applied this transformation to the Lindley distribution to derive the induced-Lindley (i-Lindley) model, which features the cumulative distribution function (CDF):

$$F(x; \xi) = 1 - \frac{1 + \xi + \xi x}{1 + \xi} e^{-\xi x} \quad (3)$$

The above mathematical formulation in equation (3) was found to be identical to the Garima distribution proposed by Shanker (2016) [8], which later prompted subsequent studies on induced Garima (i-Garima) properties. Driven by these foundational advancements, the novelty of this paper lies in extending this exact line of reasoning beyond the Lindley family to the Xgamma distribution introduced by Sen et al. (2016) [10].

While existing literature provides a vast array of generalised families (such as the beta and Kumaraswamy-G classes), these traditional approaches often introduce many parameters, making estimation mathematically cumbersome. Furthermore, while the equilibrium transformation has been successfully applied to the Lindley and Garima frameworks to capture specific hazard structures, a distinct gap remains in the literature regarding its application to other flexible baseline models. Specifically, the standard Xgamma distribution, as described by Sen et al. (2016) [10], has emerged as a powerful model for lifetime data, yet the literature has not explored how its unique mathematical properties behave under an equilibrium transformation. This leaves a gap in available models that can natively inherit the beneficial traits of the Xgamma baseline while utilising an induced, weighted interpretation to model complex survival structures.

To bridge this gap, this paper explicitly contributes the derivation and introduction of a novel induced distribution known as the Induced Xgamma (i-Xg) model. Based on the framework of the Induced Xgamma (i-Xg) model, our contributions are as follows:

- **Derivation of a Novel Model:** Formulates and introduces the Induced Xgamma (*i-Xg*) distribution as a new mathematical model within lifetime and waiting-time statistics.
- **Methodological Expansion:** Successfully applies equilibrium theory and weighted distribution transformations to the standard Xgamma baseline, extending the application of induced distributions beyond classical Lindley and Garima frameworks.

- **Enhanced Analytical Flexibility:** Contributes a robust, highly flexible alternative that does not belong to the standard exponential family, making it uniquely suited to model complex ageing mechanisms.
- **Improved Practical Data Application:** Provides data practitioners in queueing systems, actuarial science, and survival analysis with an optimised statistical tool engineered specifically for fitting right-skewed lifetime data.

The rest of the article is framed as follows. Section 2 develops the proposed i-Xg distribution and its properties, such as moments, reliability, stochastic ordering, order statistics and their extreme domains, stress-strength reliability, entropy, and inequality curves. Section 3 is devoted to checking the identifiability condition of the proposed distribution. Seven estimation procedures are described in Section 4. In Section 5, simulation study results are presented and compared in terms of bias and mean squared error. Five data sets have been analysed to validate the proposed distribution in Section 6. Section 7 makes some concluding remarks.

2. THE PROPOSED DISTRIBUTION

By applying the equilibrium transformation to the Xgamma distribution, we obtain the induced Xgamma (i - Xg) model as

$$g(x; \xi) = \frac{\xi}{3 + \xi} \left(1 + \xi + \xi x + \frac{\xi^2 x^2}{2} \right) e^{-\xi x}, \quad x > 0, \xi > 0. \quad (4)$$

$$G(x; \xi) = 1 - \left[1 + \frac{\xi x}{\xi + 3} \left(2 + \frac{\xi x}{2} \right) \right] e^{-\xi x}. \quad (5)$$

It is important to note that the i - Xg distribution can be expressed as a mixture of three well-known distributions such as the exponential distribution with rate ξ , the gamma distribution with shape 2 and rate ξ , and the gamma distribution with shape 3 and rate ξ . Specifically,

$$g(x; \xi) = p_1 g_1(x) + p_2 g_2(x) + p_3 g_3(x),$$

where $g_1(x) = \text{Exp}(\xi)$, $g_2(x) = \text{Gamma}(2, \xi)$, $g_3(x) = \text{Gamma}(3, \xi)$, and the mixing weights are

$$p_1 = \frac{1 + \xi}{3 + \xi}, \quad p_2 = \frac{1}{3 + \xi}, \quad p_3 = \frac{1}{3 + \xi}.$$

Although the i-Xg admits a finite mixture representation, it is not reducible to the baseline Xgamma or any known special case. The plot of some PDF and CDF is given in Figure 1.

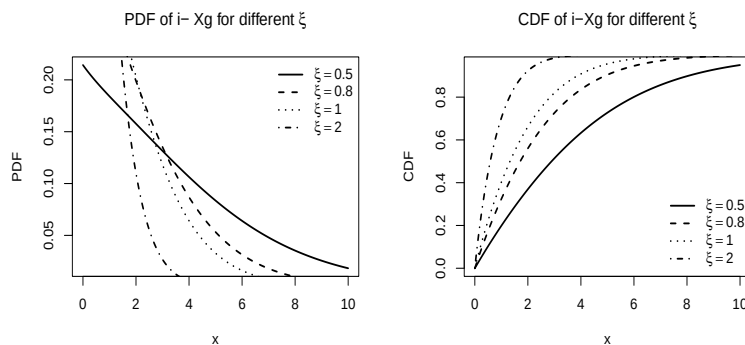


Figure 1: Plot of i-Xg for different values of ξ

2.1. Moments

The r^{th} moment about the origin is

$$\mu'_r = E[X^r] = \frac{r!}{\xi^r(\xi+3)} \left[(1+\xi) + \frac{(r+1)(r+4)}{2} \right].$$

For $r = 1, 2$ and 3 we get,

$$\begin{aligned}\mu'_1 &= \frac{\xi+6}{\xi(\xi+3)} = E(X) = \mu, \\ \mu'_2 &= \frac{2(\xi+10)}{\xi^2(\xi+3)}, \\ \mu'_3 &= \frac{6(\xi+15)}{\xi^3(\xi+3)}.\end{aligned}$$

and from above the variance can be obtained as,

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = \frac{\xi^2 + 14\xi + 24}{\xi^2(\xi+3)^2}.$$

The index of dispersion (ID), $ID(X) = \frac{\mu_2}{\mu} = \frac{\xi^2 + 14\xi + 24}{\xi(3+\xi)(6+\xi)}$ indicates the suitability of the distribution for under or over-dispersed data sets. $ID(X) \geq 1$ implies $\xi \geq 1.399$. Therefore, for $0 < \xi < 1.399$, the distribution is suitable for fitting over-dispersed data, whereas for $\xi > 1.399$, under-dispersed data fitting is appropriate. Here, the coefficient of variation (CV) is less than 1 for all values of ξ , also coefficient of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) are found to be more than 1 and 3 respectively for different values of ξ ; therefore the proposed distribution is positively skewed and leptokurtic (see, Figure 2).

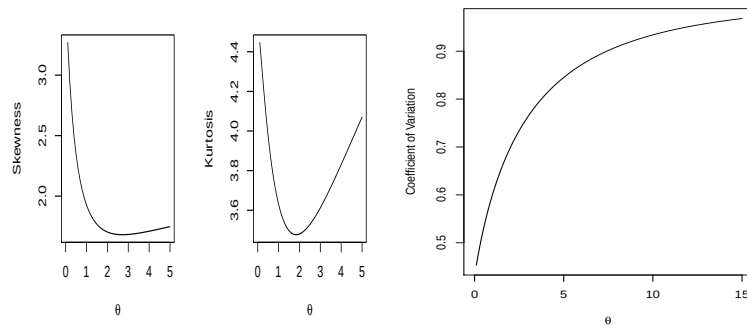


Figure 2: The graph of skewness, kurtosis and Coefficient of variation of i -Xg for different values of ξ

- The harmonic mean is obtained from

$$E\left(\frac{1}{X}\right) = \frac{\xi}{\xi+3}$$

The moment generating function is

$$M_X(t) = \frac{\xi(3\xi^2 + t^2 - 3\xi t)}{(\xi+3)(\xi-t)^3}$$

The central moments of i -Xg distribution is given by

$$E(X - \mu)^k = \frac{\Gamma(k+1)}{\xi^k(3+\xi)} \left(1 + \xi + \frac{(k+1)(k+4)}{2} \right) \sum_{j=0}^k \binom{k}{j} (-1)^{j-k} \mu^{j-k}$$

2.2. Mean Deviation about mean and median

• Mean Deviation about mean

$$\begin{aligned} MD_{mean} &= 2\mu F(\mu) - 2 \int_0^\mu xf(x)dx \\ &= 2\mu \left[1 - \left(1 + \frac{\xi}{\xi+3} (2\mu + \frac{\xi^2\mu^2}{2}) e^{-\xi\mu} \right) \right] \\ &\quad - \frac{2}{\xi(\xi+3)} \left[(6+\xi) \left((1+\mu\xi)(6+\xi) + \frac{5\mu^2\xi^2}{2} + \frac{\xi^3\mu^3}{2} \right) e^{-\mu\xi} \right] \end{aligned}$$

• Mean Deviation about median

$$\begin{aligned} MD_{median} &= \mu - 2 \int_0^\mu xf(x)dx \\ &= \mu - \frac{2}{\xi(\xi+3)} \left[(6+\xi) \left((1+\mu\xi)(6+\xi) + \frac{5\mu^2\xi^2}{2} + \frac{\mu^3\xi^3}{2} \right) e^{-\mu\xi} \right] \end{aligned}$$

2.3. Reliability properties

The reliability function is given by

$$R(x) = 1 - F(x) = \left[1 + \frac{\xi}{\xi+3} (2x + \frac{\xi x^2}{2}) \right] e^{-\xi x}.$$

The hazard rate function is given by

$$r(x) = \frac{f(x)}{R(x)} = \frac{\xi \left(1 + \xi + \xi x + \frac{\xi^2 x^2}{2} \right)}{\left(3 + \xi + 2\xi x + \frac{\xi^2 x^2}{2} \right)}.$$

It is to be noted that $\frac{\xi(1+\xi)}{3+\xi} \leq r(x) < \xi$.

The MRL function is given by

$$mrl(x) = P(X - x | X > x) = \frac{1}{\xi} \left(\frac{6 + \xi + 3\xi x + \frac{\xi^2 x^2}{2}}{3 + \xi + 2\xi x + \frac{\xi^2 x^2}{2}} \right)$$

Since

$$6 + \xi + 3\xi x + \frac{\xi^2 x^2}{2} > 3 + \xi + 2\xi x + \frac{\xi^2 x^2}{2}$$

we found that $mrl(x) > \frac{1}{\xi}$ and $\frac{1}{\xi} < mrl(x) \leq \frac{6+\xi}{\xi(3+\xi)}$. Note that $mrl(0) = \frac{\xi+6}{\xi(\xi+3)} = \mu'_1$.

3. STOCHASTIC ORDERING

Theorem 1. Let $X_i \sim \text{i-Xg}(\xi_i)$, $i = 1, 2$. If $\xi_1 \geq \xi_2$, then

$$X_1 \leq_{lr} X_2 \Rightarrow X_1 \leq_{hr} X_2 \Rightarrow X_1 \leq_{st} X_2 \text{ and } X_1 \leq_{cx} X_2.$$

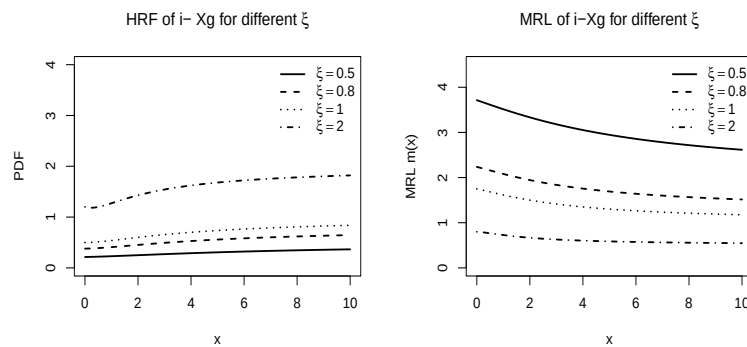


Figure 3: The graph of hazard rate and mean remaining life function of i -Xg for different values of ξ

Proof. Consider the likelihood ratio

$$\begin{aligned}\phi(x) &= \frac{g(x; \xi_1)}{g(x; \xi_2)} \\ &= \frac{1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2}}{1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2}} e^{-(\xi_1 - \xi_2)x}.\end{aligned}$$

Now, differentiating the $\log \phi(x)$ w.r.t. x , we get

$$\begin{aligned}\frac{d}{dx} \log \phi(x) &= \frac{\xi_1 + \xi_1^2 x}{1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2}} - \frac{\xi_2 + \xi_2^2 x}{1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2}} - (\xi_1 - \xi_2). \\ &= \frac{\xi_2^2 \left(1 + \frac{\xi_2 x}{2}\right)}{1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2}} - \frac{\xi_1^2 \left(1 + \frac{\xi_1 x}{2}\right)}{1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2}}.\end{aligned}$$

Let $A(\xi; x) = \frac{\xi^2 \left(1 + \frac{\xi x}{2}\right)}{1 + \xi + \xi x + \frac{\xi^2 x^2}{2}}$. A calculus argument shows $A(\xi; x)$ increases in ξ for $x \geq 0$.

Thus, if $\xi_1 \geq \xi_2$ then $A(\xi_1; x) \geq A(\xi_2; x)$ and

$$\frac{d}{dx} \log \phi(x) = A(\xi_2; x) - A(\xi_1; x) \leq 0.$$

Hence $\phi(x)$ is decreasing in x , proving $X_1 \leq_{lr} X_2$. Standard results give $X_1 \leq_{hr} X_2$ and $X_1 \leq_{st} X_2$, and for this model the convex order also follows. ■

From the above it is found that ξ acts as a failure-intensity control. A larger value of ξ corresponds to stochastically shorter lifetimes. When $\xi_1 > \xi_2$, the system modeled by X_1 tends to fail earlier, exhibits a uniformly higher hazard rate, and has lower survival probability than the system modeled by X_2 . Thus, ξ governs the rate of aging and reliability behavior of the i -Xg model.

3.1. Distribution of order statistics

Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the corresponding order statistics, then the PDF and CDF of r^{th} order statistics say, $Y = X_{(r)}$ is given by,

$$f_{r,n}(y) = \frac{n! \xi \left(1 + \xi + \xi y + \frac{\xi^2 y^2}{2}\right) e^{-\xi y}}{(\xi + 3)(r-1)!(n-r)!} \sum_{l=0}^{n-r} \binom{n-r}{l} \left[1 - \left(1 + \frac{\xi}{\xi + 3} \left(\xi y + \frac{\xi^2 y^2}{2}\right)\right) e^{-\xi y}\right]^{r+l-1}$$

$$F_{r,n}(y) = \sum_{j=r}^n \sum_{l=0}^{n-j} \binom{n}{j} \binom{n-j}{l} (-1)^l \left[1 - \left(1 + \frac{\xi}{\xi+3} (\xi y + \frac{\xi y^2}{2}) \right) e^{-\xi y} \right]^{j+l}$$

3.2. Extreme domain of order statistics

Let the random observations $(X_1, X_2, \dots, X_n)$ come from the proposed distribution. According to the central limit theorem, the $\frac{\bar{X}_n - \mu}{\sqrt{\mu_2}}$ approaches to standard normal distribution as $n \rightarrow \infty$. The asymptotic distributions of order statistics may be of interest. The highest and smallest order statistics are $X_{(n)} = \max(X_1, X_2, \dots, X_n)$ and $X_{(1)} = \min(X_1, X_2, \dots, X_n)$, respectively. These two extreme order statistics serve as representations of the lives of parallel and series systems and are important tools in reliability analysis. For the CDF in (5), it is noticed that

$$\lim_{t \rightarrow \infty} \frac{1 - G(t+x)}{1 - G(t)} = e^{-\xi x}$$

and

$$\lim_{t \rightarrow 0} \frac{G(tx)}{G(t)} = x.$$

So,

$$\lim_{n \rightarrow \infty} P[a_n(X_{(n)} - b_n) \leq x] = e^{-e^{-\xi x}}, \quad (6)$$

where, $a_n = 1$ and $b_n = G^{-1}(1 - \frac{1}{n})$ and

$$\lim_{n \rightarrow \infty} P[c_n(X_{(1)} - d_n) \leq x] = 1 - e^{-x}, \quad (7)$$

where, $c_n = 1$ and $d_n = G^{-1}(\frac{1}{n})$. It is to be noted that for the exponential distribution, the asymptotic distributions of $X_{(n)}$ and $X_{(1)}$ also take the forms given by (6) and (7). The largest order statistic converges to the extreme value distribution, and the smallest order statistic converges to the exponential distribution.

3.3. Stress- strength Reliability

Let the stress $X \sim i-Xg(\xi_1)$ and strain $Y \sim i-Xg(\xi_2)$ be independent to each other. The stress-strength index (R) is

$$R = P(X < Y) = \int_0^\infty F_X(y; \xi_1) g_Y(y; \xi_2) dy.$$

Let,

$$\begin{aligned} g_Y(y; \xi_2) &= K (1 + \xi_2 + \xi_2 y + \frac{\xi_2^2}{2} y^2) e^{-\xi_2 y} \\ F_X(y; \xi_1) &= 1 - \left(1 + a_1 (2y + b_1 y^2) \right) e^{-\xi_1 y} \end{aligned}$$

where $K = \xi_2 / (\xi_2 + 3)$, $a_1 = \frac{\xi_1}{\xi_1 + 3}$ and $b_1 = \frac{\xi_1}{2}$. Hence

$$R = \int_0^\infty g_Y(y; \xi_2) dy - K \int_0^\infty \left(1 + a_1 (2y + b_1 y^2) \right) \left(1 + \xi_2 + \xi_2 y + \frac{\xi_2^2}{2} y^2 \right) e^{-(\xi_1 + \xi_2)y} dy.$$

Since $\int_0^\infty g_Y(y; \xi_2) dy = 1$, we have $R = 1 - I$ with

$$I = K \int_0^\infty \underbrace{\left(1 + 2a_1 y + \frac{a_1 \xi_1}{2} y^2 \right)}_{A(y)} \underbrace{\left(1 + \xi_2 + \xi_2 y + \frac{\xi_2^2}{2} y^2 \right)}_{B(y)} e^{-\lambda y} dy, \quad \lambda = \xi_1 + \xi_2.$$

Define, $A(y) = 1 + c_1y + c_2y^2$ with $c_1 = 2a_1 = \frac{2\zeta_1}{\zeta_1 + 3}$ where, $c_2 = \frac{a_1\zeta_1}{2} = \frac{\zeta_1^2}{2(\zeta_1 + 3)}$ and $B(y) = b_0 + b_1y + b_2y^2$ where $b_0 = 1 + \zeta_2$, $b_1 = \zeta_2$, $b_2 = \frac{\zeta_2^2}{2}$. Then

$$A(y)B(y) = p_0 + p_1y + p_2y^2 + p_3y^3 + p_4y^4,$$

where

$$\begin{aligned} p_0 &= b_0 = 1 + \zeta_2, \\ p_1 &= b_0c_1 + b_1 = \frac{2\zeta_1(1 + \zeta_2)}{\zeta_1 + 3} + \zeta_2, \\ p_2 &= b_0c_2 + b_1c_1 + b_2 = \frac{(1 + \zeta_2)\zeta_1^2}{2(\zeta_1 + 3)} + \frac{2\zeta_1\zeta_2}{\zeta_1 + 3} + \frac{\zeta_2^2}{2}, \\ p_3 &= b_1c_2 + b_2c_1 = \frac{\zeta_2\zeta_1^2}{2(\zeta_1 + 3)} + \frac{\zeta_1\zeta_2^2}{\zeta_1 + 3}, \\ p_4 &= b_2c_2 = \frac{\zeta_1^2\zeta_2^2}{4(\zeta_1 + 3)}. \end{aligned}$$

Using $\int_0^\infty y^n e^{-\lambda y} dy = \frac{n!}{\lambda^{n+1}}$ for $\lambda > 0$, we get

$$I = K \left(\frac{p_0}{\lambda} + \frac{p_1}{\lambda^2} + \frac{2p_2}{\lambda^3} + \frac{6p_3}{\lambda^4} + \frac{24p_4}{\lambda^5} \right), \quad K = \frac{\zeta_2}{\zeta_2 + 3}, \quad \lambda = \zeta_1 + \zeta_2.$$

Therefore,

$$R = 1 - \frac{\zeta_2}{\zeta_2 + 3} \left(\frac{p_0}{\lambda} + \frac{p_1}{\lambda^2} + \frac{2p_2}{\lambda^3} + \frac{6p_3}{\lambda^4} + \frac{24p_4}{\lambda^5} \right), \quad \lambda = \zeta_1 + \zeta_2,$$

with $p_0, \dots, p_4$ given above. This closed form follows directly from the polynomial expansion and elementary gamma-integrals.

Since larger ζ yields stochastically smaller lifetimes for i-Xg, R is increasing in ζ_1 (stress becomes smaller) and decreasing in ζ_2 (strength becomes smaller). Hence designs with *large* ζ_1 and *small* ζ_2 achieve higher stress–strength reliability.

3.4. Inequality curves

The Lorenz curve is given by

$$\begin{aligned} L_F(x) &= \frac{1}{E(x)} \int_0^x t f(t) dt \\ &= \frac{1}{\zeta + 9} \left[(\zeta + 6) - \left((\zeta + 6)(x + 1) + \frac{(5 + \zeta x)}{2} \zeta^2 x^2 \right) e^{-\zeta x} \right] \end{aligned}$$

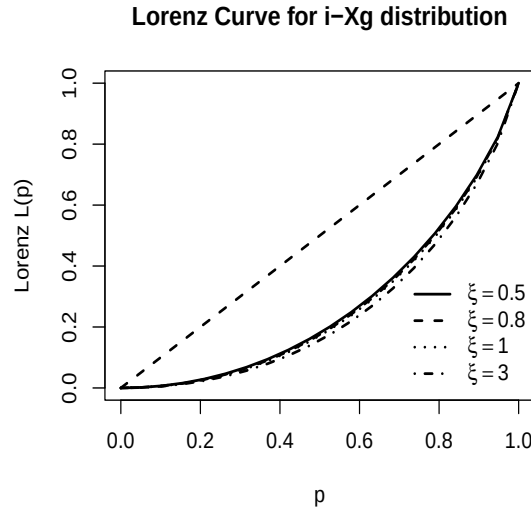


Figure 4: The Lorenz curve of i -Xg for different values of ξ

4. IDENTIFIABILITY CONDITION

We have to check whether the i -xgamma model is identifiable. Assume for two different parameter ξ_1 and ξ_2 we have the same distribution i.e. $f(x; \xi_1) = f(x; \xi_2)$. The distributions can be shown as,

$$f(x; \xi_1) = \frac{\xi_1}{3 + \xi_1} \left(1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2} \right) e^{-\xi_1 x}; \quad x > 0, \xi_1 > 0$$

$$f(x; \xi_2) = \frac{\xi_2}{3 + \xi_2} \left(1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2} \right) e^{-\xi_2 x}; \quad x > 0, \xi_2 > 0.$$

Now, from our assumption, we get

$$f(x; \xi_1) = f(x; \xi_2)$$

$$\text{or, } \frac{\xi_1}{3 + \xi_1} \left(1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2} \right) e^{-\xi_1 x} = \frac{\xi_2}{3 + \xi_2} \left(1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2} \right) e^{-\xi_2 x}$$

Taking "log" on both sides, we get

$$\log \left(\frac{\xi_1}{\xi_1 + 3} \right) + \log \left(1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2} \right) - \xi_1 x = \log \left(\frac{\xi_2}{\xi_2 + 3} \right) +$$

$$\log \left(1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2} \right) - \xi_2 x$$

$$\text{or, } \log \left(\frac{\xi_1(\xi_2 + 3)}{\xi_2(\xi_1 + 3)} \right) + \log \left(\frac{1 + \xi_1 + \xi_1 x + \frac{\xi_1^2 x^2}{2}}{1 + \xi_2 + \xi_2 x + \frac{\xi_2^2 x^2}{2}} \right) - (\xi_1 - \xi_2)x = 0$$

Since the expression equals zero for all $x > 0$, the coefficient of x must be zero, which implies $\xi_1 = \xi_2$. Thus, the parameters ξ_1 and ξ_2 are equal, which contradicts our assumption that they are different. This proves that the parameter ξ of the given statistical distribution i -Xg is identifiable, and that we can obtain precise estimates of its value.

5. ESTIMATION OF ξ USING DIFFERENT METHODS

Let $x_1, \dots, x_n$ be an i.i.d. sample from $i\text{-Xg}(\xi)$. Here we have used seven methods of estimation, namely,

1. Maximum Likelihood Estimation (MLE),
2. Ordinary Least Squares Estimation (OLSE),
3. Weighted Least Squares Estimation (WLSE),
4. Cramér-von Mises Estimation (CvME),
5. Maximum Product of Spacings Estimation (MPSE),
6. Anderson-Darling Estimation (ADE) and
7. Right-Tail Anderson Darling Estimation (RADE)

5.1. MLE

The likelihood and log-likelihood are

$$L(\xi) = \prod_{i=1}^n g(x_i; \xi) = \left(\frac{\xi}{\xi+3}\right)^n \prod_{i=1}^n \left(1 + \xi + \xi x_i + \frac{\xi^2 x_i^2}{2}\right) e^{-\xi x_i},$$

$$\ell(\xi) = n\{\log \xi - \log(\xi + 3)\} + \sum_{i=1}^n \log\left(1 + \xi + \xi x_i + \frac{\xi^2 x_i^2}{2}\right) - \xi \sum_{i=1}^n x_i.$$

Differentiating *w.r.t* ξ ,

$$\begin{aligned} \ell'(\xi) &= \frac{n}{\xi} - \frac{n}{\xi+3} + \sum_{i=1}^n \frac{1 + x_i + \xi x_i^2}{1 + \xi + \xi x_i + \frac{\xi^2 x_i^2}{2}} - \sum_{i=1}^n x_i \\ &= \frac{3n}{\xi(\xi+3)} + \sum_{i=1}^n \frac{1 + x_i + \xi x_i^2}{1 + \xi + \xi x_i + \frac{\xi^2 x_i^2}{2}} - n\bar{x}. \end{aligned}$$

The MLE $\hat{\xi}_{\text{MLE}}$ solves $\ell'(\xi) = 0$. A closed form is unavailable; use a 1-D root finder (e.g., Newton, BFGS) starting from a positive initial value.

5.2. OLSE

Let the ordered sample as $x_{1:n} < \dots < x_{n:n}$. All minimum-distance estimators below minimize a sum of squared (or weighted) differences between model CDF values $F(x_{i:n}; \xi)$. The normal equations use the CDF derivative

$$\eta_1(x | \xi) := \frac{\partial}{\partial \xi} F(x; \xi).$$

From $F(x; \xi) = 1 - A(\xi)e^{-\xi x}$ where $A(\xi) = 1 + \frac{\xi}{\xi+3}\left(2x + \frac{\xi x^2}{2}\right)$,

$$\eta_1(x | \xi) = -A'(\xi)e^{-\xi x} + A(\xi)x e^{-\xi x},$$

with

$$A'(\xi) = \frac{3}{(\xi+3)^2} \left(2x + \frac{\xi x^2}{2}\right) + \frac{\xi}{\xi+3} \cdot \frac{x^2}{2}.$$

So,

$$\eta_1(x | \xi) = \left[x \left(1 + \frac{\xi}{\xi + 3} \left(2x + \frac{\xi x^2}{2} \right) \right) - \frac{3}{(\xi + 3)^2} \left(2x + \frac{\xi x^2}{2} \right) - \frac{\xi}{\xi + 3} \cdot \frac{x^2}{2} \right] e^{-\xi x} \quad (8)$$

Define

$$S(\xi) = \sum_{i=1}^n \left[F(x_{i:n}; \xi) - \frac{i}{n+1} \right]^2.$$

The LSE $\hat{\xi}_{\text{LSE}}$ minimizes $S(\xi)$ and solves the normal equation

$$\sum_{i=1}^n \left[F(x_{i:n}; \xi) - \frac{i}{n+1} \right] \eta_1(x_{i:n} | \xi) = 0.$$

where $\eta_1(x_{i:n} | \xi)$ is given in equation 8.

5.3. WLSE

With Beta-variance weights $w_i = \frac{(n+1)^2}{(n+2)i(n-i+1)}$,

$$W(\xi) = \sum_{i=1}^n w_i \left[F(x_{i:n}; \xi) - \frac{i}{n+1} \right]^2,$$

and the WLSE $\hat{\xi}_{\text{WLSE}}$ satisfies

$$\sum_{i=1}^n w_i \left[F(x_{i:n}; \xi) - \frac{i}{n+1} \right] \eta_1(x_{i:n} | \xi) = 0.$$

where $\eta_1(x_{i:n} | \xi)$ is given in equation 8.

5.4. CvME

Using the standard plotting positions $\frac{2i-1}{2n}$,

$$C(\xi) = \frac{1}{12n} + \sum_{i=1}^n \left[F(x_{i:n}; \xi) - \frac{2i-1}{2n} \right]^2,$$

and the CvM estimator $\hat{\xi}_{\text{CvME}}$ solves

$$\sum_{i=1}^n \left[F(x_{i:n}; \xi) - \frac{2i-1}{2n} \right] \eta_1(x_{i:n} | \xi) = 0.$$

where $\eta_1(x_{i:n} | \xi)$ is given in equation 8.

5.5. MPSE

Define spacings on the transformed scale

$$D_i(\xi) = F(x_{i:n}; \xi) - F(x_{i-1:n}; \xi), \quad F(x_{0:n}; \xi) = 0, \quad F(x_{n+1:n}; \xi) = 1.$$

The MPS estimator maximizes the geometric mean

$$G(\xi) = \left(\prod_{i=1}^{n+1} D_i(\xi) \right)^{1/(n+1)} \iff H(\xi) = \frac{1}{n+1} \sum_{i=1}^{n+1} \log D_i(\xi) \text{ is maximized.}$$

Differentiating $H(\xi)$ w.r.t. ξ gives

$$\frac{\partial H}{\partial \xi} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{\eta_1(x_{i:n} | \xi) - \eta_1(x_{i-1:n} | \xi)}{D_i(\xi)} = 0.$$

5.6. ADE

Define,

$$A(\xi) = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \left\{ \log F(x_{i:n}; \xi) + \log [1 - F(x_{n+1-i:n}; \xi)] \right\},$$

We obtain ADE by minimising $A(\xi)$ w.r.t. ξ and the normal equation as

$$\sum_{i=1}^n (2i-1) \left[\frac{\eta_1(x_{i:n} | \xi)}{F(x_{i:n}; \xi)} - \frac{\eta_1(x_{n+1-i:n} | \xi)}{1 - F(x_{n+1-i:n}; \xi)} \right] = 0.$$

5.7. RADE

The RADE uses

$$R(\xi) = \frac{n}{2} - 2 \sum_{i=1}^n F(x_{i:n}; \xi) - \frac{1}{n} \sum_{i=1}^n (2i-1) \log [1 - F(x_{n+1-i:n}; \xi)],$$

and satisfies

$$-2 \sum_{i=1}^n \eta_1(x_{i:n} | \xi) + \frac{1}{n} \sum_{i=1}^n (2i-1) \frac{\eta_1(x_{n+1-i:n} | \xi)}{1 - F(x_{n+1-i:n}; \xi)} = 0.$$

Here all estimating equations are one-dimensional in ξ and can be solved by a bracketed method (e.g., Brent) in latest R-4.5.2 which enforces $\xi > 0$. We take,

$$\xi_{\text{MLE}}^{(0)} = \min \left\{ \frac{1}{\bar{x}}, \frac{3}{\bar{x}} \right\}, \quad \xi_{\text{MD}}^{(0)} : \text{minimize the objective on a coarse grid (say, } [10^{-3}, 10]).$$

For stability in MPS/AD/RAD, clip F to $[\epsilon, 1 - \epsilon]$ with $\epsilon \approx 10^{-12}$ to avoid $\log 0$, and regularize tiny spacings $D_i(\xi)$ similarly.

6. SIMULATION STUDY

To evaluate the performance of the proposed estimation methods (MLE, LSE, WLSE, CvME, MPS, ADE, and RADE) for the i-Xg distribution, a Monte Carlo simulation experiment is conducted. The primary objective of this study is to compare the bias and mean squared error (MSE) of each estimator across different sample sizes and parameter values.

Random samples from the i-Xg distribution are generated using its mixture representation. For a given parameter $\xi > 0$, the distribution can be expressed as a mixture of three components:

$$p_1 = \frac{1 + \xi}{\xi + 3}, \quad p_2 = \frac{1}{\xi + 3}, \quad p_3 = \frac{1}{\xi + 3},$$

corresponding respectively to an exponential(ξ) distribution, a gamma(2, ξ) distribution, and a gamma(3, ξ) distribution. Thus, a simulated observation from the i-Xg distribution is obtained by

1. Generate $U \sim \text{Uniform}(0, 1)$.
2. If $U \leq p_1$, set $X \sim \text{Exponential}(\xi)$.
3. Else if $p_1 < U \leq p_1 + p_2$, set $X \sim \text{Gamma}(2, \xi)$.
4. Otherwise, set $X \sim \text{Gamma}(3, \xi)$.

This algorithm produces exact samples from the i-Xg model without relying on numerical inversion or approximation techniques.

The simulation is carried out for sample sizes

$$n = 10, 20, 40, 70, 100,$$

and parameter values

$$\zeta = 0.1, 0.5, 1.0, 1.5.$$

For each (n, ζ) combination, $N = 5000$ independent datasets are generated. Each dataset is analyzed using all seven estimation techniques implemented in R, ensuring that the parameter space restriction $\zeta > 0$ is enforced during optimization and two evaluation metrics are computed as

$$\text{Bias}(\hat{\zeta}) = E[\hat{\zeta}] - \zeta, \quad \text{MSE}(\hat{\zeta}) = E[(\hat{\zeta} - \zeta)^2].$$

The Bias and MSE plot for different ζ are shown in Figure 5.

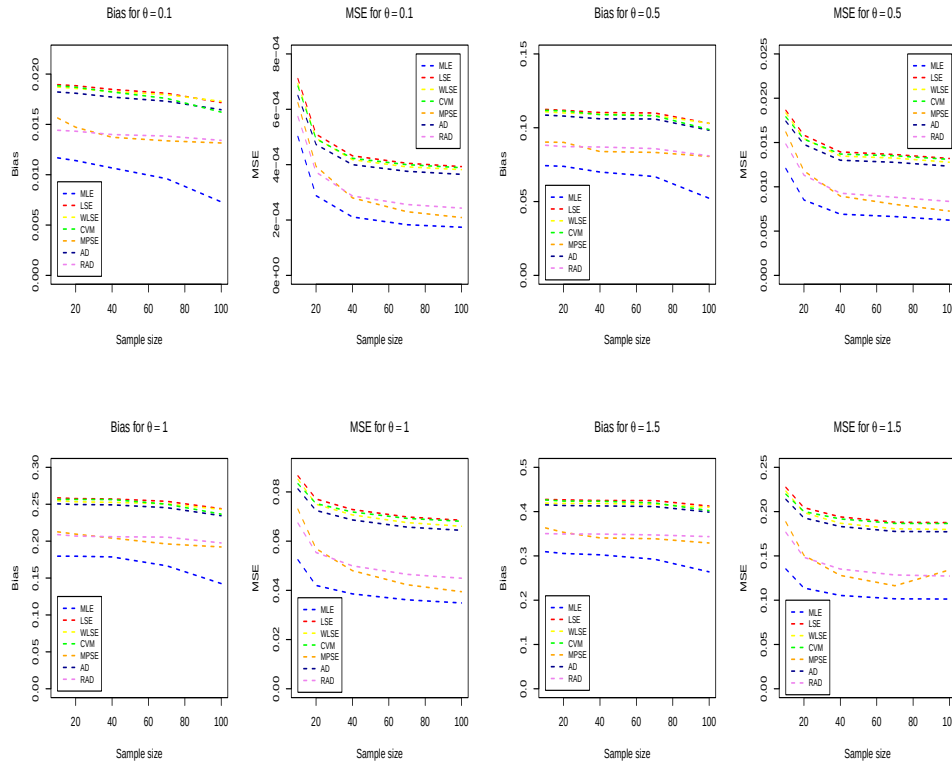


Figure 5: Bias and MSE plot for different ζ

From the Table 1 and Table 2 it is observed that lower values of both bias and MSE indicate superior estimator performance. All estimators are compared to determine their convergence behaviour as the sample size increases. In general, consistent estimators are expected to exhibit decreasing bias and MSE as n increases, confirming desirable asymptotic properties. Though, the minor increase observed for ADE and RADE at $n=100$ relative to $n=70$ is attributable to Monte Carlo variability and does not indicate estimator inconsistency, since the overall trend remains decreasing.

Table 1: True value of ξ , average bias and std. error (in parentheses) of the different estimation procedures for i -the Xg distribution

ξ	n	MLE	LSE	WLSE	CVM	MPSE	AD	RAD
0.1	10	0.007179 (0.021762)	0.017169 (0.025945)	0.017319 (0.025674)	0.016201 (0.025425)	0.015536 (0.024456)	0.016349 (0.024942)	0.013307 (0.023437)
	20	0.009822 (0.017093)	0.018036 (0.022530)	0.018014 (0.022286)	0.017531 (0.022166)	0.014954 (0.020020)	0.017309 (0.021735)	0.013917 (0.019335)
	40	0.010687 (0.014472)	0.018472 (0.020766)	0.018224 (0.020405)	0.018213 (0.020547)	0.013738 (0.016686)	0.017703 (0.019986)	0.014001 (0.016922)
	70	0.011344 (0.013455)	0.018718 (0.019995)	0.018477 (0.019678)	0.018569 (0.019859)	0.013314 (0.015090)	0.017980 (0.019254)	0.014222 (0.015884)
	100	0.011490 (0.012979)	0.018755 (0.019654)	0.018535 (0.019382)	0.018650 (0.019556)	0.012974 (0.014274)	0.018024 (0.018923)	0.014210 (0.015394)
	10	0.056032 (0.110675)	0.105953 (0.138717)	0.105772 (0.137173)	0.101781 (0.135948)	0.093744 (0.128869)	0.101434 (0.133955)	0.084459 (0.123437)
0.5	20	0.065629 (0.093265)	0.108759 (0.125887)	0.108006 (0.124361)	0.106598 (0.124140)	0.088872 (0.109063)	0.104644 (0.121699)	0.085509 (0.106873)
	40	0.071072 (0.084370)	0.110747 (0.118904)	0.109299 (0.117141)	0.109636 (0.117902)	0.084927 (0.095776)	0.106809 (0.114990)	0.086766 (0.097357)
	70	0.073885 (0.081264)	0.112102 (0.116606)	0.110621 (0.114993)	0.111463 (0.116003)	0.082866 (0.089290)	0.108224 (0.112772)	0.087687 (0.093691)
	100	0.074492 (0.079493)	0.112324 (0.115408)	0.110623 (0.113586)	0.111874 (0.114975)	0.081251 (0.085745)	0.108439 (0.111545)	0.087702 (0.091832)
	10	0.148040 (0.231379)	0.247865 (0.296582)	0.247289 (0.294643)	0.240555 (0.291037)	0.217815 (0.273609)	0.238763 (0.287509)	0.202601 (0.262300)
	20	0.165271 (0.205865)	0.252654 (0.277219)	0.249990 (0.274027)	0.248808 (0.273896)	0.208287 (0.239175)	0.244093 (0.268959)	0.203821 (0.235294)
1	40	0.173788 (0.192446)	0.255933 (0.267270)	0.251542 (0.262755)	0.253951 (0.265415)	0.199349 (0.215060)	0.247569 (0.259085)	0.205316 (0.220271)
	70	0.179196 (0.190050)	0.257952 (0.264772)	0.253069 (0.259756)	0.256809 (0.263674)	0.195831 (0.205500)	0.249809 (0.256724)	0.206894 (0.215931)
	100	0.182646 (0.189699)	0.260250 (0.264776)	0.255288 (0.259718)	0.259445 (0.263993)	0.195054 (0.201513)	0.252224 (0.256801)	0.209211 (0.215159)
	10	0.267098 (0.365240)	0.416412 (0.475352)	0.413645 (0.471172)	0.406140 (0.467127)	0.366197 (0.433115)	0.401970 (0.461382)	0.346370 (0.418505)
	20	0.288168 (0.333633)	0.419473 (0.448210)	0.414442 (0.442480)	0.414138 (0.443416)	0.349526 (0.384698)	0.406525 (0.435454)	0.344970 (0.381363)
	40	0.302127 (0.323665)	0.425893 (0.439539)	0.418370 (0.431905)	0.423143 (0.436928)	0.338568 (0.356992)	0.413160 (0.427010)	0.348785 (0.366494)
1.5	70	0.310101 (0.322629)	0.428870 (0.436937)	0.420506 (0.428518)	0.427278 (0.435392)	0.333651 (0.344970)	0.416635 (0.424826)	0.351497 (0.362039)
	100	0.308388 (0.316863)	0.427658 (0.4329708)	0.418520 (0.4238929)	0.426545 (0.4318791)	0.339384 (0.3645649)	0.415331 (0.4207491)	0.348979 (0.3560576)

7. APPLICATION TO LIFETIME DATA

The i -Xg distribution has been used to fit five data sets, and it has been compared to the i -garima and x gamma distributions. Dataset I is taken from Lawless (2003), pp. 204 [12], which contains the failure times (mm) for a sample of 15 electronic components in an accelerated life test. Dataset II is obtained from Lawless (2003), pp. 263 [12], which reports the number of cycles to failure for 25 (100 cm) yarn specimens tested at a particular strain level. Dataset III contains remission times (in weeks) for 20 leukaemia patients randomly assigned to a specific treatment, taken from Atia et al. (2023) [13]. Dataset IV is about the survival time (in days) of 72 guinea pigs infected with virulent tubercle bacilli reported by Bjerkedal, T. (1960) [14], and Dataset V is about the time to

Table 2: True value of ξ , MSE and std. error (in parentheses) of the different estimation procedures for the i -Xg distribution

ξ	n	MLE	LSE	WLSE	CVM	MPSE	AD	RAD
0.1	10	0.000474	0.000673	0.000659	0.000646	0.000598	0.000622	0.000549
		(0.021762)	(0.025945)	(0.025674)	(0.025425)	(0.024456)	(0.024942)	(0.023437)
	20	0.000292	0.000508	0.000497	0.000491	0.000401	0.000472	0.000374
		(0.017093)	(0.022530)	(0.022286)	(0.022166)	(0.020020)	(0.021735)	(0.019335)
	40	0.000209	0.000431	0.000416	0.000422	0.000278	0.000399	0.000286
		(0.014472)	(0.020766)	(0.020405)	(0.020547)	(0.016686)	(0.019986)	(0.016922)
	70	0.000181	0.000400	0.000387	0.000394	0.000228	0.000371	0.000252
		(0.013455)	(0.019995)	(0.019678)	(0.019859)	(0.015090)	(0.019254)	(0.015884)
	100	0.000168	0.000386	0.000376	0.000382	0.000204	0.000358	0.000237
		(0.012979)	(0.019654)	(0.019382)	(0.019556)	(0.014274)	(0.018923)	(0.015394)
	0.5	10	0.012249	0.019242	0.018816	0.018482	0.016607	0.017944
			(0.110675)	(0.138717)	(0.137173)	(0.135948)	(0.128869)	(0.133955)
		20	0.008698	0.015847	0.015466	0.015411	0.011895	0.014811
			(0.093265)	(0.125887)	(0.124361)	(0.124140)	(0.109063)	(0.121698)
		40	0.007118	0.014138	0.013722	0.013901	0.009173	0.013223
			(0.084370)	(0.118904)	(0.117141)	(0.117902)	(0.095776)	(0.114990)
		70	0.006604	0.013597	0.013223	0.013457	0.007973	0.012717
			(0.081264)	(0.116606)	(0.114993)	(0.116003)	(0.089290)	(0.112772)
		100	0.006319	0.013319	0.012902	0.013219	0.007352	0.012442
			(0.079493)	(0.115408)	(0.113586)	(0.114975)	(0.085745)	(0.111545)
1	10	0.053536	0.087961	0.086814	0.084702	0.074862	0.082661	0.068801
		(0.231379)	(0.296582)	(0.294643)	(0.291037)	(0.273609)	(0.287509)	(0.262300)
	20	0.042380	0.076850	0.075091	0.075019	0.057205	0.072339	0.055363
		(0.205865)	(0.277219)	(0.274027)	(0.273896)	(0.239175)	(0.268959)	(0.235294)
	40	0.037035	0.071433	0.069040	0.070445	0.046251	0.067125	0.048519
		(0.192446)	(0.267270)	(0.262755)	(0.265415)	(0.215060)	(0.259085)	(0.220271)
	70	0.036119	0.070104	0.067473	0.069524	0.042230	0.065907	0.046626
		(0.190050)	(0.264772)	(0.259756)	(0.263674)	(0.205500)	(0.256724)	(0.215931)
	100	0.035986	0.070106	0.067454	0.069692	0.040608	0.065947	0.046293
		(0.189699)	(0.264776)	(0.259718)	(0.263993)	(0.201513)	(0.256801)	(0.215159)
	1.5	10	0.133400	0.225959	0.222003	0.218208	0.187589	0.212874
			(0.365240)	(0.475352)	(0.471172)	(0.467127)	(0.433115)	(0.461382)
		20	0.111311	0.200892	0.195789	0.196617	0.147992	0.189620
			(0.333633)	(0.448210)	(0.442480)	(0.443416)	(0.384698)	(0.435454)
		40	0.104759	0.193195	0.186542	0.190906	0.127443	0.182337
			(0.323664)	(0.439539)	(0.431905)	(0.436928)	(0.356992)	(0.427010)
		70	0.104089	0.190914	0.183628	0.189566	0.119005	0.180477
			(0.322629)	(0.436937)	(0.428518)	(0.435391)	(0.344970)	(0.424826)
		100	0.100402	0.187464	0.179685	0.186520	0.132908	0.177030
			(0.316863)	(0.432971)	(0.423893)	(0.431879)	(0.364565)	(0.420749)

failure of 18 electronic devices used in Wang (2000) [15].

We can observe that the means of all the datasets are higher than their medians, indicating that the data are positively (right) skewed with positive support. All of these datasets can be used as lifetime data is evidence from Table 3. The first four datasets are nearly concave, indicating that the data are drawn from a distribution with an increasing failure rate (IFR). Consequently, it has a decreasing mean remaining life (DMRL). For Dataset V, it is approximately convex-concave, indicating that the data come from a distribution with a bathtub failure rate. For each data set, the probability model is chosen using Akaike's Information Criterion ($AIC = -2 \log\text{-likelihood} +$

2k, where k is the number of parameters in the model). A lower AIC indicates a better fit for the model. Table 3 presents some fundamental summary statistics, and calculated AICs are shown in Table 4 and demonstrate that the i-Xg is a better fit. For each data set, the histogram and the three fitted distributions such as xgamma, i-garima, and i-Xg, are plotted in Figures 6 for comparison.

Table 3: Summary of the Datasets

	Min	Q_1	Q_2 (Median)	Mean	Q_3	Max
Dataset I	1.40	11.45	22.20	27.55	41.80	66.20
Dataset II	15.0	86.0	175.0	178.3	224.0	653.0
Dataset III	1.00	7.00	16.50	19.55	28.25	49.00
Dataset IV	12.00	54.75	70.00	99.82	112.75	376.00
Dataset V	5.00	53.25	155.00	172.11	281.00	420.00

Table 4: Fitting of different distributions and corresponding negative log-likelihood and AIC values

Dataset	Xgamma			I-Garima			I-Xg		
	$\hat{\xi}$	$-\log L$	AIC	$\hat{\xi}$	$-\log L$	AIC	$\hat{\xi}$	$-\log L$	AIC
Dataset I	0.0066	65.4170	134.834	0.0489	64.4600	132.920	0.0725	64.0600	132.120
Dataset II	0.0189	154.8800	313.760	0.0076	154.0100	312.020	0.0115	152.9850	309.970
Dataset III	0.1620	80.1163	164.233	0.0687	79.0970	162.194	0.1017	78.5750	161.150
Dataset IV	0.3167	325.8400	655.680	0.1360	326.8540	657.708	0.2017	323.0910	650.182
Dataset V	0.0189	116.0758	236.152	0.0078	110.3970	224.794	0.0115	110.0940	224.188

8. CONCLUDING REMARKS

This study presents an innovative one-parameter lifetime model, the Induced Xgamma (i-Xg) distribution, formulated through an equilibrium (induced) transformation of the existing Xgamma distribution. The model is very flexible and can accurately capture both under-dispersed and over-dispersed behaviour in reliability data. Its capacity to depict various hazard shapes, including increasing, decreasing, bathtub, and inverted-bathtub patterns, highlights its practical utility in reliability engineering and longevity studies. We examine important distributional properties, including moments, stochastic ordering, inequality measures, and reliability characteristics. We also show that the model is identifiable. We examine several traditional estimation methods and assess their performance in simulations. The empirical findings underscore the robustness and efficacy of the estimation methodologies. The i-Xg model is a desirable tool for practitioners evaluating duration and survival-type data, offering greater flexibility and better goodness-of-fit than competing models, as demonstrated in applications across several real-world datasets. All things considered, the suggested distribution adds something useful to the family of continuous reliability models and might encourage further developments and applications in related statistical fields.

ACKNOWLEDGEMENT

The authors are very thankful to the anonymous referee and the editor for their valuable comments to improve the quality of the article.

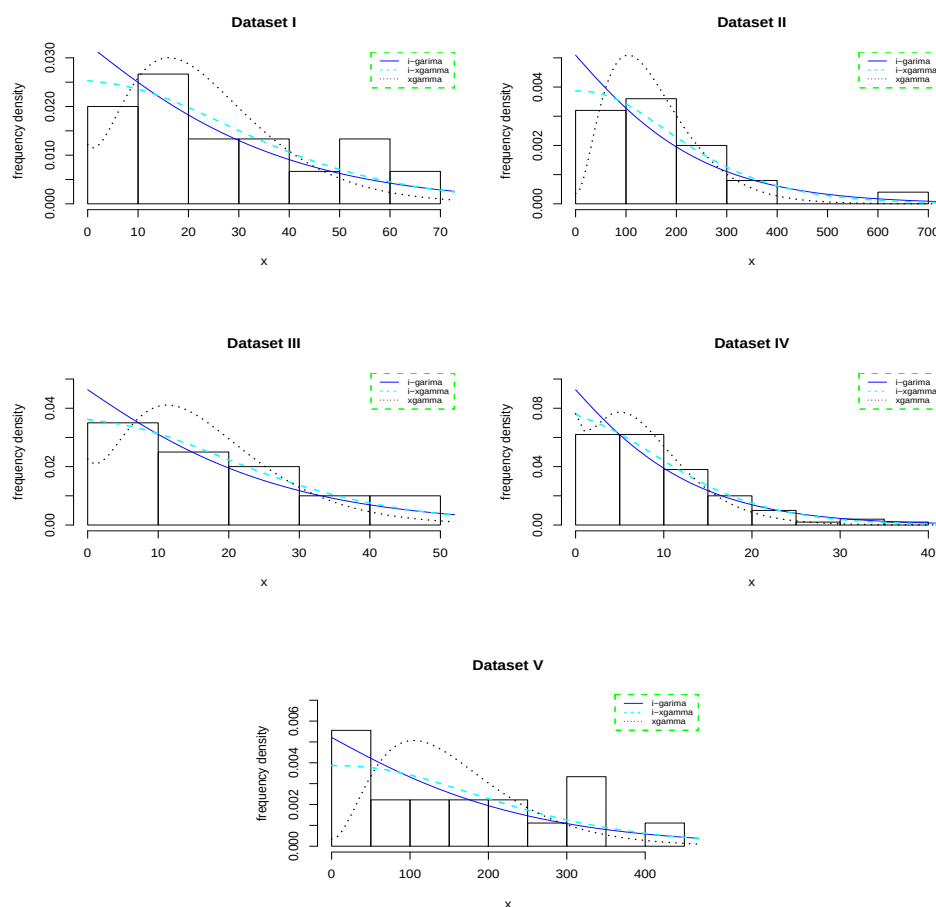


Figure 6: Histogram and fitted distributions for Data set I-V

CONFLICT OF INTEREST

All authors declare no conflicts of interest.

DISCLAIMER

Use of Gen AI: During the preparation of this work, the author(s) utilised ChatGPT from OpenAI to improve language and readability in the Introduction section. Following the use of this tool, the author(s) reviewed and edited the content as necessary and take full responsibility for the entire publication's content.

REFERENCES

- [1] Patil, G. et al. (1977). The weighted distribution: A survey of their applications.
- [2] Gupta, R. C. and Kirmani, S. (1990). The role of weighted distributions in stochastic modeling. Communications in Statistics-Theory and methods, 19(9), 3147-3162. <https://doi.org/10.1080/03610929008830371>
- [3] Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. Biometrika, 36(1/2), 149-176. <https://doi.org/10.1093/biomet/36.1-2.149>

- [4] Pearson, K. (1894). Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A, 185, 71-110. <https://doi.org/10.1098/rsta.1894.0003>
- [5] Eugene, N., Lee, C., and Famoye, F. (2002). Beta-normal distribution and its applications. Communications in Statistics-Theory and methods, 31(4), 497- 512. <https://doi.org/10.1081/STA-120003130>
- [6] Cordeiro, G. and de Castro, M. (2011). A new family of generalized distributions. Journal of Statistical Computation and Simulation, 81(7), 883-898. <https://doi.org/10.1080/00949650903530745>
- [7] Alzaghal, A., Famoye, F., and Lee, C. (2013). Exponentiated T-X Family of Distributions with Some Applications. International Journal of Statistics and Probability, 2(3), 31. <https://doi.org/10.5539/ijsp.v2n3p31>
- [8] Shanker, R. (2016). Garima distribution and its application to model behavioral science data. Biometrics & Biostatistics International Journal, 4(7), 1-9. <https://doi.org/10.15406/bbij.2016.04.00116>
- [9] Singh, Brijesh P., and Dhar Das, Utpal. (2023). On an Induced Distribution and its Statistical Properties. Revstat - Statistical Journal. 21. 303-319. <https://doi.org/10.57805/revstat.v21i3.360>
- [10] Sen, S., Maiti, S. S., and Chandra, N. (2016). The xgamma distribution: statistical properties and application. Journal of Modern Applied Statistical Methods, 15(1):38. <https://doi.org/10.22237/jmasm/1462077420>
- [11] Rényi, Alfréd. (1961). On measures of entropy and information." Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: Contributions to the theory of statistics. Vol. 4. University of California Press.
- [12] Lawless, Jerald F. (2011). Statistical models and methods for lifetime data. John Wiley & Sons. <https://doi.org/10.1002/9781118033005>
- [13] Atia, Ahmed Galal, et al. (2023). Forecasting Remission Time of a Treatment Method for Leukaemia as an Application to the Statistical Inference Approach. <https://doi.org/10.18576/isl/120208>
- [14] Bjerkedal, Tor. (1960). Acquisition of Resistance in Guinea Pigs Infected with Different Doses of Virulent Tubercle Bacilli, 130-48. <https://doi.org/10.1093/oxfordjournals.aje.a120129>
- [15] Wang, Fu-Kwun. (2000). A new model with a bathtub-shaped failure rate using an additive Burr XII distribution. Reliability Engineering & System Safety, 305-312. [https://doi.org/10.1016/S0951-8320\(00\)00066-1](https://doi.org/10.1016/S0951-8320(00)00066-1)

THE AKAIKE INFORMATION CRITERION AND ARTIFICIAL INTELLIGENCE

HENDRIK SCHÄBE

•

TÜV Rheinland InterTraffic GmbH, Germany

dr.hendrik.schaebe@gmail.com

Abstract

In this paper we use the Akaike information criterion to compare the use of parametric models with the use of very flexible models with a large parameter set, as in artificial intelligence models. It turns out that information is lost in models with many parameters, compared with using the real model.

Keywords: Akaike Information Criterion, Artificial Intelligence, Likelihood

1. INTRODUCTION

The use of artificial intelligence (AI) systems is on the rise. They are widely applied and used. So, there seems to be a big advantage of AI use. On the other hand, one must also ask: what do we lose?

In an earlier paper Schäbe and Shubinsky [1] considered what would be the correct approach to use artificial intelligence in order to gain efficiency but not to lose too much.

In this paper we will apply the Akaike information criterion to find an answer to the question, what we would lose. However, AI models are complex and there are many different classes of models. So, we can restrict ourselves to rather simple cases. AI models are mostly just statistical models, judging from the fundamental publications [2, 3], so we will use the Akaike information criterion to compare the behaviour of a AI model with the one of an ordinary statistical model. In section two, we explain the Akaike information criterion [4] and give an example, how it is applied for a kernel estimator comparing it to a parametric model. In section three, we consider a general approach. The fourth section is dedicated to the conclusions.

2. THE AKAIKE INFORMATION CRITERION

2.1. The Definition of the Akaike Information Criterion

The Akaike information criterion has been introduced by Akaike in 1974 [4]. Assume, we have a sample of size n , say $X_1, \dots, X_n$ and two different statistical models, with densities $f(x)$ and $g(x)$ having a different number of parameters k_1 and k_2 . Here θ and η are the parameters (θ being a k_1 -dimensional vector and η being a k_2 -dimensional vector) of the densities $f(x | \theta)$ and $g(x | \eta)$, whereas $\hat{\theta}$ and $\hat{\eta}$ are the coinciding values obtained as maximum likelihood estimators. Here and throughout the paper we assume that the densities exist and that they fulfil the regularity conditions as given e.g. in [5] for the maximum likelihood estimators to exist, converge in probability and being asymptotically unbiased.

Then the Akaike information criterion (AIC) compares the following quantities

$$AIC_f = k_1 - \sum_{i=1}^n \ln f(X_i | \hat{\theta}), \quad (1)$$

$$\text{AIC}_g = k_2 - \sum_{i=1}^n \ln g(X_i | \hat{\eta}), \quad (2)$$

for both models. The model with the minimal AIC value is preferred. Should the number of parameters be the same for both models, simply the model with the largest likelihood function computed for the parameter values coinciding with maximum likelihood estimators is chosen. For different numbers of parameters, a penalty term is introduced – just the number of parameters.

In the following two subsections we assume that the sample comes from a normal distribution and we will apply two different estimators. The one will be the normal distribution, i.e. the real underlying distribution of the sample. The other model is the kernel estimator. Vapnik [3] has considered this as kind of an AI application.

2.2. The AIC for a normal model

The AIC for a normal distribution model is computed as

$$\text{AIC}_N = 2k - 2 \sum_{i=1}^n \ln f(X_i | \hat{m}, \hat{\sigma}). \quad (3)$$

Since the normal distribution has 2 parameters, $k = 2$. Here $f(X_i | \hat{m}, \hat{\sigma})$ is the normal density with estimated mean and standard deviation, with

$$f(x | m, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right). \quad (4)$$

Now we compute the expectation of the AIC_N . We will neglect higher order terms in $1/n$, by partially substitution estimators by its values, e.g. for the standard deviation. Correlation between sample elements, X_i and the estimated deviation will be neglected as higher order terms. We get

$$\begin{aligned} E(\text{AIC}_N) &= 4 + 2n \ln(\sqrt{2\pi}\sigma) + nE\left[\left(\frac{X_i - m}{\sigma}\right)^2\right] \\ &= 4 + 2n \left(\ln(\sqrt{2\pi}\sigma) + \frac{1}{2}\right). \end{aligned} \quad (5)$$

The expected AIC_N for the normal distribution still depends on the variance of the distribution. Also, one can see that the AIC is proportional to n .

2.3. AIC for the Kernel Estimator

In the next step, we compute the AIC for a kernel estimator. The kernel estimator is defined by Rosenblatt and Parzen [6, 7] as

$$\tilde{f}(x) = \sum_{i=1}^n \frac{1}{nh} k\left(\frac{x - X_i}{h}\right). \quad (6)$$

For simplicity, we take a normal kernel, i.e.

$$k(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad (7)$$

which leads to

$$\tilde{f}(x) = \sum_{i=1}^n \frac{1}{nh\sqrt{2\pi}} \exp\left(-\frac{(x - X_i)^2}{2h^2}\right). \quad (8)$$

For the bandwidth we use the approximation proposed by Silverman [8], i.e.

$$h = \left(\frac{4}{3}\right)^{1/5} \hat{\sigma} n^{-1/5} \approx 1.06 \hat{\sigma} n^{-1/5}. \quad (9)$$

where $\hat{\sigma}$ is the standard deviation estimated from the sample. A more precise expression has been given by [9] using a Gram-Charlier approximation for the unknown distribution. Now, we compute the AIC_k for the kernel estimator as

$$AIC_k = 2n - 2 \sum_{i=1}^n \ln \tilde{f}(X_i). \quad (10)$$

Here the number of parameters is n , since in fact each value of the sample acts as a parameter. Since the bandwidth is derived from the standard deviation and this is also computed from the same n values X_i , the number of parameters stays n . Now we compute the expectation of AIC_k . We will again neglect higher order terms in $1/n$, by partially substitution estimators by its values, e.g. for the standard deviation. Correlation between sample elements, X_i and the estimated deviation will be neglected as higher order terms.

$$\begin{aligned} E(AIC_k) &= 2n - 2nE[\ln \tilde{f}(X_i)] \\ &= 2n + 2n \ln(nh\sqrt{2\pi}) + 2nE\left[\frac{(X_j - X_i)^2}{2h^2}\right]. \end{aligned} \quad (11)$$

$$E(AIC_k) = 2n + 2n \ln(nh\sqrt{2\pi}) + 2n \frac{\sigma^2}{h^2}. \quad (12)$$

$$\begin{aligned} E(AIC_k) &= 2n + 2n \ln(\sqrt{2\pi}) + 2n \ln\left(n^{4/5} \left(\frac{4}{3}\right)^{1/5} \sigma\right) \\ &\quad + 2n \left(\frac{4}{3n}\right)^{2/5}. \end{aligned} \quad (13)$$

$$\begin{aligned} E(AIC_k) &= 2n \left(1 + \ln(\sqrt{2\pi}) + \frac{4}{5} \ln n + \frac{1}{5} \ln \frac{4}{3} \right. \\ &\quad \left. + \ln \sigma + \left(\frac{4}{3n}\right)^{2/5}\right). \end{aligned} \quad (14)$$

$$\begin{aligned} E(AIC_k) &= 2n \left(1 + \ln(\sqrt{2\pi}) + \frac{1}{5} \ln \frac{4}{3} + \ln \sigma \right. \\ &\quad \left. + \frac{4}{5} \ln n + \left(\frac{4}{3n}\right)^{2/5}\right). \end{aligned} \quad (15)$$

$$E(AIC_k) = O(n \ln n). \quad (16)$$

2.4. Comparison

The expected AIC_k does not depend on any parameters of the underlying distribution function that we assumed to be the normal one. It can be seen that it grows with $n \ln(n)$, i.e. faster than linearly in n . Comparing this fact with the AIC for the normal model, we can see that the AIC of the kernel estimator grows faster in n than the AIC for the parametric model. The AIC shows a relative information loss and the most preferable model is always that with the smaller AIC. So we can see that with growing sample size the kernel estimator will perform worse with growing sample size, assuming the underlying model is the normal distribution.

On the one hand side this result seems to be trivial, a parametric model will always perform better than a non-parametric one – provided it is correct. On the other hand side we see that the difference in information loss for the non-parametric approach is only of the order $n^{1/5}$.

Figure 1 shows then difference of $E(AIC_k) - E(AIC_N)$. It can be seen that the difference grows with increasing n , but decreases with increasing variance of the population.

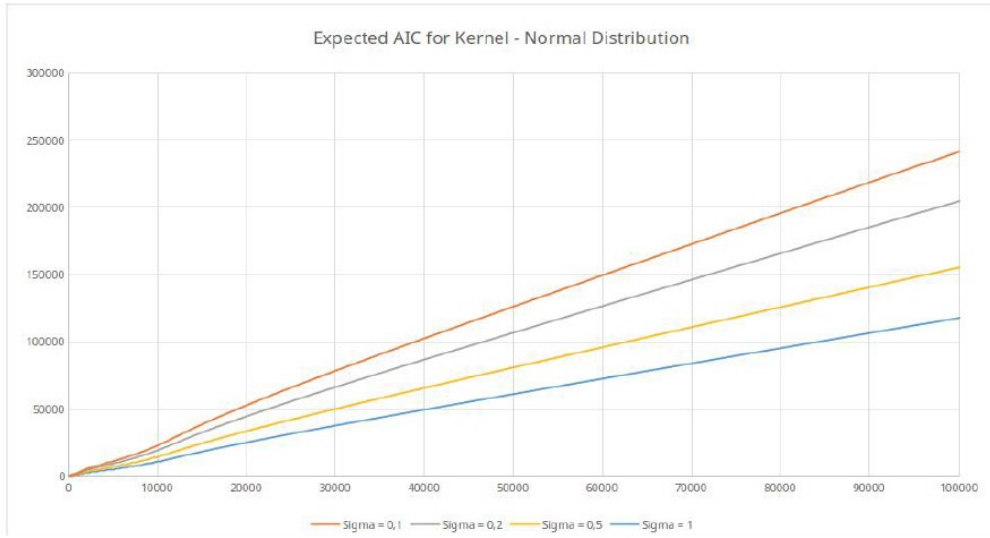


Figure 1: Comparison of the expected AIC values for the kernel estimator versus the normal model

3. A GENERALISATION

In this chapter we generalise the considerations from the previous chapter. Assume again that we have a sample $X_1, \dots, X_n$ from a distribution with density $f(x | \theta)$, where θ is a k_1 dimensional vector of parameters. The values can be multidimensional and the only restriction is, that the coinciding sigma algebra is measurable and that the density exists and fulfils the necessary regularity conditions for the existence and convergence of the maximum likelihood estimator.

Assume now that the second model is the AI model, having density $g(x | \eta)$ with parameter vector η , having k_2 dimensions and in most cases $k_2 \gg k_1$.

Now we have

$$AIC_f = k_1 - \sum_{i=1}^n \ln f(X_i | \hat{\theta}), \quad (17)$$

$$AIC_{AI} = k_2 - \sum_{i=1}^n \ln g(X_i | \hat{\eta}). \quad (18)$$

Computing the expectations we arrive at

$$E(AIC_f) = k_1 - \int_{-\infty}^{\infty} \ln f(x) f(x) dx, \quad (19)$$

$$E(AIC_{AI}) = k_2 - \int_{-\infty}^{\infty} \ln g(x) f(x) dx. \quad (20)$$

So, the difference of the expectations turns out to be

$$\begin{aligned} E(AIC_{AI}) - E(AIC_f) &= k_2 - k_1 - \int_{-\infty}^{\infty} \ln g(x) f(x) dx \\ &\quad + \int_{-\infty}^{\infty} \ln f(x) f(x) dx. \end{aligned} \quad (21)$$

The second term is just the Kullback–Leibler distance

$$D(f \parallel g) = \int_{-\infty}^{\infty} \ln \left(\frac{f(x)}{g(x)} \right) f(x) dx, \quad (22)$$

i.e.

$$E(AIC_{AI}) - E(AIC_f) = k_2 - k_1 + D(f \parallel g). \quad (23)$$

This expression is positive, if the parametric model outperforms the AI model. Now we have $k_2 > k_1$ and $D(f \parallel g) \geq 0$ for $f(x)$ and $g(x)$ being different. This shows, that using an AI model in place of the genuine parametric model leads to information loss measured by the Akaike criterion caused first by the increased number of parameters and second by using another model. This other model (here described by the density $g(x)$) might be very flexible, but is not identical to the real model $f(x)$.

4. CONCLUSION

Then result of the application of the Akaike information criterion shows, that application of an AI model always leads to information loss, compared with the use of the real correct model.

So, AI models might help for a first guess or if the real underlying model is hard to find. Nevertheless, one must always keep in mind that the use of an AI model will not use the entire information. This is also caused by the fact, as pointed out by Schäbe and Shubinsky [1], that an AI model will not use a set of sufficient statistics, which would distillate all information from the sample. The question is then, how to apply an AI model and a good first step would be to ask it, in assisting for finding the underlying model, then testing for this model and then to apply the real parametric model.

REFERENCES

- [1] Schäbe, H., and Shubinsky, I. B. (2024). How to properly apply systems of artificial intelligence. *Reliability: Theory & Applications*. 19(4):31–35. <https://doi.org/10.24412/1932-2321-2024-480-31-35>
- [2] Vapnik, V. N., and Chervonenkis, A. Y. (1974). Pattern recognition theory: Statistical problems of learning. Nauka. Moscow
- [3] Vapnik, V. N. (2010). The nature of statistical learning theory. Springer. <https://doi.org/10.1007/978-1-4757-3264-1>
- [4] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19(6):716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- [5] Mann, N. R., et al. (1974). Methods for statistical analysis of reliability and life data. John Wiley & Sons
- [6] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*. 27(3):832–837. <https://doi.org/10.1214/aoms/1177728190>
- [7] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*. 33(3):1065–1076. <https://doi.org/10.1214/aoms/1177704472>
- [8] Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman & Hall/CRC. London. <https://doi.org/10.1201/9781315140919>
- [9] Schäbe, H. (1988). Zur bestimmung des glättungsparameters für kernschätzungen mit zensierten stichproben. *Wiss. Z. HfV, Dresden*. 25(3):573–591