SCOPUA
Journal of Applied Statistical Research
Editor in Chief | Dr. Muhammad Aslam

# SCOPUA Journal of Applied Statistical Research

**Volume 2 (Issue 2) Published on 2026-06-09**

https://doi.org/10.64060/JASRV2i2

| | |
|---|---|
| ISSN (e) | 3104-4794 |
| DOI | 10.64060/JASR |
| Frequency | Quarterly |
| Editor in Chief | Prof. Dr. Muhammad Aslam |
| Contact | jasr@journals.scopua.com |

**Journal Homepage**

https://journals.scopua.com/index.php/JASR

**Publisher**

Scientific Collaborative Online Publishing Universal Academy

## Aims & Scope

SCOPUA Journal of Applied Statistical Research (JASR), ISSN 3104-4794, aims to serve as a comprehensive, peer-reviewed, open-access platform dedicated to advancing knowledge and fostering innovation in the field of applied statistical science. The journal publishes high-quality, original research articles, reviews, and case studies that explore the development and application of statistical methodologies to solve complex real-world problems. Targeting a global audience of statisticians, researchers, and data analysts, the journal seeks to be a leading forum for the dissemination of cutting-edge statistical knowledge and to support the growing demand for advanced statistical applications in various sectors.

Its scope covers a wide range of topics, including statistical theory and methods, biostatistics, statistical computing, data analytics, neutrosophic statistics, and statistical quality control. JASR is particularly focused on interdisciplinary research that bridges statistics with fields such as engineering, social sciences, business, environmental science, and computer science, promoting the integration of statistical tools in diverse areas. The journal welcomes submissions that address contemporary challenges in big data analytics, artificial intelligence, uncertainty modelling, and decision science, with special attention to innovative applications that have tangible societal or industrial impact.

# Contents (Volume 2 & Issue 2 – 2026)

SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 2, June 2026
DOI: https://doi.org/10.64060/JASR2V2i1


# Socioeconomic and Demographic Study of the Determinants of Anaemia Prevalence Among Women in Nigeria

Onyenekwe Chukwuenyem E [1*], Oladuti Mercy. O [2], Adubiaro Dunsin [2], Onyenekwe Amala Joy [3]

[1]Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

[2]Department of Statistics, Federal University of Technology, Akure, Ondo State, Nigeria

[3]Department of Environmental Health Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria

* Corresponding Email: ce.onyenekwe@unizik.edu.ng

Received: 27 February 2026 / Revised: 05 April 2026 / Accepted: 11 April 2026 / Published online: 15 April 2026



ABSTRACT

Anaemia remains a significant public health issue among women of reproductive age, especially in developing countries like Nigeria. This study examines the influence of socio-economic and demographic factors on anaemia among women in Nigeria using data from the 2018 Nigeria Demographic and Health Survey (NDHS). Variables analyzed include age, marital status, body mass index (BMI), education level, wealth index, pregnancy status, contraceptive use, region, place of residence, ethnicity, media exposure, and access to improved water. We first used descriptive statistics and Pearson's chi-square tests to explore links between women's anaemia status and key factors like residence, education, and wealth. Variables showing significant associations at this stage were then examined further with a partial proportional odds model, chosen to properly handle the ordered categories of anaemia severity (non-anaemic, mild, moderate, severe).Among the women surveyed, 41% had no anaemia, while 27% showed mild anaemia, 30% moderate anaemia, and 2% severe anaemia. Rural living, lower education evels, poorer household wealth, current pregnancy, lack of improved water access, and certain regional differences all raised the odds of severe anaemia significantly. The study concludes that socio-economic inequalities play a significant role in the burden of anaemia among women in Nigeria. Targeted interventions focusing on education, economic empowerment, and improved healthcare access are essential to reduce anaemia prevalence.

**Keywords**: Anaemia; Socio-economic Factors; Nigeria; NDHS; Ordinal Regression

## 1. Introduction

Anaemia is a state in which one's body is short of red blood cells or the capacity to transport oxygen from the lungs to the other parts of the body. The causes of anaemia include, but are not limited to, nutritional deficiencies, genetic conditions, illnesses, parasites, and menstruation. Research has it that anaemia affects women more than men all over the world, including Nigeria. From the findings of Gardner and Kassebaum (2020), anaemia is a major public health problem in Nigeria, where it was estimated that 48.2% of women aged 15- 49 years are anemic.

More than 500 million women of childbearing age around the world have anaemia. Nigeria has one of the highest rates in sub-Saharan Africa, where more than 50% of women have it. This ongoing problem harms the health of mothers, raises the chances of bad pregnancy outcomes, and keeps the cycle of poverty and stunted child development going from one generation to the next. It is a major public health issue. Investigating the socio-economic and demographic determinants of anaemia in Nigeria is pertinent for policy formulation, as it guides targeted interventions within the National Anaemia Reduction Policy, addresses the deficiency in ordinal analyses utilising recent NDHS data, and rectifies the absence of partial proportional odds modelling to reflect the differential impacts across varying levels of anaemia severity.

Anaemia is prevalent in Nigerian women of childbearing age, and it is related to morbidity and mortality because it has adverse effects on pregnancy outcomes and children's learning abilities (Karami et al., 2022). Anaemia is a significant health problem, prevalent among women in Nigeria and it has social and economic impacts on them. A cross-sectional

study conducted by Keokenchanh et al. (2021) revealed that anaemia is highly prevalent among women in the reproductive age in Nigeria, with a prevalence of 42% nationally, while the highest prevalence was reported in the Northern region of the country. The high prevalence of anaemia can be due to the following reasons including low diet, poor health facilities and low economic status. Also, according to Jasim et al. (2020), anaemia affects women in pregnancy and childbirth since it acts as a risk factor for both domains. Anaemia during pregnancy leads to maternal morbidity and mortality, preterm birth, and low birth weight; therefore, it is essential to address it (Majeed et al., 2022).Effiong and Udom (2025) study looked at pregnant women visiting antenatal clinics in Calabar. IT checked how common anaemia and malaria were, and how often they occurred together. They found that both problems are still very widespread there, showing malaria's heavy toll on anaemia in Nigeria's high-risk areas. This points to the importance of tackling connected causes—like parasites, poor nutrition, and social factors—in strategies to fight anaemia among women of childbearing age.

## 2. Methodology

This study used secondary data from the 2018 Nigeria Demographic and Health Survey (NDHS). The Partial Proportional Odds Model (PPOM) was selected because anaemia severity is ordinal (non-anaemic, mild, moderate, severe), but the proportional odds assumption of the standard ordered logit model was violated by key variables. This assumption requires the effect of each predictor to be consistent across all cumulative odds splits; violations occur when relationships differ by severity threshold. Variables such as rural residence, education level, wealth index, pregnancy status, and improved water access violated this assumption, as confirmed by Brant tests ($p < 0.05$ for these predictors), necessitating PPOM to relax proportionality for those while maintaining it for others. Data were analyzed using descriptive statistics, Pearson's chi-square tests for associations, and PPOM via the 'vgam' package in R

### *2.1 Variable Description*

2.1.1 Dependent variable

The respondent's anaemia level is the dependent variable. Women with a hemoglobin level of less than 12 g/dL were considered anemic. The anaemia level of respondents was recoded into an ordinal variable as mild, moderate, severe, and non-anemic. According to Mayang S. et al. (2001), anaemia was categorized as mild (hemoglobin (Hb) of 10.0–11.9 g/dL), moderate (Hb of 7.0–9.9 g/dL), and severe (Hb < 7.0 g/dL).

2.1.2 Independent variables

The variables were grouped into demographic and socio-economic factors based on the conceptual framework for anaemia determinants. The demographic characteristics included the age of the respondent, marital status (never in a union, married/living with a partner, and divorced/separated/widowed), body mass index (BMI) (underweight, healthy weight, and overweight), modern contraceptive use (yes and no), region (North-Central, North-East, North-West, South-East, South-South, and South-West), place of residence (urban or rural), and ethnicity (Hausa, Yoruba, Igbo, and others). The total children ever born (0, 1, 2–4, and ⩾5) were regrouped into four categories of parity (nulliparity, primiparity, multiparity, and grand multiparity), correspondingly (Daniel Chukwuemeka et. al. 2022). BMI was converted from a numeric to a categorical variable based on the World Health Organization (WHO). BMI less than 18.5, between 18.5 and 25, and those greater than 25 were regrouped into Underweight, healthy weight, and overweight, respectively. Also, the age of the respondent was converted into five (5) categorical-based variables namely: 15 – 19, 20 – 24, 25 – 29, 30 – 34, and ≥ 40.

The socio-economic included are highest education (no education, primary, secondary, and higher), wealth index (poorest, poor, moderate, rich, richest), the main source of drinking water (unimproved and improved) and media exposure (yes and no).

### *2.2 Bivariate Analysis*

2.2.1 Chi-square ($\chi^2$)

A Pearson's Chi-square ($\chi^2$) test for independence is a non-parametric statistical technique for testing whether two categorical variables are related or not. The test compares the observed frequencies of the variables in a sample with the expected frequencies under the null hypothesis, which is usually that the variables are independent. The test statistic of is given by;

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (1)$$

where O is the observed frequency and E is the expected frequency

### *2.3 Statistical Models*

2.3.1 Ordinal Logistic Regression

Ordinal logistic regression or simply ordinal regression is a statistical modelling technique for testing hypotheses about the nature of the relationship between one or more independent variables and one ordinal dependent (response or outcome) variable (Long, J. S., and J. Freese. 2006). It is an extension of binary logistic regression and is suitable for modelling the relationship between independent variables and an ordinal outcome variable with more than two categories.

In this study, the technique is used to capture the ordinal nature of the dependent variables (non-anemic, mild, moderate, and severe anaemia). The model assumes that;

- The dependent variable is measured on an ordinal level.
- The independent variable(s) can be either continuous, categorical, or ordinal.
- No high correlation between two or more independent variables (multi-collinearity)
- Each independent variable has the same effect at each cumulative split of the ordinal dependent variable (Proportional Odds). If the relationship between all pairs of groups is consistent, only one set of coefficients is needed, meaning there is just one model. However, if this assumption is violated, different models are required to describe the relationship between each pair of outcome groups. The ordinal logistic model is also called the proportional odds model (POM).

Given an ordinal response variable Y with J categories, let X be a vector of explanatory variables with p elements, suppose $P(Y \leq j \mid X)$ is the cumulative probability of the response variable being in category j or lower, given the explanatory variables and $P(Y > j \mid X)$ is the complementary probability of the response variable being in category j+1 or higher, given the explanatory variables the model is expressed as;

$$\log\left(\frac{P(Y\leq j|X)}{P(Y>j|X)}\right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \quad (2)$$

Where, $\alpha_j$ is the intercept for the jth category, j=1..., J−1, and $\beta_i$ is the coefficient for the ith explanatory variable, i=1..., p.

2.3.2 Generalised Ordered Logistic Regression

The general ordered logistic model extends the ordered logistic regression model. It is used for ordinal response variables when the proportional odds assumption may not apply. This model allows different thresholds between categories (Richard Williams, 2006).

For this study, the partial proportional odds model was employed because the assumption of proportional odds holds for some predictors.

2.3.3 Partial Proportional Odds Model (PPOM)

This model relaxes the proportional odds assumption for some predictions while maintaining it for others. This is useful when some predictors meet the proportion odds assumption, but others do not. The PPOM permits nonproportional odds for a subset of q of the p-predictors (q ≤ p). We may define the PPOM cumulative probability as follows:

$$\log\left(\frac{P(Y\leq j|X)}{P(Y>j|X)}\right) = \alpha_j + x_i^{'}\beta + t'\gamma_j \quad (3)$$

Where $t'$is a vector $q$ covariates $(1by\ q)$ that contain the values of observation i on that subset of the $p$ independent variables for which the assumption of proportionality is either not met or is to be tested, and $\gamma_j$is the vector of the regression coefficient with the dimension $(1by\ q)$, which is associated with the $q$ covariates. As a result, $t'\gamma_j$ is the increase that is associated with the $jth$ cumulative logit $(1 \leq j \geq j-1)$, where $\gamma_1 = 0$. If the values of $\gamma_i = 0$ for all $j$, Then this model reduces to POM. (Richard Williams, 2006)

***2.4 Model Evaluation***

This section involves assessing how well a model fits the data and how accurately it predicts outcomes.

2.4.1 Test for Multicollinearity

Multicollinearity happens when two or more variables are highly correlated. Testing for multicollinearity is also important when fitting POM models, just like in other regression models. For this study, the variance inflation factor is used to assess multicollinearity.

2.4.2 Variance Inflation Factor (VIF)

Variance inflation factor (VIF) is a measure of multicollinearity in a regression analysis. It indicates the degree to which correlations among predictors inflate variance. It is expressed as,

$$VIF = \frac{1}{1-R_i^2} \quad (4)$$

Where $\boldsymbol{R_i^2}$ represents the unadjusted coefficient of determination for regressing the ith independent variable on the remaining ones.

The general rule of thumb for the VIF test is that if the VIF value is greater than 10, then there is multicollinearity.

2.4.3 Test for Proportional ODDS

This section explains the test used to evaluate the proportional odds assumption. Testing for proportional odds is important because the model's validity and interpretability depend greatly on this assumption.

The null hypothesis for the testing of the proportional odds assumption is given by

$$H_0: \beta_1 = \beta_2 = \cdots \beta_{J-1}$$

The null hypothesis is rejected when $p - value\ < 0.05$.

2.4.4 Wald Test

The Wald test is a statistical test used to assess the significance of individual coefficients in a regression model. It evaluates whether the estimated coefficient of a predictable variable is different from zero and tests the parallel lines assumption.

Given the likelihood function of the cumulative of the ordinal logistic model as;

$$L(\theta; x, y) = \prod_{i=1}^{n} \prod_{j=1}^{K} (\gamma_{ij} - \gamma_{ij-1})^{I(y_i=j)} \quad (5)$$

Where $\theta^T = (\alpha_1, \beta_1^T, \dots, \alpha_{j-1}, \beta_{j-1}^T)$ includes all parameters.

Based on the asymptotic distribution of $\hat{\theta}$, the null hypothesis for the Wald test can be written in a matrix form

$$L * \left(\beta_1^T, \beta_2^T, \dots \beta_{j-1}^T\right)^T = 0 \quad (6)$$

where $L = \begin{pmatrix} I_{s\times s} & 0_{s\times s} & \cdots & 0_{s\times s} & -I_{s\times s} \\ 0_{s\times s} & I_{s\times s} & \cdots & 0_{s\times s} & -I_{s\times s} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0_{s\times s} & 0_{s\times s} & \dots & I_{s\times s} & -I_{s\times s} \end{pmatrix}$ is an $(J-2)\times(J-1)$ block matrix, $I_{s\times s}$ is the $s \times s$ identity matrix, and $0_{s\times s}$ is the $s \times s$ with all entries 0.

Let V be the estimated asymptotic variance of $(\beta_1^T, \beta_2^T, \dots \beta_{j-1}^T)$, then the asymptotic variance of $L * \left(\beta_1^T, \beta_2^T, \dots \beta_{j-1}^T\right)^T$ is $L * V * L^T$.

Hence, Wald's statistics can be defined as

$$s_{Wald} = (\beta_1^T, \beta_2^T, \dots \beta_{j-1}^T) L^T (L * V * L^T)^{-1} L * \left(\beta_1^T, \beta_2^T, \dots \beta_{j-1}^T\right)^T \quad (7)$$

The Wald statistics follow a $\chi^2$ distribution with $(J-2)$ degrees of freedom.

## 3. Result

### *3.1 Characteristics of respondents*

Table 1 provides a summary of the characteristics of the study participants in 2018. The table details respondents' attributes, such as age, gender, education level, and region. The age distribution indicates that the 25-29 years group is the most prevalent, making up 28.26% of all respondents. Conversely, the 15-19 years group was the least represented, comprising only 3.99%. The wealth index distribution shows respondents are fairly evenly spread across different economic levels. The middle wealth group is the largest at 22.03%, followed by the poorest at 21.29%, the poorer at 20.71%, the richer at 20.21%, and the richest at 15.76%. Regarding marital status, most respondents are married or living with a partner (94.68%), with widowed individuals accounting for 3.31%. About 47.24% have had more than five children, while only around 0.80% have not given birth yet. Furthermore, 41.20% of respondents have a healthy weight. However, more than half are underweight (57.02%), with overweight individuals being the least common at 1.75%. Education levels vary widely; the highest percentage (40.19%) have no formal education, followed by those with secondary education (34.56%). Approximately 62% live in rural areas, and 59.40% have access to media. Although 25.91% of respondents are from Nigeria's northwest, 16.71% are Igbo, and about 12.36% are Yoruba. The proportions of respondents who are currently pregnant, breastfeeding, and using contraceptives are 13.09%, 52.09%, and 18.92%, respectively. Additionally, around 66% have access to improved water sources.

**Table 1**: Summary of respondent characteristics

| Characteristics | | frequency | Percentage (%) |
|---|---|---|---|
| Age group | 15 – 19 | 511 | 3.99 |
| | 20 – 24 | 2,273 | 17.73 |
| | 25 – 29 | 3,623 | 28.26 |
| | 30 - 34 | 2,975 | 23.21 |
| | 35 – 39 | 2,203 | 17.19 |
| | ≥ 40 | 1,234 | 9.63 |
| Current marital status | Never in union | 258 | 2.01 |
| | Married/living with partner | 12,137 | 94.68 |
| | widowed/divorced/separated | 424 | 3.31 |
| Body mass index | Underweight | 7,310 | 57.02 |
| | Healthy weight | 5,281 | 41.20 |
| | Overweight | 224 | 1.75 |
| Parity | Nulliparity | 102 | 0.80 |
| | Primiparity | 1,771 | 13.82 |
| | Multiparity | 4,890 | 38.15 |
| | Grand Multiparity | 6,056 | 47.24 |
| Religion | Christianity | 5,773 | 45.03 |
| | Islamic | 6,998 | 54.59 |
| | Others | 48 | 0.37 |
| Place of residence | Urban | 4,819 | 37.59 |

| | | | |
|---|---|---|---|
| | Rural | 8,000 | 62.41 |
| Wealth index | Poorest | 2,729 | 21.29 |
| | Poorer | 2,655 | 20.71 |
| | Middle | 2,824 | 22.03 |
| | Richer | 2,591 | 20.21 |
| | Richest | 2,020 | 15.76 |
| Education level | No education | 5,152 | 40.19 |
| | Primary | 2,134 | 16.65 |
| | Secondary | 4,430 | 34.56 |
| | Higher | 1,103 | 8.60 |
| Ever terminated a pregnancy | No | 11,030 | 86.04 |
| | Yes | 1,789 | 13.96 |
| Currently pregnant | No | 11,141 | 86.91 |
| | Yes | 1,678 | 13.09 |
| Currently Breastfeeding | No | 6,142 | 47.91 |
| | Yes | 6,677 | 52.09 |
| Current use of contraceptive | No | 10,394 | 81.08 |
| | Yes | 2,425 | 18.92 |
| Access to media | No | 5,204 | 40.60 |
| | Yes | 7,615 | 59.40 |
| Water source | Unimproved | 4,399 | 34.32 |
| | Improved | 8,420 | 65.68 |
| Ethnicity | Hausa | 4,955 | 38.65 |
| | Igbo | 2,142 | 16.71 |
| | Yoruba | 1,584 | 12.36 |
| | Others | 4,138 | 32.28 |
| Region | North-central | 2,230 | 17.40 |
| | North-east | 2,399 | 18.71 |
| | North-west | 3,321 | 25.91 |
| | South-east | 1,723 | 13.44 |
| | South-south | 1,337 | 10.43 |
| | South-west | 1,809 | 14.11 |

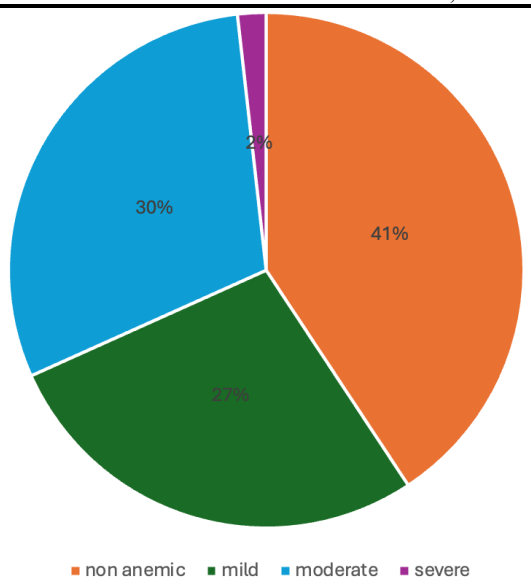


**Figure 1**: Severity of anaemia among women in Nigeria

The anaemia prevalence among 12,819 women in Nigeria in 2018 is shown in Figure 1. The figure displays the distribution of anaemia severity among women. Non-anemic women made up about 41% of the population. While 30% had moderate anaemia, 27% had mild anaemia; the severe category (2%) had the lowest percentage. This indicates that approximately 59% of the total population was anemic.

### *3.2 Prevalence of Severity of Anaemia Among Women in Nigeria*

The prevalence of non-, mild, moderate, and severe anaemia among women, as well as the relationship between anaemia levels and various characteristics, are shown in Table 4. Across age groups, the highest prevalence of severe anaemia was observed in the 15-19 years group (1.96%), while the lowest was in the 25-29 years group (1.32%). Additionally, respondents who are underweight have a higher prevalence of being non-anemic (43.31%) compared to those with a healthy weight (37.28%) and overweight individuals (36.40%). Severe anaemia is most common among those with a

healthy weight (2.39%). Furthermore, women with no education have the highest prevalence of moderate (33.99%) and severe (2.97%) anaemia, while women with higher education have the lowest prevalence of moderate (51.95%) and severe (1.27%) anaemia.

However, the chi-square test indicates there is a significant association between the severity of anaemia with women's age, body mass index, parity, place of residence, wealth index, education level, pregnancy status, use of contraceptives, access to media, water source, ethnicity, and religion (p-value < 0.01). In contrast, marital status, breastfeeding, and ever-terminated pregnancy were not significantly associated with the severity of anaemia among women.

**Table 2**: Prevalence of severity of anaemia among women

| | | Not anemic | Mild | Moderate | Severe | $\chi^2$ | P-value |
|---|---|---|---|---|---|---|---|
| | | n (%) | n (%) | n (%) | n (%) | | |
| **Characteristics** | | 15 - 19 | 177(34.64) | 152(29.75) | 172(33.66) | 10(1.96) | 33.34 |
| Age group | 20 - 24 | 906(39.86) | 608(26.75) | 721(31.72) | 38(1.67) | | 0.004 |
| | 25 – 29 | 1,535(42.37) | 985(27.19) | 1,055(29.12) | 48(1.32) | | |
| | 30 - 34 | 1,168(39.26) | 849(28.54) | 888(29.85) | 70(2.35) | | |
| | 35 – 39 | 891(40.44) | 625(28.37) | 647(29.37) | 40(1.82) | | |
| | ≥ 40 | 541(43.84) | 316(25.61) | 354(28.69) | 23(1.86) | | |
| Current marital status | Never in union | 104(40.31) | 74(28.68) | 75(29.07) | 5(1.94) | 7.75 | |
| | Married/living with partner | 4,947(40.76) | 3,358(27.67) | 3,621(29.83) | 211(1.74) | | 0.26 |
| | Widowed /divorced/ separated | 167(39.39) | 103(24.29) | 141(33.25) | 13(3.07) | | |
| Body mass index | Underweight | 3,166(43.31) | 2,005(27.43) | 2,039(27.89) | 100(1.37) | 73.44 | |
| | Healthy weight | 1,969(37.28) | 1,460(27.65) | 1,726(32.68) | 126(2.39) | | 0.00 |
| | Overweight | 83(36.40) | 70(30.70) | 72(31.58) | 3(1.32) | | |
| Parity | Nulliparity | 48(47.06) | 21(20.59) | 29(28.43) | 4(3.92) | 22.01 | |
| | Primiparity | 718(40.54) | 489(27.61) | 538(30.38) | 26(1.47) | | 0.01 |
| | Multiparity | 2,069(42.31) | 1,310(26.79) | 1,440(29.45) | 71(1.45) | | |
| | Grand Multiparity | 2,383(39.35) | 1,715(28.32) | 1,830(30.22) | 128(2.11) | | |
| Place of residence | Urban | 2,210(45.86) | 1,277(26.50) | 1,261(26.17) | 71(1.47) | 94.46 | |
| | Rural | 3,008(37.60) | 2,258(28.23) | 2,576(32.20) | 158(1.98) | | 0.00 |
| Wealth index | Poorest | 921(33.75) | 780(28.58) | 972(35.62) | 56(2.05) | 202.77 | |
| | Poorer | 1,016(38.27) | 733(27.61) | 854(32.17) | 52(1.96) | | 0.00 |
| | Middle | 1,113(39.41) | 821(29.07) | 842(29.82) | 48(1.70) | | |
| | Richer | 1,111(42.88) | 707(27.29) | 726(28.02) | 47(1.81) | | |
| | Richest | 1,057(52.33) | 494(24.46) | 443(21.93) | 26(1.29) | | |
| Education level | No education | 1,851(35.93) | 1,438(27.91) | 1,751(33.99) | 112(2.17) | 151.47 | |
| | Primary | 867(40.63) | 605(28.35) | 621(29.10) | 41(1.92) | | 0.00 |
| | Secondary | 1,927(43.50) | 1,200(27.09) | 1,241(28.01) | 62(1.40) | | |
| | Higher | 573(51.95) | 292(26.47) | 224(20.31) | 14(1.27) | | |
| Ever terminated a pregnancy | No | 4,494(40.74) | 3,050(27.65) | 3,277(29.71) | 209(1.89) | 6.70 | |
| | Yes | 724(40.47) | 485(27.11) | 560(31.30) | 20(1.12) | | 0.08 |
| Currently pregnant | No | 4,627(41.53) | 3,052(27.39) | 3,265(29.31) | 197(1.77) | 26.59 | |
| | Yes | 591(35.22) | 483(28.78) | 572(34.09) | 32(1.91) | | 0.00 |
| Currently Breastfeeding | No | 2,482(40.41) | 1,682(1,682) | 1,877(30.56) | 101(1.64) | 3.29 | |
| | Yes | 2,736(40.98) | 1,853(27.75) | 1,960(29.35) | 128(1.92) | | 0.35 |
| Current use of contraceptive | No | 4,042(38.89) | 2,868(27.59) | 3,274(31.50) | 210(2.02) | 106.49 | |
| | Yes | 1,176(48.49) | 667(27.51) | 563(23.22) | 19(0.78) | | 0.00 |
| Access to media | No | 1,989(38.22) | 1,497(28.77) | 1,606(30.86) | 112(2.15) | 26.87 | |
| | Yes | 3,229(42.40) | 2,038(26.76) | 2,231(29.30) | 117(1.54) | | 0.00 |
| Water source | Unimproved | 1,713(38.94) | 1,213(27.57) | 1,410(32.05) | 63(1.43) | 19.89 | |
| | Improved | 3,505(41.63) | 2,322(27.58) | 2,427(28.82) | 166(1.97) | | 0.00 |
| Ethnicity | Hausa | 1,862(37.58) | 1,357(27.39) | 1,622(32.73) | 114(2.30) | 136.76 | |
| | Igbo | 768(35.85) | 606(28.29) | 726(33.89) | 42(1.96) | | 0.00 |
| | Yoruba | 763(48.17) | 445(28.09) | 359(22.66) | 17(1.07) | | |
| | Others | 1,825(44.10) | 1,127(27.24) | 1,130(27.31) | 56(1.35) | | |
| Region | North-central | 967(43.36) | 607(27.22) | 635(28.48) | 21(0.94) | 149.41 | |
| | North-east | 943(39.31) | 693(28.89) | 720(30.01) | 43(1.79) | | 0.00 |
| | North-west | 1,339(40.32) | 871(26.23) | 1,031(31.04) | 80(2.41) | | |
| | South-east | 579(33.60) | 499(28.96) | 608(35.29) | 37(2.15) | | |
| | South-south | 517(38.67) | 343(25.65) | 449(33.58) | 28(2.09) | | |
| | South-west | 873(48.26) | 522(28.86) | 394(21.78) | 20(1.11) | | |

| South-west |
|---|

### *3.3 Predictors of Severity Levels of Anaemia Among Nigerian Women*

The proportional odds model was considered to determine the factors influencing the severity of anaemia among women in Nigeria. The model assumes that there is no multicollinearity and the correlation between the dependent and independent variables does not change for the categories of the dependent variable (parallel line assumption). The results from the test for multicollinearity and parallel assumptions are presented in Tables 3 and 4.

Table 3 shows the result of the multicollinearity assessment of the explanatory variables. The table shows the variance inflation factor (VIF) and tolerance, which measure how much the variance of an estimated regression coefficient increases due to collinearity. The VIFs of all the explanatory variables are below 3, indicating no multicollinearity. Also, the tolerance for the explanatory variables was relatively high, indicating no multicollinearity.

**Table 3**: Multicollinearity assessment of the explanatory variables

| Variable | VIF | Tolerance | R-Squared |
|---|---|---|---|
| Age group | 1.69 | 0.5932 | 0.4068 |
| Body mass index | 1.18 | 0.8443 | 0.1557 |
| Marital status | 1.02 | 0.9814 | 0.0186 |
| Parity | 1.66 | 0.6039 | 0.3961 |
| Place of residence | 1.43 | 0.7009 | 0.2991 |
| Wealth index | 2.52 | 0.3976 | 0.6024 |
| Education Level | 2.10 | 0.4771 | 0.5229 |
| Ever terminated a pregnancy | 1.02 | 0.9850 | 0.0150 |
| Currently pregnant | 1.25 | 0.8019 | 0.1981 |
| Currently breastfeeding | 1.27 | 0.7898 | 0.2102 |
| Use of modern Contraceptive | 1.17 | 0.8551 | 0.1449 |
| Region | 1.22 | 0.8189 | 0.1811 |
| Access to media | 1.22 | 0.8166 | 0.1834 |
| Water source | 1.27 | 0.7846 | 0.2154 |
| Race/Ethnicity | 1.20 | 0.8330 | 0.1670 |
| Mean VIF | 1.44 | | |

**Table 4:** Result for the test of parallel line assumption

| Independent variable | p>Chi-square |
|---|---|
| Age group | |
| 15-19 (Ref) | 0 |
| 20-24 | 0.238 |
| 25-29 | 0.247 |
| 30-34 | 0.642 |
| 35-39 | 0.645 |
| >=40 | 0.187 |
| | |
| Current marital status | |
| Never in union (Ref) | 0 |
| Married/living with partner | 0.685 |
| widowed/divorced/separated | 0.129 |
| Body mass group | |
| Underweight (Ref) | 0 |
| Healthy weight | 0.876 |
| Overweight | 0.034 |
| Parity | |
| Nulliparity (Ref) | 0 |
| Primiparity | 0.021 |
| Multiparity | 0.027 |
| Grand Multiparity | 0.044 |
| Place of residence | |
| Urban (Ref) | 0 |
| Rural | 0.421 |
| Wealth index | |
| Poorest (Ref) | 0 |
| Poorer | 0.392 |
| Middle | 0.254 |
| Richer | 0.073 |
| Richest | 0.135 |
| Education level | |
| No education (Ref) | 0 |
| Primary | 0.422 |
| Secondary | 0.868 |
| Tertiary | 0.308 |

| | |
|---|---|
| Ever terminated pregnancy | |
| No (Ref) | 0 |
| Yes | 0.035 |
| Currently pregnant | |
| No (Ref) | 0 |
| Yes | 0.449 |
| Currently Breastfeeding | |
| No (Ref) | 0 |
| Yes | 0.307 |
| Region | |
| North-central (Ref) | 0 |
| North-east | 0.180 |
| North-west | 0.000 |
| South-east | 0.186 |
| South-south | 0.006 |
| South-west | 0.069 |
| Access to media | |
| No (Ref) | 0 |
| Yes | 0.016 |
| Water source | |
| Unimproved (Ref) | 0 |
| Improved | 0.000 |
| Ethnicity | |
| Hausa (Ref) | 0 |
| Igbo | 0.608 |
| Yoruba | 0.624 |
| Others | 0.985 |
| Overall | $\chi^2$= 130.2<br>p-value = 0.000 |

Table 4 presents the result of Wald tests of parallel line assumption. The result indicates a significant result ($\chi^2$= 130.2, p-value < 0.05) for POM, suggesting that the parallel regression assumption is violated for the model. The violation of this assumption can lead to misinterpretation of the model coefficient. The results also show that 8 (Age group, marital status, place of residence, educational level, currently pregnant, currently breastfeeding, and ethnicity) out of the 14 variables meet the parallel lines assumption. Therefore, the partial proportional odds model (PPOM), which is a generalized ordinal regression that relaxes the parallel line assumption, was used. An insignificant test statistic ($\chi^2$ = 53.67, p-value = 0.2661) for the Wald test for the PPOM indicates that the model does not violate the parallel lines assumption.

Table 5 presents the result of the fitted partial proportional odds model (PPOM). From the result, the women aged 40 and above tend to have significantly lower odds (OR = 0.755, p-value < 0.01) compared to women in the age group (15-19) years, suggesting that a woman aged 40 and above is less likely to have severe anaemia. Also, the odd ratios for healthy-weight women are consistently positive and statistically significant but increase across thresholds. This means that a woman with a healthy weight has less likelihood to have severe anaemia compared to an underweight woman, with the greatest difference being that healthy-weight women are highly likely to be non-anemic or mild. Although the odd ratio of each parity category decreases across the thresholds, only odd ratios in the first threshold were significant, indicating that as the parity of a woman increases, the likelihood of having severe anaemia reduces. More so, women living in rural residents have significantly higher odds (OR = 1.178, p-value < 0.01) compared to women in urban residents, indicating a woman living in a rural residence is associated with a greater likelihood of having severe anaemia. Moreover, as the wealth index increases, the odds generally decrease. Women in the poorer (OR = 0.857, p-value < 0.01) and middle (OR = 0.788, p-value < 0.01) categories have significantly lower odds compared to women in the poorest category, suggesting that as the wealth index of a woman increases, the likelihood of having severe anaemia reduces while women in richer (OR = 0.699, p-value < 0.01) and richest (OR = 0.531, p-value < 0.01) categories increase across the thresholds, suggesting that a woman in this categories have less likelihood to have severe anaemia compare to a woman in the poorest category. However, women with primary (OR = 0.795, p-value < 0.01), secondary (OR = 0.763, p-value < 0.01), and tertiary (OR = 0.670, p-value < 0.01) education levels have significantly lower odds compared to those with no education, indicating that as the education level of a woman increases, the likelihood of having severe anaemia reduces. In contrast, women who are currently pregnant have higher odds (OR = 1.184, p-value < 0.01) compared to those who are not pregnant, suggesting that a pregnant woman is associated with a greater likelihood of having severe anaemia. Furthermore, the odd ratios for the variable "using modern contraceptives" are consistently positive and statistically significant but decline across thresholds. This means that a woman using modern contraceptives has more likelihood to have severe anaemia compared to a woman not using it, with the greatest difference being that woman using modern contraceptives are less likely to be non-anaemic or mild. Conversely, the improved water source effect is negative but gets larger across thresholds. Hence, women who have access to improved water tend to be less likely to have severe anaemia

compared to women who have access to unimproved water are females, with the greatest differences being that women with improved water are less likely to be in the moderate or severe anemic categories. Additionally, women who belong to the Igbo (OR = 0.894, p-value > 0.05) and Yoruba (OR = 0.831, p-value > 0.05) ethnic categories have insignificant lower odds compared to Hausa. But other ethnic groups have significantly lower odds (OR = 0.676, p-value < 0.01) compared to Hausa. This suggests that a woman from the other ethnic group is less likely to have severe anaemia.

**Table 5**: PPOM for the association between socio-economic and demographic characteristics and severity of anaemia among women in Nigeria.

| Characteristics | 1vs2,3,4 Odd ratio. (SE) | 1,2vs3,4 Odd ratio. (SE) | 1,2,3 vs4 Odd ratio. (SE) |
|---|---|---|---|
| Age group | | | |
| 15-19 (ref) | 0 | 0 | 0 |
| 20-24 | 0.902(0.084) | 0.902(0.084) | 0.902(0.084) |
| 25-29 | 0.833(0.080) | 0.833(0.080) | 0.833(0.080) |
| 30-34 | 0.926(0.094) | 0.926(0.094) | 0.926(0.094) |
| 35-39 | 0.868(0.092) | 0.868(0.092) | 0.868(0.092) |
| >=40 | 0.755**(0.086) | 0.755**(0.086) | 0.755**(0.086) |
| Current marital status | | | |
| Never in union (ref) | 0 | 0 | 0 |
| Married/living with partner | 1.006(0.122) | 1.006(0.122) | 1.006(0.122) |
| widowed/divorced/separated | 1.120(0.168) | 1.120(0.168) | 1.120(0.168) |
| Body mass group | | | |
| Underweight (Ref) | 0 | 0 | 0 |
| Healthy weight | 1.122**(0.045) | 1.168**(0.049) | 1.706**(0.243) |
| Overweight | 1.025(0.129) | 1.025(0.129) | 1.025(0.129) |
| Parity | | | |
| Nulliparity | 0 | 0 | 0 |
| Primiparity | 1.597*(0.332) | 1.149(0.254) | 0.375(0.206) |
| Multiparity | 1.579*(0.326) | 1.183(0.259) | 0.410(0.215) |
| Grand Multiparity | 1.680*(0.351) | 1.192(0.264) | 0.594(0.309) |
| Place of residence | | | |
| Urban (Ref) | 0 | 0 | 0 |
| Rural | 1.178**(0.049) | 1.178**(0.049) | 1.178**(0.049) |
| Wealth index | | | |
| Poorest (Ref) | 0 | 0 | 0 |
| Poorer | 0.857**(0.045) | 0.857**(0.045) | 0.857**(0.045) |
| Middle | 0.788**(0.046) | 0.788**(0.046) | 0.788**(0.046) |
| Richer | 0.699**(0.05) | 0.741**(0.055) | 1.111(0.215) |
| Richest | 0.531**(0.045) | 0.592**(0.053) | 1.043(0.263) |
| Education level | | | |
| No education | 0 | 0 | 0 |
| Primary | 0.795**(0.043) | 0.795**(0.043) | 0.795**(0.043) |
| Secondary | 0.763**(0.043) | 0.763**(0.043) | 0.763**(0.043) |
| Tertiary | 0.670**(0.056) | 0.670**(0.056) | 0.670**(0.056) |
| Ever terminated pregnancy | | | |
| No (Ref) | 0 | 0 | 0 |
| Yes | 1.068(0.057) | 1.088(0.061) | 0.583(0.138) |
| Currently pregnant | | | |
| No (Ref) | 0 | 0 | 0 |
| Yes | 1.184**(0.065) | 1.184**(0.065) | 1.184**(0.065) |
| Currently Breastfeeding | | | |
| No (Ref) | 0 | 0 | 0 |
| Yes | 0.968(0.036) | 0.968(0.036) | 0.968(0.036) |
| Use of modern contraceptive | | | |
| No (Ref) | 0 | 0 | 0 |
| Yes | 0.792**(0.040) | 0.713**(0.041) | 0.453**(0.114) |
| Region | | | |
| North-central (Ref) | 0 | 0 | 0 |
| North-east | 0.808**(0.050) | 0.808**(0.050) | 0.808**(0.050) |
| North-west | 0.655**(0.046 | 0.743**(0.053) | 1.142(0.186) |
| South-east | 1.638**(0.170) | 1.638**(0.170) | 1.638**(0.170) |
| South-south | 1.537**(0.112) | 1.830**(.136) | 2.061**(0.45) |
| South-west | 0.971(0.081) | 0.971(0.081) | 0.971(0.081) |
| Access to media | | | |
| No (Ref) | 0 | 0 | 0 |
| Yes | 1.013(0.042) | 1.079(0.047) | 0.749*(0.107) |

| | | | |
|---|---|---|---|
| Water source | | | |
| Unimproved (Ref) | 0 | 0 | 0 |
| Improved | 1.158**(0.051) | 1.104**(0.05) | 1.878**(0.296) |
| Ethnicity | | | |
| Hausa (Ref) | 0 | 0 | 0 |
| Igbo | 0.894(0.096) | 0.894(0.096) | 0.894(0.096) |
| Yoruba | 0.831(0.082) | 0.831(0.082) | 0.831(0.082) |
| Others | 0.676**(0.039) | 0.676**(0.039) | 0.676**(0.039) |
| Constant | 1.556(0.395) | 0.598(0.158) | 0.031**(0.017) |

*1= Not Anemic, 2= Mild, 3= Moderate, 4=Severe. *Significant at 95% confidence level. **significant at 99% confidence level. Wald test ($\chi^2 = 53.67$, p-value = 0.2661) is insignificant at a 5% confidence level.*

## 4. Discussion

The high anaemia prevalence among women in the current study tallies with evidence of high prevalence from previous African studies.

In contrast, other African studies found low anaemia prevalence among women. (Kibret KT et al., 2019; Gona PN et al., 2021; Hakizimana D, et al., 2019; Habyarimana F, et al., 2020; Hailu BA, et al., 2021). Since anaemia prevalence greater than 40% constitutes a severe public health problem. (McLean E, et al., 2009) These study findings indicate that anaemia among women is a fatal public health problem in Nigeria. Consequently, increased maternal mortality, poor birth outcomes, and reduced productivity due to anaemia among reproductive-age women might persist in Nigeria. (Horton S, Ross J., 2003; Cook RL, et al. 2017) The study findings of significant regional differences in anaemia prevalence and severity among women are consistent with evidence that residing in specific regions or provinces increased the odds of being anemic in other low and middle-income countries (LMICs). (Zegeye B, 2021; Nankinga O, Aguta D., 2019) Similarly to this finding, anaemia prevalence among women was higher in the Northern regions than in Southern regions in a previous Nigerian study.

This study finding that rural residence predicted an increased risk of anaemia and its severity is consistent with increased odds of having severe anemic among women (Assefa E., 2021; Kibret KT, 2019; Abate TW, 2016; Nankinga O and Aguta D, 2019) in prior studies but differs from other studies where a rural residence is protective (Teshale AB, et al., 2020) or urban residence is a risk factor. (Correa-Agudelo E, 2021) Low access to mass media in rural areas resulting in inequitable access to health information might contribute to the risk of being anemic even though media exposure was only significant for moderate anaemia among women in the fitted model. (Abubakar I et al., 2022) Second, Nigerian women residing in urban areas are more likely to be overweight/obese, while rural women are more likely to be either underweight or overweight. (Akokuwebe ME and Idemudia ES., 2021) Third, rural women are less likely to use modern contraceptives compared to urban women. (Ekholuenetale M, et al. 2021) In this study, being underweight and non-use of contraceptives increased the odds of being anemic among women.

This study findings that low education increased the odds of being anemic and anaemia severity among women agreed with evidence from prior studies on women, (Hailu BA, et al., 2021; Ogunsakin RE, et al., 2021) whereas educational attainment was protective of being severely anaemic. (Owais A, et al., 2021; Nti J, et al., 2021; Teshale AB, et al., 2020) As stated elsewhere, (Ogunsakin RE, et al., 2021; Kibret KT, et al., 2019) high education helps women to improve their nutrition and sanitary practices, hence reducing the risk of anaemia.

This study's findings that household socio-economic status predicted severity among women highlight the protective role of wealth and the influence of poverty in being anemic. This finding agrees with existing evidence that being rich reduced the likelihood of severe anaemia among women (Owais A, et al., 2021; Assefa E., 2021; Nti J, et al., 2021; Teshale AB, et al., 2020; Hakizimana D, et al., 2019; Nankinga O, Aguta D., 2019) High socio-economic status improves women's access to improved water, education, healthy weight, as well as better media exposure, which contribute to the prevention of anaemia among women. (Chaparro CM, Suchdev PS., 2019)

The use of modern contraceptives reduces the risk of severe anaemia as reported in several prior studies among women (Owais A, et al., 2021; Teshale AB et al., 2020; Hakizimana D, et al., 2019;) Conversely, contraceptive use increased the likelihood of being severely anemic. (Nti J, et al., 2021) Evidence from the literature indicates that the prevalence of anaemia declines with increasing duration of use of modern contraceptives in one of three ways. (Gebremedhin S. and Asefa A, 2019) First, spacing births reduces nutritional stress associated with successive pregnancies, preventing maternal iron depletion. Second, some contraceptives modify the iron status, which lowers the risk of being anemic. Third, iron-containing contraceptives provide iron supplementation to prevent iron deficiency anaemia. Further studies are needed to determine which of these factors plays a greater role in predisposing women in Nigeria to the severity of anaemia. (Gebremedhin S. and Asefa A, 2019)

Similar to findings in Ethiopia, normal body weight reduced the likelihood of severe anaemia among women in this study. (Tsehayu A, et al., 2022) However, being underweight increases the risk of anaemia among women. (Assefa E, 2021; Hakizimana D, et al., 2019; Correa-Agudelo E, 2021; Habyarimana F, et al., 2020) Anaemia, due to being underweight,

might result from undernourishment, but poor dietary diversity was not included as a predictor of being anemic in this study. Further studies are required to determine the relationship between micronutrient-rich diets, BMI and severity of anaemia among women in Nigeria.

Furthermore, our finding that lack of media exposure increased the likelihood of being moderately anemic among women is consistent with evidence from a preceding study. (Keokenchanh S, et al. 2021) Access to media is associated with awareness of the value and availability of health services, (Abubakar I, et al., 2022; Igbinoba AO, et al., 2020) information on diets and nutrition, and an increased likelihood of contraceptive use. (Ajaero CK, et al., 2016; Yaya S, Bishwajit G, 2020) Furthermore, the finding that age 40 and above reduced the odds of developing severe anaemia among women contrasts evidence of reduced odds of anaemia among this age group in Ethiopia. (Kibret KT, et al. 2019).

## 5. Conclusion

Anaemia is still a big problem in Nigeria. Policies should be centred on education, alleviation of poverty, and enhancement of maternal health. Interventions based on evidence include:

- Improve access to healthcare in rural areas by sending out mobile clinics and paying for iron supplementation programs. This is in line with Nigeria's National Anaemia Reduction Policy, which says that the risk of severe anaemia is 1.8 times higher in rural areas.
- Promote education and nutrition literacy: Include anaemia prevention in school curricula and community health campaigns, focusing on groups with low levels of education (OR=2.1 for severe anaemia).
- To help poor people get out of poverty, expand conditional cash transfer programs that are linked to maternal health check-ups for poor families (OR=2.5 risk).
- Improve the infrastructure for water and sanitation: Speed up WASH investments in areas with poor water quality (OR=1.4 risk).
- Integrating maternal health: Require preconception and antenatal screening with complimentary treatment during pregnancy (OR=1.6 risk), thereby endorsing SDG 2 and 3.
- To lower Nigeria's 58% anaemia rate among women and reach national health goals, we need these multisectoral policies that deal with socio-economic inequalities right away.

## References


Accinelli, R., & Leon-Abarca, J. A. (2020). Age and altitude of residence determine anaemia prevalence in Peruvian 6 to 35 months old children. *PLoS ONE*, 15.

Adelman, S., Gilligan, D., Konde-Lule, J., & Alderman, H. (2019). School Feeding Reduces Anaemia Prevalence in Adolescent Girls and Other Vulnerable Household Members in a Cluster Randomized Controlled Trial in Uganda. *Journal of NutriLife*, 149, 659–666.

Agustina, R., Wirawan, F., Sadariskar, A. A., Setianingsing, A. A., Nadiya, K., Prafiantini, E., Asri, E., Purwanti, T., Kusyuniati, S., Karyadi, E., & Raut, M. K. (2021). Associations of Knowledge, Attitude, and Practices toward Anaemia with Anaemia Prevalence and Height-for-Age Z-Score among Indonesian Adolescent Girls. *Food and Nutrition Bulletin*, 42, S92–S108.

Asghari, S., Mohammadzadegan-Tabrizi, R., Rafraf, M., Sarbakhsh, P., & Babaie, J. (2020). Prevalence and predictors of iron-deficiency anaemia: Women's health perspective at reproductive age in the suburb of dried Urmia Lake, Northwest of Iran. *Journal of Education and Health Promotion,* 9.

Assefa E. Multilevel analysis of anaemia levels among reproductive age groups of women in Ethiopia. SAGE Open Med 2021; 9: 0987375

Bolarinwa, O. A., Tessema, Z., Frimpong, J. B., Seidu, A., & Ahinkorah, B. (2021). Spatial distribution and factors associated with modern contraceptive use among women of reproductive age in Nigeria: A multilevel analysis. *PLoS ONE*, 16.

Correa-Agudelo E, Kim HY, Musuka GN, et al. The epidemiological landscape of anaemia in women of reproductive age in sub-Saharan Africa. Sci Rep 2021; 11(1): 11955. in Nigeria. J Public Health 2022; 44: 111–120.

da Silva, F. A. C., Ferreira, A. L., Pimentel, L., Lyra, C. H. M. M., Hazin-Costa, M. F., Guerra, G., Araújo, A., & Souza, A. I. (2021). The Use of Hydroxyurea During Pregnancy in Sickle Cell Anaemia Women: A Case Series and Literature Review. *Journal of Hematology Research.*

Effiong, J. O., & Udom, V. B. (2025). *Prevalence of anaemia and malaria parasitaemia among pregnant women in Nigeria: A cross-sectional analysis.* International Journal of Interdisciplinary Research in Medical and Health Sciences, 12(4), 17–22. https://doi.org/10.5281/zenodo.17514411

Fawole, O., Okedare, O. O., & Reed, E. (2020). Home was not a safe haven: women's experiences of intimate partner violence during the COVID-19 lockdown in Nigeria. *BMC Women's Health,* 21.

Gardner, W., & Kassebaum, N. (2020). Global, Regional, and National Prevalence of Anaemia and Its Causes in 204 Countries and Territories, 1990–2019. *Current Developments in Nutrition.*

Girum, T., & Wasie, A. (2018). The Effect of Deworming School Children on Anaemia Prevalence: A Systematic Review and Meta-Analysis. *Open Nursing Journal*, 12, 155–161.

Gona PN, Gona CM, Chikwasha V, et al. Intersection of HIV and anaemia in women of reproductive age: a 10-year analysis of three Zimbabwe demographic health surveys, 2005–2015. BMC Public Health 2021; 21(1): 41

Habyarimana F, Zewotir T, Ramroop S. Prevalence and risk factors associated with anaemia among women of childbearing age in Rwanda. Afr J Reprod Health 2020; 24(2): 141–151. [PubMed] [Google Scholar]
Hailu BA, Laillou A, Chitekwe S, et al. Subnational mapping for targeting anaemia prevention in women of reproductive age in Ethiopia: a coverage-equity paradox. Matern Child Nutr 2021; 2021: e13277. [PMC free article] [PubMed] [Google Scholar]
Hakizimana D, Nisingizwe MP, Logan J, et al. Identifying risk factors of anaemia among women of reproductive age in Rwanda – a cross-sectional study using secondary data from the Rwanda demographic and health survey 2014/2015. BMC Public Health 2019; 19(1): 1662. [PMC free article] [PubMed] [Google Scholar]
Jasim, S., Al-Momen, H., & Al-asadi, F. (2020). Maternal Anaemia Prevalence and Subsequent Neonatal Complications in Iraq. *Open Access Macedonian Journal of Medical Sciences*, 8, 71–75.
Karami, M., Chaleshgar, M., Salari, N., Akbari, H., & Mohammadi, M. (2022). Global Prevalence of Anaemia in Pregnant Women: A Comprehensive Systematic Review and Meta-Analysis. *Maternal and Child Health Journal*, 26, 1473–1487.
Keokenchanh, S., Kounnavong, S., Tokinobu, A., Midorikawa, K., Ikeda, W., Morita, A., Kitajima, T., & Sokejima, S. (2021). Prevalence of Anaemia and Its Associate Factors among Women of Reproductive Age in Lao PDR: Evidence from a Nationally Representative Survey. *Anaemia,* 2021.
Kibret KT, Chojenta C, D'Arcy E, et al. Spatial distribution and determinant factors of anaemia among women of reproductive age in Ethiopia: a multilevel and spatial analysis. BMJ Open 2019; 9(4): e027276.
Kinyoki, D. K., Osgood-Zimmerman, A., Bhattacharjee, N. V., Letournea, L. E. A. D. S. B. K. M. I. D. M. M. S. L.-A. L. E. D., Schaeffer, L. E., Lazzar-Atwood, A., Lu, D., Ewald, S. B., Donkers, K. M., Letourneau, I. D., Collison, M. L., Schipp, M. F., Abajobir, A., Abbasi, S., Abbasi, N., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., … Hay, S. (2021). Anaemia prevalence in women of reproductive age in low- and middle-income countries between 2000 and 2018. *Nature Network Boston*, 27, 1761–1782.
Li, H., Xiao, J., Liao, M., Huang, G., Zheng, J., Wang, H., Huang, Q., & Wang, A. (2020). Anaemia prevalence, severity and associated factors among children aged 6–71 months in rural Hunan Province, China: a community-based cross-sectional study. *BMC Public Health*, 20.
Liyew, A. M., Tesema, G., Alamneh, T. S., Worku, M., Teshale, A. B., Alem, A. Z., Tessema, Z., & Yeshaw, Y. (2021). Prevalence and determinants of anaemia among pregnant women in East Africa; A multi-level analysis of recent Demographic and Health Surveys. *PLoS ONE*, 16.
Long, J. S., and J. Freese. 2006. Regression Models for Categorical Dependent Variables Using Stata. 2nd ed. College Station, TX: Stata Press.https://en.m.wikipedia.org/wiki/Stata
Majeed, M., Farooq, S., Zaib, S., Niazi, Z., & Rasheed, T. (2022). Prevalence of Anaemia in Pregnant Women. *Pakistan Journal of Medical & Health Sciences*.
Mayang S, Saskia P, Elviyanti M, Susilowati H, Sugiatmi, Martin WB, Ray Y. Estimating the prevalence of anaemia: A comparison of three method. Bulletin of the World Health Organization, 2001;79:506–511. PMid:11436471, PMCid:2566437
McLean E, Cogswell M, Egli I, et al. Worldwide prevalence of anaemia, WHO vitamin and mineral nutrition information system, 1993–2005. Public Health Nutr 2009; 12(4): 444–454
Nankinga O, Aguta D. Determinants of anaemia among women in Uganda: further analysis of the Uganda demographic and health surveys. BMC Public Health 2019; 19(1): 1757.
Nti J, Afagbedzi S, da-Costa Vroom FB, et al. Variations and determinants of anaemia among reproductive age women in five sub-Saharan Africa countries. Biomed Res Int 2021; 2021: 9957160.
Obayelu, O. A., & Chime, A. C. (2020). Dimensions and drivers of women's empowerment in rural Nigeria. *International Journal of Social Economics*, 47, 315–333.
Odimegwu, C., Alabi, O., Wet, N. D., & Akinyemi, J. (2018). Ethnic heterogeneity in the determinants of HIV/AIDS stigma and discrimination among Nigeria women. *BMC Public Health,* 18.
Omotade, A. A., A.A, O., & J.O, F. (2015). The Contributions of Nigeria Women Towards National Development. *International Journal for Innovation Education and Research*, 3.
Orpin, J., Papadopoulos, C., & Puthussery, S. (2020). The Prevalence of Domestic Violence Among Pregnant Women in Nigeria: A Systematic Review. *Trauma, Violence, & Abuse*, 21, 15–3.
Orsango, A., Loha, E., Lindtjørn, B., & Engebretsen, I. (2020). Efficacy of processed amaranth-containing bread compared to maize bread on hemoglobin, anaemia and iron deficiency anaemia prevalence among two-to-five-year-old anemic children in Southern Ethiopia: A cluster randomized controlled trial. *PLoS ONE*, 15.
Rappaport, A. I., Karakochuk, C. D., Whitfield, K. C., Kheang, K. M., & Green, T. (2017). A method comparison study between two hemoglobinometer models (Hemocue Hb 301 and Hb 201+) to measure hemoglobin concentrations and estimate anaemia prevalence among women in Preah Vihear, Cambodia. *International Journal of Laboratory Hematology*, 39.
Richard Williams. Generalized ordered logit/partial proportionalodds models for ordinal dependent variables, 2006.
Sharif, N., Das, B., & Alam, A. (2023). Prevalence of anaemia among reproductive women in different social group in India: Cross-sectional study using nationally representative data. *PLoS ONE,* 18.
Tesema, G., Worku, M., Tessema, Z., Teshale, A. B., Alem, A. Z., Yeshaw, Y., Alamneh, T. S., & Liyew, A. M. (2021). Prevalence and determinants of severity levels of anaemia among children aged 6–59 months in sub-Saharan Africa: A multilevel ordinal logistic regression analysis. *PLoS ONE*, 16.

Zegeye B, Ahinkorah BO, Ameyaw EK, et al. Determining prevalence of anaemia and its associated factors in Cameroon: a multilevel analysis. Biomed Res Int 2021; 2021: 9912549.n

## Declaration

**Conflict of Study:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
**Funding Statement:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
**Author Contribution Statement**: All the authors worked together as a cell group and designed the research; Analyzed and interpreted the data, and wrote the paper.
**Availability of Data and Material**: Data will be made available by the corresponding author on reasonable request.
**Ethical Approval**: Not Applicable
**Consent to Publish**: Not Applicable
**Consent to Participate**: Not Applicable
**Acknowledgement**: Not Applicable

## Author(s) Bio / Authors' Note

***Onyenekwe Chukwuenyem E*** *is from the Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria. Email:* ce.onyenekwe@unizik.edu.ng

***Oladuti Mercy. O*** *is from the Department of Statistics, School of Physical Sciences, Federal University of Technology, Akure, Ondo State, Nigeria. Email:* omoladuti@futa.edu.ng

***Adubiaro Dunsin*** *is from the Department of Statistics, School of Physical Sciences, Federal University of Technology, Akure, Ondo State, Nigeria. Email:* adibexdunsin@gmail.com

***Onyenekwe Amala Joy*** *is from. Department of Environmental Health Sciences, Faculty of Health Sciences and Technology, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria. Email:* aj.mmadubugwu@unizik.edu.ng

SCOPUA Journal of Applied Statistical Research

Vol.2, Issue 2, June 2026

**DOI:** https://doi.org/10.64060/JASR2V2i2

# Estimation for a Regression Model based on Log Generalized Inverse Generalized Weibull distribution under Type I Censoring

Neetu Singla [1], Kanchan Jain [2*]

[1] Altimetric India Pvt Ltd, Bangalore, India
[2] Formerly at Department of Statistics, Panjab University, Chandigarh, India
* Corresponding Email: jaink14@gmail.com

Received: 18 February 2026 / Revised: 09 April 2026 / Accepted: 15 April 2026 / Published online: 25 April 2026



**ABSTRACT**

The Log Generalized Inverse Generalized Weibull (LGIGW) distribution, obtained by taking the logarithm of a Generalized Inverse Generalized Weibull (GIGW) random variable, is introduced and studied in this work. The proposed distribution is particularly useful for modeling lifetime, financial, and reliability data due to its flexibility in capturing various data behaviours. It encompasses several well-known classical distributions as special cases. In addition, a corresponding regression model based on the LGIGW distribution is developed. This model generalizes a number of existing regression models available in the literature, which arise as special cases under specific parameter settings. Parameter estimation is carried out using the method of maximum likelihood, and the performance of the estimators is assessed through Monte Carlo simulation studies.The applicability of the proposed model is illustrated using two real-life data sets. Model comparison is conducted using standard information criteria, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and the corrected Akaike Information Criterion (AICc). Furthermore, sensitivity analysis and residual diagnostic techniques are employed to identify influential observations.

**Keywords:** Regression, Censoring, Maximum Likelihood Estimation, Sensitivity Analysis, Residual Techniques

## 1 Introduction

Modeling lifetime data under censoring is a central problem in survival analysis, reliability theory, and biomedical research. Classical parametric models such as the Weibull distribution remain popular due to their flexibility and tractability. However, standard Weibull models are often inadequate for describing complex hazard rate shapes observed in practice, particularly in the presence of non-monotonic hazard functions. To address such limitations, several generalizations of the Weibull distribution have been proposed. The Inverse Weibull (IW) distribution introduced by Keller and Kamath (1982) has been widely used in reliability and survival studies. Subsequent works including Singh et al. (2013); Sultan et al. (2014); Nassar and Abo-Kasem (2017) investigated classical and Bayesian estimation under various censoring schemes. Krishna et al. (2019) further explored stress–strength reliability inference for the IW model. The logarithmic transformation of lifetime distributions has led to an important class of log-location-scale regression models. Kumar and Nair (2018) studied the Log Inverse Weibull (LIW) distribution whose applicability is limited in cases where the hazard function is non-decreasing. A major advancement was made by De Gusmao et al. (2011), who introduced the Generalized Inverse Weibull (GIW) and Log Generalized Inverse Weibull (LGIW) distributions, providing greater flexibility in modeling hazard behavior. More recently, Jain et al. (2014) proposed the Generalized Inverse Generalized Weibull (GIGW) distribution, which further extends the flexibility of inverse Weibull-type models. Motivated by this development, we introduce the Log Generalized Inverse Generalized Weibull (LGIGW) distribution, obtained by logarithmic transformation of a GIGW random variable. The proposed distribution accommodates increasing hazard rates and contains LGIW and LIW distributions as special cases. Other references in the area include Alsaggaf et al. (2024), Frolov et al. (2024), Aljeddani and Mohammed (2023), Temraz et al. (2022), El-Morshedy et al. (2022), Alotaibi et al. (2022).

There has been substantial progress in log-type regression modeling. Notable contributions include Log Birnbaum–Saunders regression model (Tabassum and Altaf (2025)) , Log-weighted exponential regression

model (Altun (2021)) , Log-beta generalized half-normal regression model (Pescim et al. (2013)), Log-Kumaraswamy Generalized Gamma (Pascoa et al. (2013)), Log Generalized Modified Weibull (Ortega et al. (2011)), Log-Birnbaum-Saunders (Qu and Xie (2011)), Log-Exponentiated Weibull regression model (Hashimoto et al. (2010)).

Statistical inference in high-dimensional binary GLMs was addressed by Cai et al. (2023) and the `SIHR` package by Rakshit et al. (2021). The ordered beta regression model was proposed by Kubinec (2023). Penalized multinomial logistic regression with DP weights was developed in Fuetterer et al. (2025). A recent roadmap for logistic regression modeling was given in Bui and Ding (2025). Despite these advances, existing models may lack sufficient flexibility for modeling lifetime data exhibiting strong skewness. The LGIGW regression model proposed in this paper provides a unified and highly flexible framework encompassing several existing models as subcases. Specifically, the LGIW and LIW regression models arise as special cases when $\alpha = 1$ and $(\alpha, \gamma) = (1, 1)$, respectively. This hierarchical structure places the LGIGW model as a natural generalization of existing inverse Weibull-based regression frameworks.

The main contributions of this paper are Introduction of the LGIGW distribution and its properties, Development of the LGIGW regression model under Type I censoring, Derivation of likelihood equations and observed information matrix, Asymptotic inference including confidence interval construction, Simulation study for estimation of parameters. Empirical comparison with competing submodels, Sensitivity and residual diagnostics for detection of influential or extreme observations.For detection of influential observations, a lot of work has been done. Few recent references in this context are Khaleeq et al. (2021), Khan et al. (2025), Soale (2025).

The paper is organized as follows. In Section 2, the Log Generalized Inverse Generalized Weibull (LGIGW) distribution has been introduced. The corresponding regresssion model has been discussed in Section 3. Nonlinear equations for finding maximum likelihood estimates and the elements of information matrix have also been derived in this section. In Section 4, simulation study results are included. Section 5 consists of application to two real life data sets for which the proposed regression model has been compared with Log Generalized Inverse Weibull (LGIW) and Log Inverse Weibull (LIW) regression models using AIC, BIC and AICc. Section 6 discusses some sensitivity techniques viz global and local influence techniques under various perturbation schemes for identifying influential observations. To check the outliers, residual analysis is explained in Section 7. Conclusions are reported in Section 8.

## 2 Log Generalized Inverse Generalized Weibull Distribution

Let X be a random variable following Generalized Inverse Generalized Weibull (GIGW) distribution proposed by Jain et al. (2014). For $X \sim \text{GIGW}(\alpha, \gamma, \lambda, \beta)$, the density function of GIGW is given by

$$f(x) = \alpha\lambda^{\beta}\beta\gamma x^{-(\beta+1)} e^{-\gamma(\lambda/x)^{\beta}} (1 - e^{-\gamma(\lambda/x)^{\beta}})^{(\alpha-1)}; x > 0. \tag{2.1}$$

Then $Y = \log(X)$ follows Log Generalized Inverse Generalized Weibull (LGIGW) distribution. The probability density function (pdf) of Y is derived as

$$g(y) = \alpha\gamma\beta\lambda^{\beta} e^{-(\beta+1)y} e^{-\gamma\lambda^{\beta}e^{-y\beta}} \left\{1 - e^{-\gamma\lambda^{\beta}e^{-y\beta}}\right\}^{\alpha-1} e^{y}$$

$$= \alpha\gamma\beta\{e^{\beta\log\lambda - y\beta - \gamma e^{(\beta\log\lambda - y\beta)}}\}\{1 - e^{-\gamma e^{(\beta\log\lambda - y\beta)}}\}^{\alpha-1}, -\infty < y < \infty \tag{2.2}$$

Putting $\sigma = 1/\beta$ and $\mu = \log\lambda$ in (2.2), we write

$$g(y) = \frac{\alpha\gamma}{\sigma} e^{\left\{-\left(\frac{y-\mu}{\sigma}\right) - \gamma e^{\left[-\left(\frac{y-\mu}{\sigma}\right)\right]}\right\}} \left\{1 - e^{-\gamma e^{\left[-\left(\frac{y-\mu}{\sigma}\right)\right]}}\right\}^{\alpha-1}, -\infty < y < \infty \tag{2.3}$$

where $\alpha, \gamma, \sigma > 0$ and $-\infty < \mu < \infty$.

Accordingly, the LGIGW distribution is parameterized by $(\alpha, \gamma, \mu, \sigma)$, with parameter space given by

$$\alpha > 0, \quad \gamma > 0, \quad \sigma > 0, \quad \mu \in \mathbb{R}.$$

where $\sigma = 1/\beta$ and $\mu = \log \lambda$.
The corresponding survival function is

$$S(y) = \int_y^\infty \frac{\alpha\gamma}{\sigma} \exp\left\{-(\frac{m-\mu}{\sigma}) - \gamma \exp\left[-(\frac{m-\mu}{\sigma})\right]\right\} \left\{1 - \exp^{-\gamma e^{\left[-(\frac{m-\mu}{\sigma})\right]}}\right\}^{\alpha-1} dm$$

$$= \int_0^{e^{\left\{-(\frac{y-\mu}{\sigma})\right\}}} \alpha\gamma e^{(-\gamma t)} \left\{1 - e^{(-\gamma t)}\right\}^{\alpha-1} dt \text{ by taking } e^{\left\{-(\frac{m-\mu}{\sigma})\right\}} = t.$$

Putting $e^{(-\gamma t)} = s$ and simplifying, we get

$$S(y) = \left\{1 - e^{-\gamma e^{\left[-(\frac{y-\mu}{\sigma})\right]}}\right\}^{\alpha}. \tag{2.4}$$

The pdf of the standardized random variable $Z = \frac{Y-\mu}{\sigma}$ can be written as

$$\pi(z; \alpha, \gamma) = \alpha\gamma e^{\{-z-\gamma e^{-z}\}}[1 - e^{(-\gamma e^{-z})}]^{\alpha-1}, -\infty < z < \infty, \alpha > 0, \gamma > 0. \tag{2.5}$$

The $k^{th}$ order moment of Z is

$$E(Z^k) = \int_{-\infty}^{\infty} \alpha\gamma z^k e^{\{-z-\gamma e^{-z}\}}[1 - e^{(-\gamma e^{-z})}]^{\alpha-1} dz.$$

By substituting $u = e^{-z}$ and using the following result

$$\int_0^\infty (\log u)^k e^{(-\gamma u)} du = \left.\frac{\partial^k [\gamma^{-a}\Gamma a]}{\partial a^k}\right|_{a=1},$$

we get

$$E(Z^k) = \gamma(-1)^k \Gamma(\alpha+1) \sum_{j=0}^{\infty} \frac{(-1)^j}{\Gamma(\alpha-j)j!} \left\{\frac{1}{j} \left.\frac{\partial^k [\gamma^{-a}\Gamma a]}{\partial a^k}\right|_{a=1} - \frac{\log(j+1)^k}{\gamma j}\right\}. \tag{2.6}$$

The skewness and kurtosis are given by

$$\text{Skewness} = \frac{\mathbb{E}\left[(Z - \mathbb{E}(Z))^3\right]}{(\text{Var}(Z))^{3/2}}, \quad \text{Kurtosis} = \frac{\mathbb{E}\left[(Z - \mathbb{E}(Z))^4\right]}{(\text{Var}(Z))^2}.$$

Numerical computations of the first four moments can be carried out for selected parameteric values and skewness and kurtosis obtained. The following table gives values of skewness and kurtosis for distribution of Z and also comments about the shape of the distribution.

Table 1: Skewness and Kurtosis for selected parameter values

| $\alpha$ | $\gamma$ | Skewness | Kurtosis | Shape |
|---|---|---|---|---|
| 1.0 | 0.6 | 1.28 | 4.75 | Highly right-skewed, heavy-tailed |
| 1.5 | 1.0 | 0.89 | 3.62 | Moderately right-skewed |
| 1.9 | 1.3 | 0.63 | 3.18 | Mild skewness |
| 2.0 | 1.8 | 0.48 | 3.05 | Slightly skewed, near normal |
| 2.3 | 2.0 | 0.28 | 2.91 | Almost symmetric, slightly platykurtic |

As parameter values increase, the distribution of Z exhibits low skewness and nearly mesokurtic behavior, indicating that it approaches symmetry and resembles the normal distribution.

## 2.1 Properties of LGIGW distribution

Figure 1 illustrates the density for selected parameter combinations.

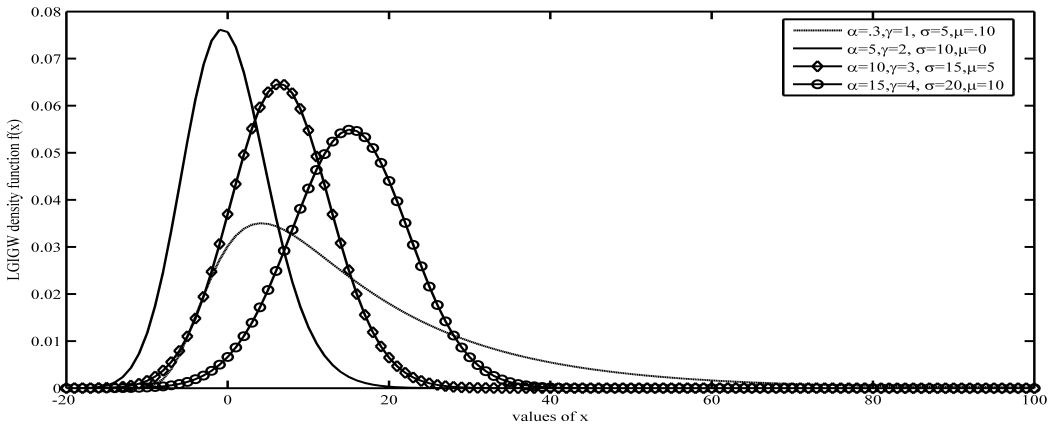


**Figure 1: LGIGW density plots**

It is observed that

- for small $\alpha$, the distribution is more right-skewed. An increase in $\alpha$ leads to a reduction in right skewness.
- An increase in $\gamma$ results in a more pronounced peak and lighter tail of the distribution.
- Higher values of $\sigma$ increase dispersion while preserving the direction of skewness.
- The plots confirm that LGIGW accommodates a wide range of shapes, including those corresponding to LGIW ($\alpha$ = 1) and LIW (($\alpha$ = $\gamma$ = 1)

**Remark 1:**

- LGIW distribution De Gusmao et al. (2011) is a particular case of LGIGW distribution for $\alpha = 1$.
- For $\alpha = \gamma = 1$,((2.5)) takes the form $e^{\{-z-e^{-z}\}}$, which is the pdf of Inverse Extreme Value distribution or Log Inverse Weibull (LIW) distribution De Gusmao et al. (2011).

In survival studies, some observations are lost to further follow up and are considered to be censored Klein and Moeschberger (2003). In the next section, we discuss LGIGW regression model in case of censoring. The expressions for elements of score vector and information matrix have been derived.

## 3 Log Generalized Inverse Generalized Weibull Regression Model

In real life, covariates such as age, weight, smoking status and diabetes level affect the lifetimes. If there are n individuals under study, then $t_i$ denotes the lifetime and $c_i$, the censoring time for $i^{th}$ individual, $i = 1, ..., n$. The lifetimes and the censoring times are assumed to be independent. The response variable $y_i$ is taken to be $\min(\log(t_i), \log(c_i))$. In the regression model, it is assumed that each response variable depends upon p covariates.
A linear regression model for the response variable $y_i$ is written as

$$y_i = \boldsymbol{x_i}^T\boldsymbol{\beta} + \sigma z_i, \quad i = 1, ..., n \tag{3.1}$$

where

- $z_i$ is the random error with pdf given by (2.5);
- $x_i = (x_{i1}, ..., x_{ip})^T$ is the vector of covariates for $i = 1, ..., n$;
- $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is the vector of unknown regression coefficients;
- $\alpha > 0, \gamma > 0$ and $\sigma > 0$ are unknown scalar parameters;
- the vector $\boldsymbol{x}_i$ models the location parameter $\mu_i$ given by $\boldsymbol{x}_i{}^T\boldsymbol{\beta}$.

In matrix notation, (3.1) can be written as

$$\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{\beta} + \sigma\boldsymbol{Z} \tag{3.2}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \left[\beta_1, \beta_2, \vdots, \beta_p\right], \qquad \boldsymbol{Z} = \left[z_1, z_2, \ldots z_n\right]$$

The survival function of $y_i$ can be estimated as

$$S(y_i, \widehat{\alpha}, \widehat{\gamma}, \widehat{\sigma}, \widehat{\boldsymbol{\beta}}^{\mathbf{T}}) = \left[1 - e^{\left\{-\widehat{\gamma} e^{(-(\frac{y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}))}\right\}}\right]^{\widehat{\alpha}}. \tag{3.3}$$

**Remark 2:**

- For $\alpha = 1$, equation (3.1) simplifies to the LGIW regression model introduced by De Gusmao et al. (2011).
- For $\gamma = 1$, the proposed model reduces to the Log Inverse Generalized Weibull (LIGW) regression model.
- For $\alpha = 1$ and $\gamma = 1$, the model further reduces to the Log Inverse Weibull (LIW) regression model.

## 3.1 Estimation of Parameters

Let

$$F = \{\, i \mid \delta_i = 1 \,\}, \quad C = \{\, i \mid \delta_i = 0 \,\}$$

denote the sets of observed failures and right-censored observations respectively. Using (2.5) and (3.1), the total log likelihood function for parameters $\boldsymbol{\theta} = (\alpha, \gamma, \sigma, \boldsymbol{\beta}^T)^T$ can be written as

$$\begin{aligned} l(\boldsymbol{\theta}) = r\left\{\log\alpha + \log\gamma - \log\sigma\right\} - \sum_{i\in F}(\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}) - \gamma \sum_{i\in F} e^{(-(\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}))} \\ + (\alpha - 1)\sum_{i\in F}\log\left[1 - e^{\{-\gamma e^{(-(\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}))}\}}\right] + \alpha\sum_{i\in C}\log\left[1 - e^{\{-\gamma e^{(-(\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}))}\}}\right], \end{aligned} \tag{3.4}$$

where r is the number of observed failures.
For simplicity, we write

$$e^{\{-\gamma e^{(-(\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}))}\}} = m_i \text{ and } (\frac{y_i - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}) = z_i \text{ in the sequel.}$$

The MLE of $\boldsymbol{\theta}$ can be obtained by maximizing log-likelihood function given by (3.4).This is attained by solving $\frac{\partial l}{\partial \alpha} = 0$, $\frac{\partial l}{\partial \gamma} = 0$, $\frac{\partial l}{\partial \sigma} = 0$ and $\frac{\partial l}{\partial \beta_j} = 0$, where

$$\frac{\partial l}{\partial \alpha} = \frac{r}{\alpha} + \sum_{i\in F}\log\left[1 - m_i\right] + \sum_{i\in C}\log\left[1 - m_i\right] \ ;$$

$$\frac{\partial l}{\partial \gamma} = \frac{r}{\gamma} - \sum_{i\in F} e^{-z_i} + (\alpha - 1)\sum_{i\in F}\frac{m_i e^{-z_i}}{[1 - m_i]} + \alpha\sum_{i\in C}\frac{m_i e^{-z_i}}{[1 - m_i]} \ ;$$

$$\frac{\partial l}{\partial \sigma} = \frac{-r}{\sigma} + \sum_{i\in F} z_i - \gamma\sum_{i\in F} e^{-z_i}\frac{z_i}{\sigma} + (\alpha - 1)\gamma\sum_{i\in F}\frac{m_i e^{-z_i} z_i}{[1 - m_i]} + \alpha\gamma\sum_{i\in C}\frac{m_i e^{-z_i} z_i}{[1 - m_i]} \ ;$$

$$\frac{\partial l}{\partial \beta_j} = \sum_{i\in F}\frac{x_{ij}}{\sigma} - \gamma\sum_{i\in F} e^{-z_i}\frac{x_{ij}}{\sigma} + (\alpha - 1)\gamma\sum_{i\in F}\frac{m_i e^{-z_i}\frac{x_{ij}}{\sigma}}{[1 - m_i]} + \alpha\gamma\sum_{i\in C}\frac{m_i e^{-z_i}\frac{x_{ij}}{\sigma}}{[1 - m_i]} \ .$$

As it is difficult to find the MLEs analytically, hence we shall be using numerical methods for finding the estimates in Section 4.

For interval estimation and testing of hypotheses for the parameters, the $4 \times 4$ observed information matrix is obtained as

$$\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{\theta}) = -\begin{bmatrix} \mathrm{J}_{\alpha,\alpha} & \mathrm{J}_{\alpha,\gamma} & \mathrm{J}_{\alpha,\sigma} & \mathrm{J}_{\alpha,\beta_j} \\ & \mathrm{J}_{\gamma,\gamma} & \mathrm{J}_{\gamma,\sigma} & \mathrm{J}_{\gamma,\beta_j} \\ & & \mathrm{J}_{\sigma,\sigma} & \mathrm{J}_{\sigma,\beta_j} \\ & & & \mathrm{J}_{\beta_j,\beta_s} \end{bmatrix}.$$

The expression for elements of matrix J have been included in the Appendix.
The parameter space is defined such that all scale and shape parameters are strictly positive, while location-type parameters are real-valued, and the true parameter lies in the interior of the parameter space. The LGIGW model is identifiable, as distinct parameter values correspond to distinct distributions due to the identifiability of its component distributions and the one-to-one transformation structure. The log-likelihood function is assumed to be twice continuously differentiable, and the Fisher information matrix is finite and positive definite. In the presence of censored observations, a non-informative and independent censoring mechanism is assumed, ensuring the validity of the likelihood function. Under these regularity conditions, the maximum likelihood estimator exists, is consistent, and asymptotically normal.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}_{p+3}\left(0, I(\theta)^{-1}\right), \tag{3.5}$$

where $I(\theta)$ is the expected information matrix, approximated by the observed information matrix $J(\hat{\theta})$. Thus,

$$\hat{\theta} \approx \mathcal{N}_{p+3}\left(\theta, J(\hat{\theta})^{-1}\right). \tag{3.6}$$

An approximate $100(1-\xi)\%$ confidence interval for parameter $\theta_j$ is

$$\hat{\theta}_j \pm z_{\xi/2}\sqrt{\widehat{\mathrm{Var}}(\hat{\theta}_j)}, \tag{3.7}$$

where

- $z_{\xi/2}$ is the standard normal quantile,
- $\widehat{\mathrm{Var}}(\hat{\theta}_j)$ is the $j^{th}$ diagonal element of $J(\hat{\theta})^{-1}$.

These intervals allow statistical inference for regression coefficients and shape parameters.

# 4 Simulation Study

## 4.1 Design of the Study

A Monte Carlo simulation study was conducted to evaluate the finite-sample performance of the maximum likelihood estimators (MLEs) of the LGIGW regression model under Type I censoring.
Data were generated from model (3.1) with two covariates. The true parameter values were chosen as $\alpha = 0.50$, $\gamma = 1.50, \quad \sigma = 0.80, \quad \beta = (0.50, -0.30)^T$.
Covariates were generated as $x_{i1} \sim \text{Uniform}(0,1), \quad x_{i2} \sim \text{Uniform}(0,1)$.
Sample sizes considered were $n = 30, 50, 100$.
Type I censoring was imposed to achieve approximately 20% and 40% censoring rates.
For each configuration, 1000 replications were performed.

## 4.2 Performance Measures

The following metrics were computed:

- Bias: $E(\hat{\theta} - \theta)$
- Mean Squared Error (MSE): $E[(\hat{\theta} - \theta)^2]$
- Coverage Probability (CP) of 95% confidence intervals

The model parameters were estimated via numerical maximization of the log-likelihood function using a quasi-Newton algorithm in R software. Convergence was declared when the change in parameter estimates and the log-likelihood values between successive iterations fell below $10^{-6}$ and $10^{-8}$ respectively, or when the maximum number of iterations $(1000)$ was reached. The optimization routine was implemented in R using the optim function, and convergence was assessed based on both relative tolerance and gradient norm criteria.

### 4.3 Results

Tables 2 and 3 present results under 20% and 40% censoring respectively.

Table 2: Simulation Results under 20% Censoring

| Parameter | n | Bias | MSE | CP (95%) |
|---|---|---|---|---|
| $\alpha$ | 30 | 0.021 | 0.018 | 0.93 |
| | 50 | 0.012 | 0.010 | 0.94 |
| | 100 | 0.005 | 0.004 | 0.95 |
| $\gamma$ | 30 | 0.084 | 0.126 | 0.92 |
| | 50 | 0.042 | 0.068 | 0.94 |
| | 100 | 0.018 | 0.022 | 0.95 |
| $\sigma$ | 30 | 0.016 | 0.012 | 0.94 |
| | 50 | 0.008 | 0.006 | 0.95 |
| | 100 | 0.003 | 0.002 | 0.95 |
| $\beta_1$ | 30 | 0.014 | 0.008 | 0.94 |
| | 50 | 0.006 | 0.004 | 0.95 |
| | 100 | 0.002 | 0.001 | 0.95 |
| $\beta_2$ | 30 | -0.018 | 0.009 | 0.93 |
| | 50 | -0.007 | 0.004 | 0.94 |
| | 100 | -0.003 | 0.002 | 0.95 |

Table 3: Simulation Results under 40% Censoring

| Parameter | n | Bias | MSE | CP (95%) |
|---|---|---|---|---|
| $\alpha$ | 30 | 0.038 | 0.031 | 0.91 |
| | 50 | 0.020 | 0.018 | 0.93 |
| | 100 | 0.009 | 0.007 | 0.94 |
| $\gamma$ | 30 | 0.131 | 0.218 | 0.89 |
| | 50 | 0.064 | 0.108 | 0.92 |
| | 100 | 0.025 | 0.042 | 0.94 |
| $\sigma$ | 30 | 0.028 | 0.021 | 0.92 |
| | 50 | 0.013 | 0.010 | 0.94 |
| | 100 | 0.006 | 0.004 | 0.95 |
| $\beta_1$ | 30 | 0.021 | 0.014 | 0.92 |
| | 50 | 0.009 | 0.006 | 0.94 |
| | 100 | 0.004 | 0.002 | 0.95 |
| $\beta_2$ | 30 | -0.026 | 0.015 | 0.90 |
| | 50 | -0.012 | 0.007 | 0.93 |
| | 100 | -0.005 | 0.003 | 0.94 |

The simulation results indicate that

- the MLEs are approximately unbiased.
- Bias and MSE decrease as sample size increases.
- Coverage probabilities approach the nominal 95% level for $n \geq 50$.
- Higher censoring increases variability, particularly for $\gamma$, but does not induce systematic bias.

## 5 Applications

Application of LGIGW regression model has been explored for two cases
(i) n =15 and
(ii) n =40.
(i) The bone marrow transplant censored data reported in Klein and Moeschberger (2003) are used to illustrate the applicability of the LGIGW regression model. The data set consists of survival times (in days) to death or relapse for 15 patients diagnosed with non-Hodgkin lymphoma (NHL) who received allogeneic (Allo) bone marrow transplants from HLA-matched sibling donors. The primary objective is to assess the effect of covariates on the logarithm of the survival times.

We fit the LGIGW regression model without intercept

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \sigma z_i, \quad i = 1, ..., 15,$$

where the errors $z_1, ..., z_{15}$ are independent random variables with density function given in (2.5).

MLEs of the parameters with their Standard Errors (SEs) are reported in Table 4 for LGIGW, LGIW and LIW regression models fitted to the data set. The p-values have also been reported for testing the significance of regression coefficients.

**Table 4: Estimates of the parameters and standard errors (SEs)**

| Model | estimates | SE | p-value |
|---|---|---|---|
| LGIGW | $\widehat{\alpha} = .1603$ | .0044 | - |
| | $\widehat{\gamma} = 1.2466$ | .7727 | - |
| | $\widehat{\sigma} = .5434$ | .0270 | - |
| | $\widehat{\beta_1} = .0581$ | .0001 | .000 |
| | $\widehat{\beta_2} = -.0236$ | .0002 | .000 |
| LGIW | $\widehat{\alpha} = 1.0000$ | - | - |
| | $\widehat{\gamma} = 3.1157$ | 9.9616 | - |
| | $\widehat{\sigma} = .4111$ | .0076 | - |
| | $\widehat{\beta_1} = .0546$ | .0000 | .000 |
| | $\widehat{\beta_2} = -.0218$ | .0001 | .000 |
| LIW | $\widehat{\alpha} = 1$ | - | - |
| | $\widehat{\gamma} = 1$ | - | - |
| | $\widehat{\sigma} = .4602$ | .0117 | - |
| | $\widehat{\beta_1} = .0598$ | .0000 | .000 |
| | $\widehat{\beta_2} = -.0153$ | .0000 | .000 |

The results presented in Table 4 provide important insights into the performance of the LGIGW regression model in comparison with its submodels. It is observed that the regression coefficients $\beta_1$ and $\beta_2$ are statistically significant at the 5% level across all models, indicating that the selected covariates have a significant impact on the logarithm of survival times. The parameter estimates under the LGIGW model appear stable with relatively small standard errors, suggesting reliable estimation. In particular, the additional shape parameter $\alpha$ in the LGIGW model enhances flexibility, allowing the model to capture more complex data structures as compared to the LGIW and LIW models.
The sample size (n = 15) is relatively small. Therefore, inference has to be interpreted cautiously. The asymptotic normal approximation may be less accurate in such small samples. However, the analysis is presented primarily to illustrate model applicability rather than to draw definitive clinical conclusions.
LGIGW regression model has been compared with LGIW and LIW regression models and AIC, BIC and AICc values for these models are provided in Table 5.

**Table 5: AIC, BIC and AICc for LGIGW, LGIW and LIW regression models**

| Model | AIC | BIC | AICc |
|---|---|---|---|
| LGIGW | 46.232 | 50.406 | 44.865 |
| LGIW | 112.466 | 115.299 | 110.133 |
| LIW | 92.904 | 95.028 | 90.504 |

It is seen that the values of AIC, BIC and AICc are lowest for LGIGW regression model which establishes the superiority of our regression model.

(ii) A dataset consisting of 40 financial loan records was analyzed, where the response variable represents time to default in months. Right censoring corresponds to loans that remained active at the end of the study. Two covariates were considered: normalized credit score and borrower risk category. The dataset exhibits skewness and heavy-tailed behavior, making it suitable for flexible parametric modeling. We fit the LGIGW regression model with intercept

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sigma z_i, \quad i = 1, ..., 40,$$

where the errors $z_1, ..., z_{40}$ are independent random variables with density function given in (2.5).
The LGIGW regression model provided the best fit in terms of AIC and BIC when compared with its submodels. Dataframe is
time$(t_i)$ = (4, 6, 7, 9, 11, 14, 18, 22, 5, 8, 10, 13, 17, 21, 26, 30, 6, 9, 12, 15, 19, 23, 28, 35, 7, 11, 14, 18, 24, 32, 40, 50, 8, 12, 16, 20, 27, 36, 45, 60),
status $(\delta_i)$ = (1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0),
$x_1$ (Credit Score) = (0.40, 0.50, 0.45, 0.55, 0.60, 0.62, 0.75, 0.80, 0.42, 0.48, 0.53, 0.58, 0.72, 0.78, 0.85, 0.88, 0.44, 0.49, 0.57, 0.61, 0.74, 0.79, 0.86, 0.90, 0.46, 0.52, 0.59, 0.73, 0.81, 0.87, 0.92, 0.95, 0.47, 0.54, 0.60, 0.76, 0.84, 0.91, 0.94, 0.97),
$x_2$ (Risk) = (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0)

Maximum likelihood estimates, standard errors, and p-values for LGIGW, LGIW, LIW, and Weibull regression models are presented in Table 6.

Table 6: Maximum likelihood estimates, standard errors, and p-values for LGIGW, LGIW, LIW, and Weibull regression models

| Model | Parameter | Estimate | Std. Error | p-value |
|---|---|---|---|---|
| LGIGW | $\beta_0$ | 1.245 | 0.312 | 0.0002 |
| | $\beta_1$ | -2.138 | 0.845 | 0.0115 |
| | $\beta_2$ | 0.965 | 0.402 | 0.0189 |
| | $\alpha$ | 1.872 | 0.556 | 0.0011 |
| | $\lambda$ | 0.743 | 0.221 | 0.0009 |
| | $\gamma$ | 1.356 | 0.478 | 0.0047 |
| LGIW | $\beta_0$ | 1.102 | 0.295 | 0.0005 |
| | $\beta_1$ | -1.845 | 0.801 | 0.0190 |
| | $\beta_2$ | 0.812 | 0.376 | 0.0302 |
| | $\alpha$ | 1.645 | 0.498 | 0.0023 |
| | $\gamma$ | 1.210 | 0.442 | 0.0068 |
| LIW | $\beta_0$ | 0.954 | 0.270 | 0.0012 |
| | $\beta_1$ | -1.502 | 0.756 | 0.0450 |
| | $\beta_2$ | 0.690 | 0.350 | 0.0485 |
| | $\gamma$ | 1.085 | 0.410 | 0.0095 |
| Weibull | $\beta_0$ | 0.875 | 0.260 | 0.0015 |
| | $\beta_1$ | -1.320 | 0.710 | 0.0620 |
| | $\beta_2$ | 0.580 | 0.330 | 0.0785 |
| | Shape | 1.950 | 0.520 | 0.0008 |

Note: Significant parameters (p-value $< 0.05$) indicate important covariate effects. The LGIGW model shows more statistically significant parameters, reflecting its flexibility and better fit.

The following table compares LGIGW, LGIW, LIW and Weibull regression models on basis of goodness-of-fit criteria.

Table 7: Comparison of LGIGW, LGIW, LIW and Weibull regression models based on goodness-of-fit criteria

| Model | Log-Likelihood | AIC | BIC |
|---|---|---|---|
| LGIGW | 13.88 | **-15.769** | **-7.786** |
| LGIW | 4.449 | 1.101 | 9.546 |
| LIW | 11.207 | -14.415 | -7.659 |
| Weibull | -30.532 | 69.064 | 75.819 |

Among the competing models, the LGIGW regression model provides the lowest AIC and BIC values, indicating superior goodness-of-fit. The Weibull model shows comparatively poorer performance, suggesting that it fails to capture the underlying heavy-tailed and skewed behaviour of the data. Since outliers in any data set can vary the results of analysis, hence an important step in the analysis of data is to detect the outliers and influential observations using sensitivity and residual analysis. For this, several approaches exist in literature given by Chatterjee and Hadi (1988); Cook (1977, 1986); Weisberg and Cook (1982).
In the following section, we briefly outline the methods used for sensitivity analysis and for deriving the necessary mathematical expressions.

# 6 Sensitivity Analysis

Different approaches for detecting the outliers are discussed in the following subsections.

## 6.1 Global Influence

The impact of removing the $i^{\text{th}}$ observation from the data set is assessed through the case-deletion approach, a widely used method in global influence analysis (Cook, 1977). The case-deletion model corresponding to model (3.1) is given by

$$y_j = \boldsymbol{x_j}^T\boldsymbol{\beta} + \sigma z_j, \quad j = 1, ..., n \text{ and } j \neq i. \tag{6.1}$$

The subscript 'i' indicates the original quantity but with deletion of $i^{th}$ observation. If the estimates are seriously affected by deleting an observation, then that observation needs attention.
If $\widehat{\boldsymbol{\theta}}_{(i)}$ : estimate of parameters obtained by deleting $i^{th}$ observation;
$\widehat{\boldsymbol{\theta}}$ : original estimate of parameters; then to measure the global influence,
Generalized Cook Distance is given by

$$GD_i(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}})^T \boldsymbol{J}(\widehat{\boldsymbol{\theta}})^{-1}(\widehat{\boldsymbol{\theta}}_{(i)} - \widehat{\boldsymbol{\theta}}), \tag{6.2}$$

Another measure to assess the difference between $\widehat{\boldsymbol{\theta}}_{(i)}$ and $\widehat{\boldsymbol{\theta}}$ is the likelihood distance given by

$$LD_i(\theta) = 2\{l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{(i)})\}, \tag{6.3}$$

where
$l(\widehat{\boldsymbol{\theta}})$ : original log-likelihood function;
$l(\widehat{\boldsymbol{\theta}}_{(i)})$ : log-likelihood function obtained after deleting $i^{th}$ observation.
If $\widehat{\boldsymbol{\theta}}_{(i)}$ is far from $\widehat{\boldsymbol{\theta}}$, then $i^{th}$ observation can be considered as an influential observation.

The values of $GD_i(\boldsymbol{\theta})$ and $LD_i(\boldsymbol{\theta})$ are given in Table 8.

**Table 8: Generalized Cook's Distances and Likelihood Distances**

| i | $GD_i(\boldsymbol{\theta})$ | $LD_i(\boldsymbol{\theta})$ |
|---|---|---|
| 1 | .1259 | -6.8353 |
| 2 | 13.1495 | 170.4717 |
| 3 | .7788 | -2.7328 |
| 4 | .2601 | -2.1485 |
| 5 | 13.9921 | 4.5275 |
| 6 | .0020 | -3.5609 |
| 7 | .0187 | -3.6358 |
| 8 | .0592 | .9726 |
| 9 | .0591 | 1.3945 |
| 10 | .0592 | .8294 |
| 11 | .0592 | .6291 |
| 12 | .0107 | -.2799 |
| 13 | .2870 | -.7998 |
| 14 | .0011 | -1.5854 |
| 15 | .0286 | .4901 |

The global influence measures as given in Table 8 clearly identify observations 2 and 5 as influential, as they exhibit substantially larger Generalized Cook's distances compared to other observations. The likelihood distance values further confirm this finding. The index plots of influence measures are shown in Figures 2 and 3.

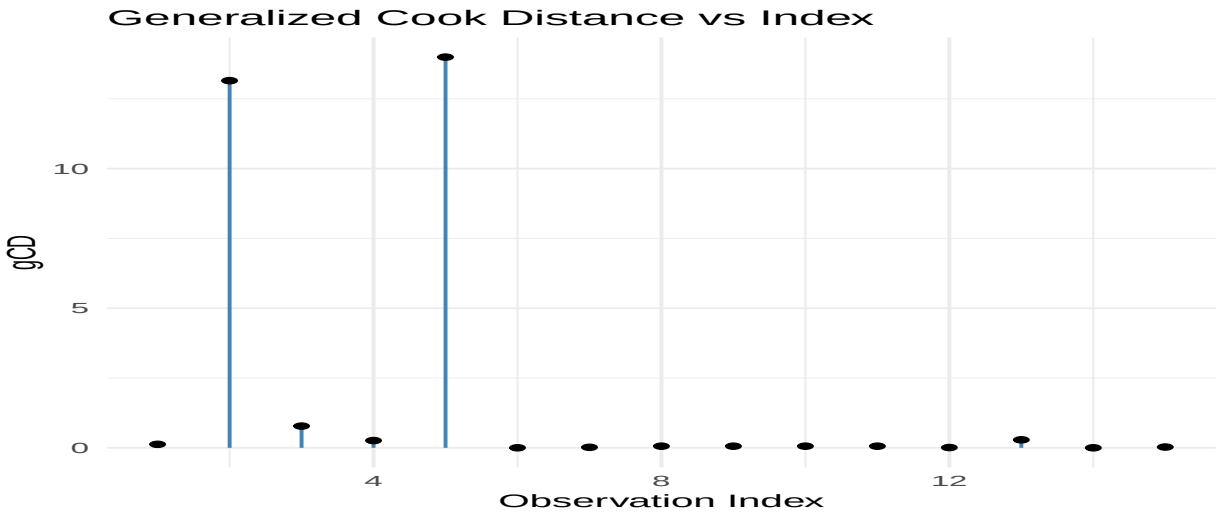


**Figure 2: Generalized Cook's Distance**

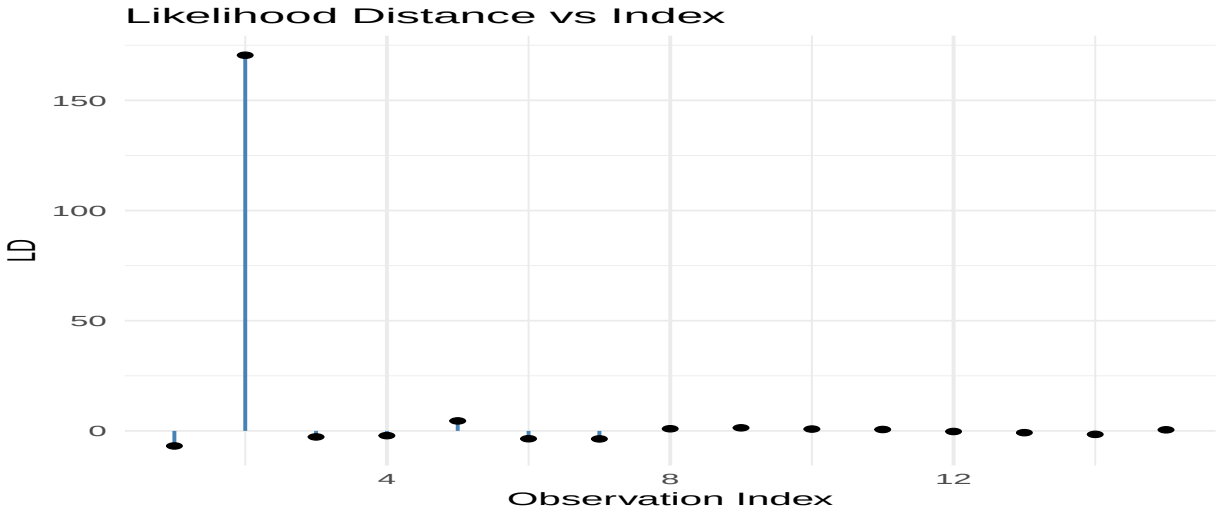


**Figure 3: Likelihood Distance**

The above figures also help in concluding that observations 2 and 5 are the influential observations.

## 6.2 Local Influence

An alternative approach to sensitivity analysis is the local influence method introduced by Cook (1986). This method involves assigning weights to observations rather than deleting them, and assessing their impact on the parameter estimates. The resulting framework is commonly referred to as a perturbed model. If $\widehat{\boldsymbol{\theta}}_w$ denotes the vector of estimates under perturbed model, then $LD(w) = 2\{l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_w)\}$ is the likelihood displacement. The normal curvature at direction $\boldsymbol{d}$ is given by

$$\boldsymbol{C_d}(\boldsymbol{\theta}) = 2|\boldsymbol{d}^T \boldsymbol{\Delta}^T J(\widehat{\boldsymbol{\theta}})^{-1} \boldsymbol{\Delta} \boldsymbol{d}| \tag{6.4}$$

such that $||\boldsymbol{d}|| = 1$ and $\boldsymbol{\Delta}$ is a (p+3)×n matrix. We can also calculate $\boldsymbol{C_d}(\alpha), \boldsymbol{C_d}(\gamma), \boldsymbol{C_d}(\sigma)$ and $\boldsymbol{C_d}(\beta)$. The elements of $\boldsymbol{\Delta}$ matrix are calculated under various perturbation schemes explained below. The largest eigen value of the matrix $\boldsymbol{B} = \boldsymbol{\Delta}^T \boldsymbol{J}(\widehat{\boldsymbol{\theta}})^{-1} \boldsymbol{\Delta}$ is $C_{d_{max}}$ where $\boldsymbol{d}_{max}$ is the eigen vector corresponding to $C_{d_{max}}$. The index plots of $\boldsymbol{d}_{max}, \boldsymbol{C_{d_i}}(\alpha), \boldsymbol{C_{d_i}}(\gamma), \boldsymbol{C_{d_i}}(\sigma)$ and $\boldsymbol{C_{d_i}}(\boldsymbol{\beta})$ together denote the total local influence Lesaffre and Verbeke (1998). Hence the curvature has the form $C_i = 2|\boldsymbol{\Delta}_i^T J(\widehat{\boldsymbol{\theta}})^{-1} \boldsymbol{\Delta}_i|$ at direction $\boldsymbol{d}$. For the vector of weights $\boldsymbol{w} = (w_1, ..., w_n)^T$, let the log-likelihood (3.4), under perturbed scheme be $l(\boldsymbol{\theta}|\boldsymbol{w})$. Then

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{ji})_{(p+3)\times n} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial \theta_j \partial w_i}, i = 1, ..., n,\ j = 1, ..., p+3 \tag{6.5}$$

is calculated under different perturbed schemes.

- **Case-weight Perturbation**: It is the basis for the study of influence. To perturb the contribution of each case, the vector of weights $\boldsymbol{w}$ is used where $w_j$=1 (inclusion) or $w_j$=0 (exclusion) of an observation from the estimation of $\boldsymbol{\theta}$. This method enables us to check the relative importance of observations in estimation.
  Under this scheme, the log-likelihood has the form

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{w}) = \{\log\alpha + \log\gamma - \log\sigma\} \sum_{i\in F} w_i - \sum_{i\in F} w_i z_i - \gamma \sum_{i\in F} w_i e^{-z_i} \\ + (\alpha - 1) \sum_{i\in F} w_i \log\left[1 - e^{\{-\gamma e^{-z_i}\}}\right] + \alpha \sum_{i\in C} w_i \log\left[1 - e^{\{-\gamma e^{-z_i}\}}\right], \end{aligned} \tag{6.6}$$

  where $z_i = (\frac{y_i - \boldsymbol{x_i}^T \boldsymbol{\beta}}{\sigma})$ and $0 \le w_i \le 1$.

  For $\widehat{z}_i = (\frac{y_i - \boldsymbol{x_i}^T \widehat{\boldsymbol{\beta}}}{\widehat{\sigma}})$, $\widehat{m}_i = e^{\{-\widehat{\gamma} e^{-\widehat{z}_i}\}}$, the elements of $\boldsymbol{\Delta}$ matrix are given in the Appendix.

- **Response Variable Perturbation**:
  Each $y_i$ is perturbed as $y_{iw} = y_i + w_i S_y$, where $S_y$ is standard deviation of elements contained in $\boldsymbol{Y}$ and $w_i \in \Re$. Under this scheme, the log-likelihood has the form

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{w}) = r\left\{\log\alpha + \log\gamma - \log\sigma\right\} - \sum_{i\in F} z_i^* - \gamma \sum_{i\in F} e^{-z_i^*} \\ + (\alpha-1)\sum_{i\in F}\log\left[1 - e^{\{-\gamma e^{-z_i^*}\}}\right] + \alpha\sum_{i\in C}\log\left[1 - e^{\{-\gamma e^{-z_i^*}\}}\right], \end{aligned} \tag{6.7}$$

  where $z_i^* = \dfrac{(y_i + w_i S_y) - \boldsymbol{x}_i{}^T\boldsymbol{\beta}}{\sigma}$.
  For $\widehat{z_i}^* = \dfrac{(y_i + w_i S_y) - \boldsymbol{x}_i{}^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}$, and $\widehat{m_i}^* = e^{\{-\widehat{\gamma}e^{-\widehat{z_i}^*}\}}$, the elements of $\boldsymbol{\Delta}$ matrix are given in the Appendix.
- **Explanatory Variable Perturbation**:
  On a particular explanatory variable, say $X_t$, we apply an additive perturbation scheme by writing $x_{itw} = x_{it} + w_i S_x$, where $S_x$ is the estimated standard deviation of perturbed explanatory variable and $w_i \in \Re$. The perturbed log-likelihood has the form

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{w}) = r\left\{\log\alpha + \log\gamma - \log\sigma\right\} - \sum_{i\in F} z_i^{**} - \gamma \sum_{i\in F} e^{-z_i^{**}} \\ + (\alpha-1)\sum_{i\in F}\log\left[1 - e^{\{-\gamma e^{-z_i^{**}}\}}\right] + \alpha\sum_{i\in C}\log\left[1 - \exp\left\{-\gamma e^{-z_i^{**}}\right\}\right], \end{aligned} \tag{6.8}$$

  where $z_i^{**} = \dfrac{y_i - \boldsymbol{x}_i^{*T}\boldsymbol{\beta}}{\sigma}$ and $\boldsymbol{x}_i^{*T}\boldsymbol{\beta} = \beta_1 + \beta_2 x_{i2} + ... + \beta_t(x_{it} + w_i S_x) + ... + \beta_p x_{ip}$.
  The elements of $\boldsymbol{\Delta}$ matrix are included in the Appendix.

By applying case weight perturbation for considered data set using LGIGW regresssion model, we obtain the value of $C_{d_{max}} = 11.823$. The plots of eight perturbation ,eigen vector corresponding to $d_{max}$ are shown in Figure 4 (a) and total influence $C_i$ are shown in Figure 4 (b).

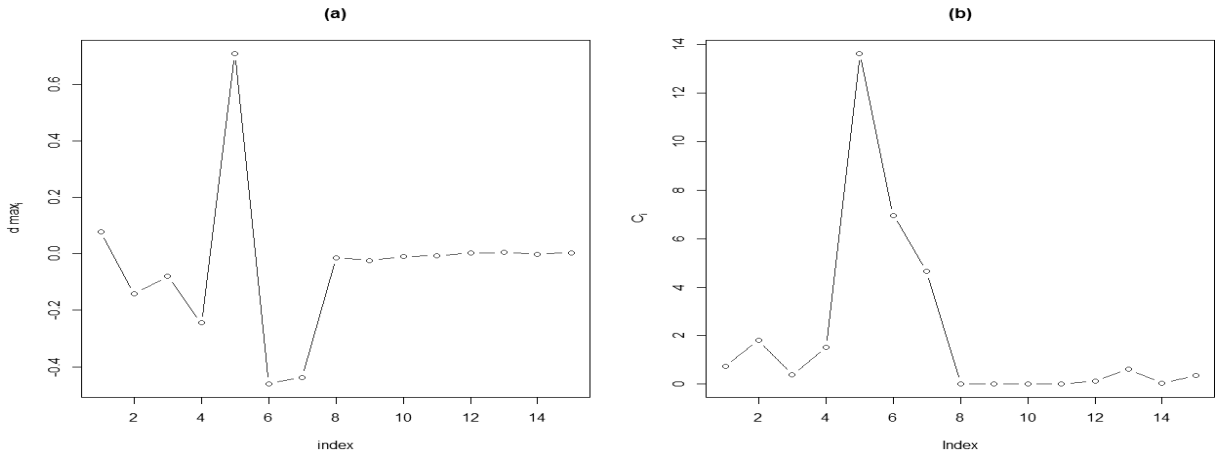


**Figure 4: Index plots of (a) $d_{max_i}(\boldsymbol{\theta})$ and (b) total local influence $C_i(\boldsymbol{\theta})$ for case-weight perturbation**

From Figures 4 (a) and 4 (b), it can be concluded that observations 2 and 5 are the influential observations. Using Response variable perturbation, maximum curvature $C_{d_{max}}$ is obtained as 3.971.

Figure 5(a) gives the plot of $d_{max}$ and it is observed that the observations 2 and 5 are far away from other observations. Total local influence $C_i$ versus observation index has been plotted in Figure 5(b) versus observation index.

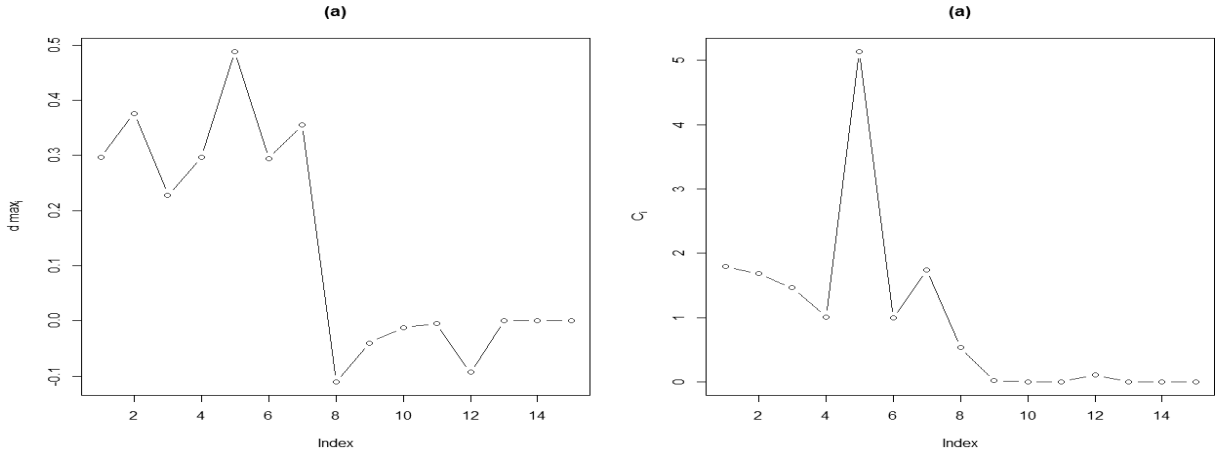


**Figure 5: Index plots of (a) $d_{max_i}(\boldsymbol{\theta})$ and (b) total local influence $C_i(\boldsymbol{\theta})$ under response variable perturbation**

Under local influence analysis, the maximum curvature values obtained from case-weight and response perturbation schemes indicate that observations 2 and 5 contribute disproportionately to parameter instability. However, subsequent refitting after removing observation 5 reveals only moderate relative changes in parameter estimates, indicating that the overall inference of the model remains stable. Thus, although influential, the detected observations do not materially alter substantive conclusions

# 7 Residual analysis

After fitting a model to the data, residual analysis (Collett (2015); Ortega et al. (2008)) is carried out to check the assumptions of the model, validity of the data, study the departures from error assumption and identify the outliers. Martingale residual was proposed by Barlow and Prentice (1988) and studied by Fleming and Harrington (2011). Therneau et al. (1990) introduced deviance residual for the Cox model with no time-dependent covariates. The two types of residuals are detailed in the following subsections.

## 7.1 Martingale residual

The residual for LGIGW regression model can be written as

$$r_{M_i} = \delta_i + \log\left(S(y_i; \widehat{\alpha}, \widehat{\gamma}, \widehat{\sigma}, \widehat{\boldsymbol{\beta}}^{\mathbf{T}})\right), \tag{7.1}$$

where

$$\delta_i = \begin{bmatrix} 1 & \text{if } i \in F, \\ 0 & \text{if } i \in C. \end{bmatrix}$$

For $\widehat{z}_i = (\dfrac{y_i - \boldsymbol{x}_i^T \widehat{\boldsymbol{\beta}}}{\widehat{\sigma}})$,

$$r_{M_i} = \begin{bmatrix} 1 + \widehat{\alpha}\left[1 - \exp\left\{-\widehat{\gamma}\exp\left(-\widehat{z}_i\right)\right\}\right] & \text{if } i \in F, \\ \widehat{\alpha}\left[1 - \exp\left\{-\widehat{\gamma}\exp\left(-\widehat{z}_i\right)\right\}\right] & \text{if } i \in C. \end{bmatrix} \tag{7.2}$$

The maximum value of residual $r_{M_i}$ is +1 and minimum is $-\infty$.

## 7.2 Modified deviance residual (martingale-type residual)

The transformation of martingale residual based on deviance has been used on LGIGW regression model to get a new residual, distributed symmetrically around zero. The modified deviance residual for LGIGW regression model can be written as

$$r_{D_i} = sgn(r_{M_i}) - 2[r_{M_i} + \delta_i \log\left(\delta_i - r_{M_i}\right)]^{1/2} = \begin{bmatrix} sgn(r_{M_i})\sqrt{-2[r_{M_i} + \log\left(1 - r_{M_i}\right)]} & \text{if } i \in F, \\ sgn(r_{M_i})\sqrt{-2r_{M_i}} & \text{if } i \in C. \end{bmatrix} \tag{7.3}$$

where $r_{M_i}$ is the martingale residual.

Martingale-type residuals $r_{D_i}$ versus fitted values have been plotted in Figure 6 for detection of the outliers.

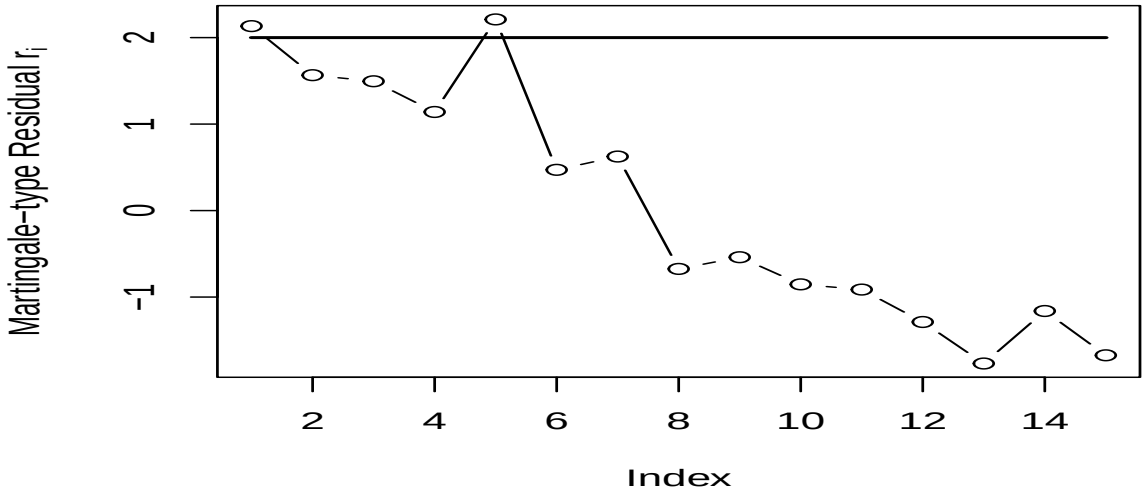


**Figure 6: Plot for martingale-type residual**

Apart from observations 1 and 5, all observations are lying in [-2,2], so we can conclude that model is fitting well to considered data set.

### 7.3 Impact of detected outliers

The relative change in estimate of a parameter can be calculated by using the formula

$$RC_{\boldsymbol{\theta_j}} = [(\widehat{\boldsymbol{\theta_j}} - \widehat{\boldsymbol{\theta}}_j(I))/\widehat{\boldsymbol{\theta_j}}] \tag{7.4}$$

where I: set of influential observations;
and $\widehat{\boldsymbol{\theta}}_j(I)$: the estimates of parameters obtained after removal of influential observations.
From sensitivity and residual analysis, we observe that observation 5 is the possible outlier. To evaluate the influence of this observation, the model was re-estimated after excluding it from the original data set. Table 9 reports the parameter estimates for both the complete data and the reduced data set obtained after removing the outlier. In addition, the relative percentage changes in the parameter estimates and their corresponding p-values are presented.

**Table 9: Parameter estimates, relative changes in parameters (in %) and p-values for regression coefficients**

| Observations | | $\widehat{\boldsymbol{\alpha}}$ | $\widehat{\boldsymbol{\gamma}}$ | $\widehat{\boldsymbol{\sigma}}$ | $\widehat{\boldsymbol{\beta}}_1$ | $\widehat{\boldsymbol{\beta}}_2$ |
|---|---|---|---|---|---|---|
| All | estimate | .16034 | 1.2466 | .5434 | .0581 | -.0236 |
| | RC (%) | - | - | - | - | - |
| | p-value | - | - | - | .000 | .000 |
| All-[#5] | estimate | .1167 | 5.6581 | .4962 | .0569 | -.0317 |
| | RC(%) | 27.245 | -353.885 | 8.691 | 2.179 | -34.349 |
| | p-value | - | - | - | .000 | .000 |

The results presented in Table 9 indicate that the parameter estimates of the LGIGW regression model are not highly sensitive to the deletion of the $5^{th}$ observation. This is evident from the fact that the removal of this potentially influential observation does not alter the statistical significance of the parameter estimates at the 5% level. Consequently, the overall inference drawn from the model remains unchanged for the data set under consideration. Table 8 further compares the parameter estimates obtained before and after the removal of the $5^{th}$ observation. Although the shape parameter $\gamma$ exhibits a relatively large percentage change, the regression

coefficients $\beta_1$ and $\beta_2$ show only moderate changes of 2.18% and 34.35%, respectively. Importantly, both coefficients remain statistically significant at the 5% level. It is worth noting that, while the present study demonstrates that the removal of the outlier does not substantially affect the model estimates for this particular data set. Different data sets especially those involving censored observations may exhibit greater sensitivity to such deletions. Thus, the analysis primarily illustrates the methodology for detecting influential observations and assessing relative percentage changes.

# 8 Conclusions

This paper introduced the Log Generalized Inverse Generalized Weibull (LGIGW) distribution and its associated regression model for censored lifetime data. The proposed model generalizes several well-known models including LGIW and LIW regression models, thereby offering a unified modeling framework. Theoretical properties of the distribution were derived, and maximum likelihood estimation was developed under Type I censoring. Asymptotic inference procedures, including confidence interval construction, were established using the observed information matrix. Through applications to a bone marrow transplant and financial data sets, the LGIGW regression model demonstrated superior performance compared to its submodels, as evidenced by lower AIC, BIC, and AICc values. Sensitivity and residual diagnostics revealed two influential observations. However, their removal did not materially affect the statistical significance of the regression coefficients, demonstrating robustness of the proposed model. Overall, the LGIGW regression model provides a flexible and powerful alternative for modeling censored lifetime data. Future research may consider Bayesian estimation, Progressive censoring schemes, Interval-censored extensions and Multivariate generalizations.

# Declarations

**Competing Interests:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
**Declaration of AI Use:** Not Applicable.
**Funding Statement:** The first author received grant provided by the University Grants Commission (UGC), Government of India.
**Data Availability Statement:** Data will be made available on request from the corresponding author.
**Ethical Approval & Consent to Participate:** Not Applicable.
**Informed Consent:** Informed consent from all participants was obtained. All participants were above the age of 18 years.
**Consent to Publish:** Consent to publish has been obtained.
**Clinical Trial Number:** Not Applicable.
**Author Contribution Statement:** Neetu Singla and Kanchan Jain, conceived the idea and designed the research; Analyzed interpreted the data and wrote the article.
**Acknowledgments:** The authors express their sincere gratitude to the worthy reviewers for their insightful comments and valuable suggestions, which have significantly improved the quality of the manuscript. The first author gratefully acknowledges the financial support provided by the University Grants Commission (UGC), Government of India.
**Disclaimer:** Clarifies that the views expressed in the article are those of the authors and do not necessarily reflect the official policy of their institutions or funding bodies.
**Availability of Materials and Code:** Not Applicable.

# References

Aljeddani, S. M. and Mohammed, M. (2023). Estimating the power generalized weibull distribution's parameters using three methods under type-ii censoring-scheme. *Alexandria Engineering Journal*, 67:219–228.

Alotaibi, R., Nassar, M., Rezk, H., and Elshahhat, A. (2022). Inferences and engineering applications of alpha power weibull distribution using progressive type-ii censoring. *Mathematics*, 10(16):2901.

Alsaggaf, I. A., Aloufi, S. F., and Baharith, L. A. (2024). A new generalization of the inverse generalized weibull distribution with different methods of estimation and applications in medicine and engineering. *Symmetry*, 16(8):1002.

Altun, E. (2021). The log-weighted exponential regression model: alternative to the beta regression model. *Communications in Statistics-Theory and Methods*, 50(10):2306–2321.

Barlow, W. E. and Prentice, R. L. (1988). Residuals for relative risk regression. *Biometrika*, 75(1):65–74.

Bui, L. N. and Ding, Q. (2025). Logistic regression modeling: methodological insights and roadmap. *Currents in Pharmacy Teaching and Learning*, 17(12):102460.

Cai, T. T., Guo, Z., and Ma, R. (2023). Statistical inference for high-dimensional generalized linear models with binary outcomes. *Journal of the american statistical association*, 118(542):1319–1332.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity analysis in linear regression*, volume 190. John Wiley & Sons.

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

Cook, R. (1986). Assessment of local influence. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(2):133–155.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.

De Gusmao, F. R., Ortega, E. M., and Cordeiro, G. M. (2011). The generalized inverse weibull distribution. *Statistical Papers*, 52(3):591–619.

El-Morshedy, M., Eliwa, M., El-Gohary, A., Almetwally, E. M., and El-Desokey, R. (2022). Exponentiated generalized inverse flexible weibull distribution: Bayesian and non-bayesian estimation under complete and type ii censored samples with applications. *Communications in mathematics and statistics*, 10(3):413–434.

Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.

Frolov, M., Tanchenko, S., and Ohluzdina, L. (2024). Parameter estimation of the weibull distribution in modeling the reliability of technical objects. *Journal of Engineering Sciences*, 11(1):A1–A10.

Fuetterer, C., Nalenz, M., Augustin, T., and Pfeiffer, R. M. (2025). A powerful penalized multinomial logistic regression approach: Cs fuetterer et al. *Computational Statistics*, 40(8):4565–4587.

Hashimoto, E. M., Ortega, E. M., Cancho, V. G., and Cordeiro, G. M. (2010). The log-exponentiated weibull regression model for interval-censored data. *Computational statistics & data analysis*, 54(4):1017–1035.

Jain, K., Singla, N., and Sharma, S. K. (2014). The generalized inverse generalized weibull distribution and its properties. *Journal of Probability*, 2014.

Keller, A. and Kamath, A. (1982). Alternate reliability models for mechanical systems. *Third International Conference on Reliability and Maintainability, France*.

Khaleeq, J., Amanullah, M., and Almaspoor, Z. (2021). Influence diagnostics in log-normal regression model with censored data. *Mathematical Problems in Engineering*, 2021(1):9612071.

Khan, J. A., Akbar, A., and Kibria, B. G. (2025). Influence of residuals on cook's distance for beta regression model: Simulation and application. *Hacettepe Journal of Mathematics and Statistics*, 54(2):618–632.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival analysis: techniques for censored and truncated data*, volume 1230. Springer.

Krishna, H., Dube, M., and Garg, R. (2019). Estimation of stress strength reliability of inverse weibull distribution under progressive first failure censoring. *Austrian Journal of Statistics*, 48(1):14–37.

Kubinec, R. (2023). Ordered beta regression: a parsimonious, well-fitting model for continuous data with lower and upper bounds. *Political analysis*, 31(4):519–536.

Kumar, C. S. and Nair, S. R. (2018). On log-inverse weibull distribution and its properties. *American Journal of Mathematical and Management Sciences*, 37(2):144–167.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, pages 570–582.

Nassar, M. and Abo-Kasem, O. (2017). Estimation of the inverse weibull parameters under adaptive type-ii progressive hybrid censoring scheme. *Journal of computational and applied mathematics*, 315:228–239.

Ortega, E. M., Cordeiro, G. M., and Carrasco, J. M. (2011). The log-generalized modified weibull regression model. *Brazilian Journal of Probability and Statistics*, 25(1):64–89.

Ortega, E. M., Paula, G. A., and Bolfarine, H. (2008). Deviance residuals in generalised log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation*, 78(8):747–764.

Pascoa, M. A., de Paiva, C. M., Cordeiro, G. M., and Ortega, E. M. (2013). The log-kumaraswamy generalized gamma regression model with application to chemical dependency data. *Journal of Data Science*, 11(4):781–818.

Pescim, R. R., Ortega, E. M., Cordeiro, G. M., Demtriod, C. G., and Hamedani, G. (2013). The log-beta generalized half-normal regression model. *Journal of Statistical Theory and Applications*, 12(4):330–347.

Qu, H. and Xie, F.-C. (2011). Diagnostics analysis for log-birnbaum–saunders regression models with censored data. *Statistica Neerlandica*, 65(1):1–21.

Rakshit, P., Wang, Z., Cai, T. T., and Guo, Z. (2021). Sihr: Statistical inference in high-dimensional linear and logistic regression models. *arXiv preprint arXiv:2109.03365*.

Singh, S. K., Singh, U., and Sharma, V. K. (2013). Bayesian prediction of future observations from inverse weibull distribution based on type-ii hybrid censored sample. *International Journal of Advanced Statistics and Probability*, 1(2):32–43.

Soale, A.-N. (2025). Detecting influential observations in single-index fréchet regression. *Technometrics*, 67(2):311–322.

Sultan, K., Alsadat, N., and Kundu, D. (2014). Bayesian and maximum likelihood estimations of the inverse weibull parameters under progressive type-ii censoring. *Journal of Statistical Computation and Simulation*, 84(10):2248–2265.

Tabassum, S. and Altaf, S. (2025). Addressing multicollinearity in log birnbaum–saunders regression model: a ridge regression estimation approach. *International Journal of Data Science and Analytics*, 20(8):6997–7008.

Temraz, N. S. Y. et al. (2022). Fuzzy stress-strength reliability subject to exponentiated power generalized weibull. *Engineering Letters*, 30(1):45–49.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160.

Weisberg, S. and Cook, R. D. (1982). Residuals and influence in regression. *Chapman and Hall*.

## Appendix

The expressions of elements of matrix $\boldsymbol{J}$ are given below:

$$\mathrm{J}_{\alpha,\alpha} = \frac{\partial^2 \mathrm{l}}{\partial \alpha^2} = \frac{-r}{\alpha^2};$$

$$\mathrm{J}_{\alpha,\gamma} = \frac{\partial^2 \mathrm{l}}{\partial \alpha \partial \gamma} = \sum_{i \in F} \frac{m_i e^{-z_i}}{[1-m_i]} + \sum_{i \in C} \frac{m_i e^{-z_i}}{[1-m_i]};$$

$$\mathrm{J}_{\alpha,\sigma} = \frac{\partial^2 \mathrm{l}}{\partial \alpha \partial \sigma} = \gamma \sum_{i \in F} \frac{m_i e^{-z_i}(\frac{z_i}{\sigma})}{[1-m_i]} + \gamma \sum_{i \in C} \frac{m_i e^{-z_i}(\frac{z_i}{\sigma})}{[1-m_i]};$$

$$\mathrm{J}_{\sigma,\sigma} = \frac{\partial^2 \mathrm{l}}{\partial \sigma^2} = \frac{r}{\sigma^2} - 2\sum_{i \in F} \frac{z_i}{\sigma^2} + \gamma \sum_{i \in F} e^{-z_i} \left[ -(\frac{z_i}{\sigma})^2 + 2\frac{z_i}{\sigma^2} \right]$$
$$+(\alpha-1) \sum_{i \in F} \left[ \frac{-\gamma^2 m_i^2 e^{(-2z_i)}(\frac{z_i}{\sigma})^2}{[1-m_i]^2} - \frac{\gamma m_i e^{-z_i}[2\frac{z_i}{\sigma^2} - (\frac{z_i}{\sigma})^2 + \gamma e^{-z_i}(\frac{z_i}{\sigma})^2]}{[1-m_i]} \right]$$
$$+\alpha \sum_{i \in C} \left[ \frac{-\gamma^2 m_i^2 e^{-2z_i}(\frac{z_i}{\sigma})^2}{[1-m_i]^2} - \frac{\gamma m_i e^{-z_i}[2\frac{z_i}{\sigma^2} - (\frac{z_i}{\sigma})^2 + \gamma e^{-z_i}(\frac{z_i}{\sigma})^2]}{[1-m_i]} \right];$$

$$\mathrm{J}_{\gamma,\gamma} = \frac{\partial^2 \mathrm{l}}{\partial \gamma^2} = \frac{-r}{\gamma^2} - (\alpha-1) \sum_{i \in F} \frac{m_i e^{-2z_i}}{[1-m_i]^2} - \alpha \sum_{i \in C} \frac{m_i e^{-2z_i}}{[1-m_i]^2};$$

$$\mathrm{J}_{\gamma,\sigma} = \frac{\partial^2 \mathrm{l}}{\partial \gamma \partial \sigma} = \sum_{i \in F} e^{-z_i}(\frac{-z_i}{\sigma}) + (\alpha-1) \sum_{i \in F} \left[ \frac{m_i^2 e^{-2z_i}\gamma(\frac{-z_i}{\sigma})}{[1-m_i]^2} + \frac{m_i \exp -z_i(\frac{-z_i}{\sigma})[1-\gamma e^{-z_i}]}{[1-m_i]} \right]$$
$$+\alpha \sum_{i \in C} \left[ \frac{m_i^2 \exp -2z_i \gamma(\frac{-z_i}{\sigma})}{[1-m_i]^2} + \frac{m_i e^{-z_i}(\frac{-z_i}{\sigma})[1-\gamma e^{-z_i}]}{[1-m_i]} \right].$$

For $j = 1, 2, ..., p$

$$\mathrm{J}_{\alpha,\beta_j} = \frac{\partial^2 \mathrm{l}}{\partial \alpha \partial \beta_\mathrm{j}} = \gamma \sum_{i \in F} \frac{m_i e^{-z_i}(\frac{x_{ij}}{\sigma})}{[1-m_i]} + \gamma \sum_{i \in C} \frac{m_i e^{-z_i}(\frac{x_{ij}}{\sigma})}{[1-m_i]};$$

$$\mathrm{J}_{\gamma,\beta_j} = \frac{\partial^2 \mathrm{l}}{\partial \gamma \partial \beta_\mathrm{j}} = \sum_{i \in F} e^{-z_i}(-\frac{x_{ij}}{\sigma}) + (\alpha-1) \sum_{i \in F} \left[ \frac{m_i^2 e^{-2z_i}\gamma(-\frac{x_{ij}}{\sigma})}{[1-m_i]^2} + \frac{m_i e^{-z_i}(\frac{x_{ij}}{\sigma})[1-\gamma e^{-z_i}]}{[1-m_i]} \right]$$
$$+\alpha \sum_{i \in C} \left[ \frac{m_i^2 e^{-2z_i}\gamma(-\frac{x_{ij}}{\sigma})}{[1-m_i]^2} + \frac{m_i e^{-z_i}(\frac{x_{ij}}{\sigma})[1-\gamma e^{-z_i}]}{[1-m_i]} \right];$$

$$\mathrm{J}_{\sigma,\beta_j} = \frac{\partial^2 \mathrm{l}}{\partial \sigma \partial \beta_\mathrm{j}} = -\sum_{i \in F} \frac{x_{ij}}{\sigma^2} + \gamma \sum_{i \in F} \left[ -e^{-z_i}(\frac{z_i}{\sigma})(\frac{x_{ij}}{\sigma}) + e^{-z_i}(\frac{x_{ij}}{\sigma^2}) \right]$$
$$+(\alpha-1) \sum_{i \in F} \left[ \frac{\gamma^2 m_i^2 e^{-2z_i}(\frac{x_{ij}}{\sigma})(\frac{z_i}{\sigma})}{[1-m_i]^2} - \frac{\gamma m_i e^{-z_i}[\frac{x_{ij}}{\sigma^2} - (\frac{x_{ij}}{\sigma})(\frac{z_i}{\sigma}) + \gamma e^{-z_i}](\frac{x_{ij}}{\sigma})(\frac{z_i}{\sigma})}{[1-m_i]} \right]$$
$$+\alpha \sum_{i \in C} \left[ \frac{\gamma^2 m_i^2 e^{-2z_i}(\frac{x_{ij}}{\sigma})(\frac{z_i}{\sigma})}{[1-m_i]^2} - \frac{\gamma m_i e^{-z_i}[\frac{x_{ij}}{\sigma^2} - (\frac{x_{ij}}{\sigma})(\frac{z_i}{\sigma}) + \gamma e^{-z_i}](\frac{x_{ij}}{\sigma})(\frac{z_i}{\sigma})}{[1-m_i]} \right];$$

For $j, s = 1, 2, ..., p$ and $j \neq s$

$$\mathrm{J}_{\beta_j,\beta_s} = \frac{\partial^2 \mathrm{l}}{\partial\beta_\mathrm{j}\partial\beta_\mathrm{s}} = -\gamma \sum_{i\in F}(\frac{x_{ij}}{\sigma})(\frac{x_{is}}{\sigma})e^{-z_i}$$

$$+(\alpha-1)\sum_{i\in F}\left[\frac{-\gamma^2 m_i^2 e^{-2z_i}(\frac{x_{ij}}{\sigma})(\frac{x_{is}}{\sigma})}{[1-m_i]^2} - \frac{\gamma m_i e^{-z_i}(\frac{x_{ij}}{\sigma})(\frac{x_{ij}}{\sigma})[\gamma e^{-z_i}-1]}{[1-m_i]}\right]$$

$$+\alpha\sum_{i\in C}\left[\frac{-\gamma^2 m_i^2 e^{-2z_i}(\frac{x_{ij}}{\sigma})(\frac{x_{is}}{\sigma})}{[1-m_i]^2} - \frac{\gamma m_i e^{-z_i}(\frac{x_{ij}}{\sigma})(\frac{x_{ij}}{\sigma})[\gamma e^{-z_i}-1]}{[1-m_i]}\right].$$

**Case Weight Perturbation**:

For $\widehat{z}_i = (\frac{y_i - \boldsymbol{x}_i{}^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}})$, and $\widehat{m}_i = e^{\{-\widehat{\gamma}e^{-\widehat{z}_i}\}}$, the elements of $\boldsymbol{\Delta}$ matrix are as given below:

$$\Delta_{1i} = \frac{\partial^2 l(\boldsymbol{\theta}|w)}{\partial\alpha\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \widehat{\alpha}^{-1} + \log[1-\widehat{m}_i] & \text{if } i \in F, \\ \log[1-\widehat{m}_i] & \text{if } i \in C. \end{cases}$$

$$\Delta_{2i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\gamma\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \widehat{\gamma}^{-1} - e^{-\widehat{z}_i} + \dfrac{(\widehat{\alpha}-1)\widehat{m}_i e^{-\widehat{z}_i}}{[1-\widehat{m}_i]} & \text{if } i \in F, \\ \dfrac{\widehat{\alpha}\widehat{m}_i e^{-\widehat{z}_i}}{[1-\widehat{m}_i]} & \text{if } i \in C. \end{cases}$$

$$\Delta_{3i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\sigma\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \widehat{\sigma}^{-1}[-1+\widehat{z}_i(1-\widehat{\gamma}e^{-\widehat{z}_i})] + \dfrac{(\widehat{\alpha}-1)\widehat{\gamma}\widehat{m}_i e^{-\widehat{z}_i}}{[1-\widehat{m}_i]} & \text{if } i \in F, \\ \dfrac{\widehat{\alpha}\widehat{\gamma}\widehat{z}_i\widehat{m}_i e^{-\widehat{z}_i}}{[1-\widehat{m}_i]} & \text{if } i \in C. \end{cases}$$

$$\Delta_{ji} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\beta_j\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},\ j = 4, ..., p+3$$

**Response Variable Perturbation**:

For $\widehat{z}_i{}^* = \dfrac{(y_i + w_i S_y) - \boldsymbol{x}_i{}^T\widehat{\boldsymbol{\beta}}}{\widehat{\sigma}}$, and $\widehat{m_i}{}^* = e^{\{-\widehat{\gamma}e^{-\widehat{z_i}^*}\}}$, the elements of $\boldsymbol{\Delta}$ matrix can be written as:

$$\Delta_{1i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\alpha\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \dfrac{-\widehat{\sigma}^{-1}\widehat{\gamma}S_y e^{-\widehat{z_i}^*}\widehat{m_i}^*}{[1-\widehat{m_i}^*]} & \text{if } i \in F, \\ \dfrac{-\widehat{\sigma}^{-1}\widehat{\gamma}S_y e^{-\widehat{z_i}^*}\widehat{m_i}^*}{[1-\widehat{m_i}^*]} & \text{if } i \in C. \end{cases}$$

$$\Delta_{2i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\gamma\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \widehat{\sigma}^{-1}S_y e^{-\widehat{z_i}^*}\left[1 + \dfrac{(\widehat{\alpha}-1)\widehat{m_i}^*[\{\widehat{\gamma}e^{-\widehat{z_i}^*}-1\}(1-\widehat{m_i}^*) + \widehat{m_i}^*\widehat{\gamma}e^{-\widehat{z_i}^*}]}{[1-\widehat{m_i}^*]^2}\right] & \text{if } i \in F, \\ \dfrac{\widehat{\alpha}\widehat{\sigma}^{-1}S_y e^{-\widehat{z_i}^*}\widehat{m_i}^*[\{\widehat{\gamma}e^{-\widehat{z_i}^*}-1\}(1-\widehat{m_i}^*) + \widehat{m_i}^*\widehat{\gamma}e^{-\widehat{z_i}^*}]}{[1-\widehat{m_i}^*]^2} & \text{if } i \in C. \end{cases}$$

$$\Delta_{3i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\sigma\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \widehat{\sigma}^{-2}S_y[1+\widehat{\gamma}\widehat{z_i}^* e^{-\widehat{z_i}^*} - \widehat{\gamma}e^{-\widehat{z_i}^*}] \\ \quad + \dfrac{(\widehat{\alpha}-1)\widehat{\gamma}\widehat{\sigma}^{-2}S_y\widehat{m_i}^* e^{-\widehat{z_i}^*}[1-\widehat{z_i}^* + \widehat{\gamma}\widehat{z_i}^* e^{-\widehat{z_i}^*}]}{[1-\widehat{m_i}^*]} & \text{if } i \in F, \\ \quad + \dfrac{(\widehat{\alpha}-1)\widehat{\gamma}^2\widehat{\sigma}^{-2}S_y\widehat{z_i}^*(\widehat{m_i}^*)^2 e^{-2\widehat{z_i}^*}}{[1-\widehat{m_i}^*]^2} \\ \dfrac{\widehat{\alpha}\widehat{\gamma}\widehat{\sigma}^{-2}S_y\widehat{m_i}^* e^{-\widehat{z_i}^*}[1-\widehat{z_i}^* + \widehat{\gamma}\widehat{z_i}^* e^{-\widehat{z_i}^*}]}{[1-\widehat{m_i}^*]} \\ \quad + \dfrac{\widehat{\alpha}\widehat{\gamma}^2\widehat{\sigma}^{-2}S_y\widehat{z_i}^*(\widehat{m_i}^*)^2 e^{-2\widehat{z_i}^*}}{[1-\widehat{m_i}^*]^2} & \text{if } i \in C. \end{cases}$$

$$\Delta_{ji} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\beta_j\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},\ j = 4, ..., p+3$$

$$= \begin{cases} x_{ij}\widehat{\sigma}^{-2}\widehat{\gamma}S_y e^{\widehat{z_i}^*} \\ \quad + \dfrac{(\widehat{\alpha}-1)\widehat{\sigma}^{-2}S_y x_{ij}\widehat{\gamma}\widehat{m_i}^* e^{-\widehat{z_i}^*}[\widehat{\gamma}e^{-\widehat{z_i}^*}-1]}{[1-\widehat{m_i}^*]} & \text{if } i \in F, \\ \quad + \dfrac{(\widehat{\alpha}-1)\widehat{\sigma}^{-2}S_y x_{ij}\widehat{\gamma}^2(\widehat{m_i}^*)^2 e^{-2\widehat{z_i}^*}}{[1-\widehat{m_i}^*]^2} \\ \dfrac{\widehat{\alpha}\widehat{\sigma}^{-2}S_y x_{ij}\widehat{\gamma}\widehat{m_i}^* e^{-\widehat{z_i}^*}[\widehat{\gamma}e^{-\widehat{z_i}^*}-1]}{[1-\widehat{m_i}^*]} \\ \quad + \dfrac{\widehat{\alpha}\widehat{\sigma}^{-2}S_y x_{ij}\widehat{\gamma}^2(\widehat{m_i}^*)^2 e^{-2\widehat{z_i}^*}}{[1-\widehat{m_i}^*]^2} & \text{if } i \in C. \end{cases}$$

**Explanatory Variable Perturbation**: If $\widehat{z_i}^{**} = \dfrac{(y_i - \boldsymbol{x_i}^T\widehat{\boldsymbol{\beta}})}{\widehat{\sigma}}$, and $\widehat{m_i}^{**} = e^{\{-\widehat{\gamma}e^{-\widehat{z_i}^{**}}\}}$, then the elements of $\boldsymbol{\Delta}$ matrix can be written in the following forms:

$$\Delta_{1i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\alpha\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \dfrac{-\widehat{\sigma}^{-1}\widehat{\gamma}S_x\beta_t e^{-\widehat{z_i}^{**}}\widehat{m_i}^{**}}{[1-\widehat{m_i}^{**}]} & \text{if } i \in F, \\ \dfrac{-\widehat{\sigma}^{-1}\widehat{\gamma}S_x\beta_t e^{-\widehat{z_i}^{**}}\widehat{m_i}^{**}}{[1-\widehat{m_i}^{**}]} & \text{if } i \in C. \end{cases}$$

$$\Delta_{2i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\gamma\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} -\widehat{\sigma}^{-1}S_x\beta_t e^{-\widehat{z_i}^{**}}\left[1 + \dfrac{(\widehat{\alpha}-1)\widehat{m_i}^{**}[\{\widehat{\gamma}e^{-\widehat{z_i}^{**}}-1\}(1-\widehat{m_i}^{**}) + \widehat{m_i}^{**}\widehat{\gamma}e^{-\widehat{z_i}^{**}}]}{[1-\widehat{m_i}^{**}]^2}\right] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } i \in F, \\ \dfrac{-\widehat{\sigma}^{-1}S_x\beta_t e^{-\widehat{z_i}^{**}}\widehat{\alpha}\widehat{m_i}^{**}[\{\widehat{\gamma}e^{-\widehat{z_i}^{**}}-1\}(1-\widehat{m_i}^{**}) + \widehat{m_i}^{**}\widehat{\gamma}e^{-\widehat{z_i}^{**}}]}{[1-\widehat{m_i}^{**}]^2} & \text{if } i \in C. \end{cases}$$

$$\Delta_{3i} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\sigma\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} \widehat{\sigma}^{-2}S_x[-1-\widehat{\gamma}\widehat{z_i}^{**}e^{-\widehat{z_i}^{**}}+\widehat{\gamma}e^{-\widehat{z_i}^{**}}] \\ -\dfrac{(\widehat{\alpha}-1)\widehat{\gamma}\widehat{\sigma}^{-2}S_x\widehat{m_i}^{**}e^{-\widehat{z_i}^{**}}[1-\widehat{z_i}^{**}+\widehat{\gamma}\widehat{z_i}^{**}e^{-\widehat{z_i}^{**}}]}{[1-\widehat{m_i}^{**}]} & \text{if } i\in F, \\ +\dfrac{(\widehat{\alpha}-1)\widehat{\gamma}^2\widehat{\sigma}^{-2}S_x\widehat{z_i}^{*}(\widehat{m_i}^{**})^2e^{-2\widehat{z_i}^{**}}}{[1-\widehat{m_i}^{**}]^2} \\ \dfrac{-\widehat{\alpha}\widehat{\gamma}\widehat{\sigma}^{-2}S_x\widehat{m_i}^{**}e^{-\widehat{z_i}^{**}}[1-\widehat{z_i}^{**}+\widehat{\gamma}\widehat{z_i}^{**}e^{-\widehat{z_i}^{**}}]}{[1-\widehat{m_i}^{**}]} \\ +\dfrac{\widehat{\alpha}\widehat{\gamma}^2\widehat{\sigma}^{-2}S_x\widehat{z_i}^{**}(\widehat{m_i}^{**})^2e^{-2\widehat{z_i}^{**}}}{[1-\widehat{m_i}^{**}]^2} & \text{if } i\in C. \end{cases}$$

$$\Delta_{ji} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\beta_j\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}},\ j=4,...,p+3\ and\ j\neq t,$$

$$= \begin{cases} -x_{ij}\widehat{\sigma}^{-2}\widehat{\gamma}S_x\widehat{\beta}_te^{-\widehat{z_i}^{**}} \\ -\dfrac{(\widehat{\alpha}-1)\widehat{\gamma}\widehat{\sigma}^{-2}S_xx_{ij}\widehat{\beta}_t\widehat{m_i}^{**}e^{-\widehat{z_i}^{**}}[-1+\widehat{\gamma}e^{-\widehat{z_i}^{**}}]}{[1-\widehat{m_i}^{**}]} & \text{if } i\in F, \\ -\dfrac{(\widehat{\alpha}-1)\widehat{\gamma}^2\widehat{\sigma}^{-2}S_xx_{ij}\widehat{\beta}_t(\widehat{m_i}^{**})^2\exp(-2\widehat{z_i}^{**})}{[1-\widehat{m_i}^{**}]^2} \\ -\dfrac{\widehat{\alpha}\widehat{\gamma}\widehat{\sigma}^{-2}S_xx_{ij}\widehat{\beta}_t\widehat{m_i}^{**}e^{-\widehat{z_i}^{**}}[-1+\widehat{\gamma}e^{-\widehat{z_i}^{**}}]}{[1-\widehat{m_i}^{**}]} \\ -\dfrac{\widehat{\alpha}\widehat{\gamma}^2\widehat{\sigma}^{-2}S_xx_{ij}\widehat{\beta}_t(\widehat{m_i}^{**})^2e^{-2\widehat{z_i}^{**}}}{[1-\widehat{m_i}^{**}]^2} & \text{if } i\in C. \end{cases}$$

and

$$\Delta_{ti} = \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{w})}{\partial\beta_t\partial w_i}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$$

$$= \begin{cases} S_x\widehat{\sigma}^{-1}[1-\widehat{\gamma}\widehat{\sigma}^{-1}\widehat{\beta}_t(x_{it}+w_iS_x)e^{-\widehat{z_i}^{**}}-\widehat{\gamma}e^{-\widehat{z_i}^{**}}] \\ +\dfrac{(\widehat{\alpha}-1)\widehat{\gamma}\widehat{\sigma}^{-1}S_x\widehat{m_i}^{**}e^{-\widehat{z_i}^{**}}[1+\widehat{\beta}_t(\frac{x_{it}+w_iS_x}{\sigma})\{1-\widehat{\gamma}e^{-\widehat{z_i}^{**}}\}]}{[1-\widehat{m_i}^{**}]} & \text{if } i\in F, \\ -\dfrac{(\widehat{\alpha}-1)\widehat{\gamma}^2\widehat{\sigma}^{-2}\widehat{\beta}_tS_x(\widehat{m_i}^{**})^2e^{-2\widehat{z_i}^{**}}(x_{it}+w_iS_x)}{[1-\widehat{m_i}^{**}]^2} \\ \dfrac{\widehat{\alpha}\widehat{\gamma}\widehat{\sigma}^{-1}S_x\widehat{m_i}^{**}e^{-\widehat{z_i}^{**}}[1+\widehat{\beta}_t(\frac{x_{it}+w_iS_x}{\sigma})\{1-\widehat{\gamma}e^{-\widehat{z_i}^{**}}\}]}{[1-\widehat{m_i}^{**}]} \\ -\dfrac{\widehat{\alpha}\widehat{\gamma}^2\widehat{\sigma}^{-2}\widehat{\beta}_tS_x(\widehat{m_i}^{**})^2e^{-2\widehat{z_i}^{**}}(x_{it}+w_iS_x)}{[1-\widehat{m_i}^{**}]^2} & \text{if } i\in C. \end{cases}$$

## Author(s) Bio / Authors' Note

- **Dr. Neetu Singla** is a Data Scientist with over 7 years of experience in machine learning, statistical analysis, and data visualization. She holds a Ph.D. in Statistics from Panjab University, Chandigarh and has led impactful projects in anomaly detection, server capacity planning, and price optimization at Altimetrik. Her career spans both industry and academia, including roles in building recommender systems, predictive models, and mentoring data science students. With multiple international publications and certifications in Statistics, ML, Tableau, and Dataiku, she combines deep statistical expertise with practical problem-solving to deliver innovative, data-driven solutions.

- **Prof. Kanchan Jain** retired from Department of Statistics, Panjab University, Chandigarh (India) in September, 2024. She is fellow of Royal Statistical Society, London and elected member of International Statistical Institute, Netherlands and National Science Academy, Allahabad. She had been a teacher and researcher since August, 1980. The areas of her specialization include Probability, Distribution Theory, Reliability Theory and Actuarial Statistics. She has to her credit a large number of research publications in highly reputed journals and has supervised sixteen candidates towards their Ph.D. degree. She has delivered a number of invited talks at international and national level.

SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 2, June 2026
DOI: https://doi.org/10.64060/JASR2V2i3


# On New Quantitative Randomized Response Techniques and a Weighted Privacy–Efficiency Measure

Shoaib Iqbal [1], Zawar Hussain [1*]

[1]Department of Statistics, Faculty of Computing, The Islamia University of Bahawalpur 63100, Pakistan

* Corresponding Email: zhlangah@yahoo.com

Received: 03 March 2026 / Revised: 01 May 2026 / Accepted: 05 May 2026 / Published online: 15 May 2026



ABSTRACT

Randomized response techniques are widely used for collecting information on sensitive variables while preserving respondent privacy. In the context of quantitative data, existing models are typically based on either additive or multiplicative scrambling mechanisms, which limit their flexibility in balancing estimation efficiency and privacy protection. This paper proposes a generalized class of quantitative randomized response models that integrates both additive and multiplicative components within a unified framework. From this formulation, three specific models ($M_1$ - $M_3$) are developed, each representing a distinct configuration of the scrambling process and corresponding to different efficiency–privacy trade-offs. In addition, a weighted measure combining efficiency and privacy is introduced to facilitate a unified comparison of competing models. The proposed measure provides a flexible and interpretable criterion for model evaluation and supports consistent ranking under varying preference weights. Theoretical properties of the estimators are derived, and comparative analysis demonstrates the improved performance of the proposed models over existing techniques under a range of parameter settings. The proposed framework enhances the adaptability of randomized response methods and offers a practical tool for selecting appropriate models in sensitive survey applications.

**Keywords**:Randomized response models; Scrambling models; Estimation of mean; Relative efficiency; Privacy protection

## 1. Introduction

Amongst the indirect questioning techniques including item count technique by Miller [1], Nominative technique by Sirken [2], List experiment by Raghavarao and Federer [3], the randomized response technique by Warner [4] have long been recognized as an effective tool for collecting reliable data on sensitive characteristics while preserving respondent confidentiality. Since the seminal contribution of Warner [4], a wide range of models have been developed to address both qualitative and quantitative survey settings. Some of the important developments may include Warner [5], Greenberg et al. [6], Horvitz et al. [7], Kuk [8], Mangat [9], Christofides [10], Kim and Warde [11], Arnab and Singh [12], Gupta et al. [13], Eichhorn and Hayre [14]. A detailed review on qualitative and quantitative randomized response models is available in Lee et al [15]. The interested readers may refer to Lee at al. [15] and the references cited therein.

In the context of quantitative variables, these techniques aim to estimate population parameters, particularly the mean, without requiring respondents to disclose their true values directly. Most existing quantitative RRT models are built upon either additive or multiplicative scrambling mechanisms. Additive approaches introduce random noise independent of the sensitive variable, whereas multiplicative schemes distort the reported value through scaling. While both strategies offer a degree of privacy protection, they often involve a delicate trade-off between estimation efficiency and confidentiality, and their structural forms remain somewhat restrictive. In particular, the literature lacks a comprehensive framework that systematically integrates these two mechanisms.

In addition to model development, considerable attention has been given to the evaluation of randomized response techniques in terms of efficiency and privacy. Traditional comparisons are often based on variance as a measure of efficiency, while privacy protection is assessed through various ad hoc or model-specific indices. More recently, Yan et al.

[16], Gupta et al. [17] and Azeem [18] measures have been proposed, which allow the incorporation of user-defined preferences. Despite their usefulness, these measures exhibit certain limitations. In particular, they often lack analytical transparency, making it difficult to characterize conditions under which one model dominates another. Moreover, the resulting rankings may not be stable across different weight choices, which complicates interpretation in practical applications. These limitations highlight the need for a more robust and interpretable combined measure that can support consistent decision-making.

The present study addresses these limitations by proposing a generalized class of quantitative randomized response models that combines additive and multiplicative scrambling within a unified formulation. This perspective allows the construction of multiple model configurations under a common structure. In particular, three specific models ( $M_1 - M_3$ ) are developed, each representing a distinct balance between privacy protection and estimation efficiency. Rather than being arbitrary variations, these models correspond to different operational regimes within the proposed framework.

In addition, this paper introduces a weighted measure of privacy and efficiency, denoted by $\psi$ , which provides a unified basis for comparing competing models. Unlike existing measures, the proposed formulation facilitates a more transparent interpretation of trade-offs and supports consistent ranking across different preference settings. The analytical properties of this measure are also examined to highlight its practical relevance.

Overall, this study contributes to the literature by developing a flexible modeling framework and a coherent evaluation criterion for quantitative randomized response techniques. The proposed approach not only clarifies the relationships among different scrambling strategies but also enhances the decision-making process in selecting an appropriate model for sensitive surveys.

The remainder of the paper is organized as follows. Section 2 introduces the generalized scrambling framework and presents the proposed models. Section 3 derives the estimators and their properties. Section 4 discusses the weighted measure and its characteristics. Section 5 provides comparative analysis, and Section 6 concludes the paper.

## 02. Generalized Scrambling Framework

In this section, we introduce a generalized framework for constructing quantitative randomized response models that integrates both additive and multiplicative scrambling mechanisms.

Let $Y_i$ denote the true value of a sensitive quantitative variable for the $i^{th}$ respondent, with mean $\mu_y = E(Y_i)$ and variance $\sigma_y^2 = Var(Y_i)$. To protect respondent privacy, the observed response is generated through a randomized mechanism involving independent scrambling variable. Let $S_i$ be a scrambling variable such that $E(S_i) = \mu_s = 0$, $Var(S_i) = \sigma_s^2$. Assume that $S_i$ and $Y_i$ are mutually independent. We define the generalized scrambling mechanism as:

$$R_i = Y_i + C_i Y_i S_i, \tag{1}$$

where $C_i$ is a random variable representing random coefficient mechanism. Using the assumptions above, and applying expectation on (1), we obtain the expected response as:

$$E(R_i) = E(Y_i + C_i Y_i S_i) = \mu_y + E(C_i)\mu_y \mu_s$$

$$E(R_i) = \mu_y + E(C_i)\mu_y (0) = \mu_y. \tag{2}$$

Using (2), the unbiased moment estimator of $\mu_y$ may be defined as

$$\hat{\mu}_y = \bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i . \tag{3}$$

The variance of the moment estimator $\hat{\mu}_y$, in (3), may be derived as follows. From (1), we have

$$Var(R_i) = Var(Y_i + C_i Y_i S_i) = Var(Y_i) + Var(C_i Y_i S_i) + 2Cov(Y_i, C_i Y_i S_i).$$

Since $E(S_i) = 0$, we have $Cov(Y_i, Y_i S_i) = 0$. This implies $Cov(Y_i, C_i Y_i S_i) = 0$. Thus, we get

$$Var(R_i) = \sigma_y^2 + Var(C_i Y_i S_i). \tag{4}$$

Now

$$Var(C_i Y_i S_i) = E(C_i^2) Var(Y_i S_i) = E(C_i^2) E(Y_i^2 S_i^2) = E(C_i^2) E(Y_i^2) E(S_i^2).$$

$$Var(C_i Y_i S_i) = E(C_i^2)(\mu_y^2 + \sigma_y^2)\sigma_s^2.$$

Substituting the above result in (4), we get

$$Var\left(R_i\right)=\sigma_y^2+E\left(C_i^2\right)\left(\mu_y^2+\sigma_y^2\right)\sigma_s^2. \tag{5}$$

Finally,

$$Var\left(\hat{\mu}_y\right)=\frac{1}{n}Var\left(R_i\right)=\frac{1}{n}\left\{\sigma_y^2+E\left(C_i^2\right)\left(\mu_y^2+\sigma_y^2\right)\sigma_s^2\right\}. \tag{6}$$

Also note that $Var\left(\hat{\mu}_y\right)=\frac{1}{n}\left\{\sigma_y^2+E\left(C_i^2\right)\left(\mu_y^2+\sigma_y^2\right)\sigma_s^2\right\}=\frac{1}{n}\left\{\sigma_y^2+KE\left(C_i^2\right)\right\}$, where $K=\left(\mu_y^2+\sigma_y^2\right)\sigma_s^2$. Thus, the efficiency of the estimator depends entirely on $E\left(C_i^2\right)$. The estimator $\bar{R}$ is more efficient when $E\left(C_i^2\right)$ is minimized. To quantify the privacy protection, we define Yan et al. [16] privacy measure as

$$\nabla=E\left(R_i-Y_i\right)^2=E\left(C_iY_iS_i\right)=E\left(C_i^2\right)\left(\mu_y^2+\sigma_y^2\right)\sigma_s^2. \tag{7}$$

So, from (5) and (7), we can write

$$Var\left(R_i\right)=\sigma_y^2+\nabla. \tag{8}$$

The expression in (8) shows that an increase in privacy (higher $\nabla$) leads to a proportional increase in variance. Hence, there exists an inherent efficiency–privacy trade-off governed by $E\left(C_i^2\right)$. Three special cases of this generalized framework will be discussed in section 5.

## 3. Some Existing Comparison Measures

This section discusses the available measures of privacy and efficiency commonly used for comparing quantitative randomized response models.

### *3.1 The Relative Efficiency Measure*

Relative efficiency is one of the most widely used methods for performance comparison and has been applied by nearly all researchers in evaluating randomized response models. Let Model *S* be the model of interest and Model *O* is the reference model used for comparison. Denote the mean squared error (*MSE*) of the mean estimator under Model *S* and Model *O* as $MSE_S$ and $MSE_O$, respectively. The relative efficiency is then calculated as:

$$E=\frac{MSE_O}{MSE_s}.$$

Since the MSE is always positive, $E\geq 0$. If $E\geq 1$, it indicates that Model *S* is more efficient than Model *O*.

### *3.2 The Yan et al. [16] Measure of Privacy*

Yan et al. [16] proposed a metric to assess the privacy level of a given quantitative model, denoted by $\nabla$ and defined as:

$$\nabla=\left(Z-Y\right)^2$$

For any randomized response model, $\nabla\geq 0$, with higher values indicating greater privacy protection. The metric developed by Yan et al. [16] emphasizes privacy protection exclusively.

### *3.3 The Gupta et al. [17] Unified Measure of privacy and efficiency*

Separate measures for quantifying privacy level and efficiency have been used by different authors prior to the publication of the work by Gupta et al. [17]. To make simultaneous use of privacy and efficiency, Gupta et al. [17] proposed a combined metric that integrates both privacy and efficiency, as outlined below:

$$\delta=\frac{MSE}{\nabla}.$$

The value of $\delta$ ranges from 0 to ∞. A smaller δ reflects lower mean squared error, a higher privacy level, or both. This measure provides an overall assessment of a randomized response model as a single δ value. However, it does not assign relative weights to privacy or efficiency, which can limit its usefulness.

### *3.4 The Azeem [18] Weighted Privacy Measure*

Let $\nabla_S$ and $\nabla_O$ be the Yan et al. [16] measure of privacy protection for Model *S* and Model *O*, respectively. Then define

$$P=\frac{\nabla_S}{\nabla_O}.$$

Since $\nabla_S$ is in the numerator, so a higher value of $P$ indicates higher privacy protection for Model *S* than Model *O*. If the two models under consideration offer equal privacy protection, then $P=1$. Azeem [18] defined weighted privacy measure as

$$\phi = \left( \frac{w_1 E + w_2 P}{w_1 + w_2} \right). \tag{9}$$

The proposed measure, $\phi$, given in (9), ranges from 0 to ∞. A $\phi$ value greater than 1 signifies that Model *S* outperforms Model *O*. Conversely, if $0 < \phi < 1$, it indicates that Model *S* is less effective than Model *O* overall.

## 4. The Proposed Weighted Measure of Privacy

In an effort to overcome the inadequacies of existing measures, we introduce the weighted measure, denoted by $\psi$, as:

$$\psi = P^{w_1} \times E^{w_2}, \tag{10}$$

where $E$ and $P$ are as defined above and $w_1, w_2$ are non-negative weights such that $w_1 + w_2 = 1$.

The weighted measure defined in (10), is an inclusive and unified way of evaluating the randomized response models based on whether they combine efficiency and privacy as a single factor. Unlike the existing measures, which either used only efficiency or only privacy, the proposed measure strikes a balance between the two based on its definition in terms of being a weighted geometric mean of the two. Another striking strength of the proposed measure is its ease of reading: the sign of $\log(\psi)$ clearly indicates whether the proposed model provides improvement, equal, or worse performance than the competing model. The $\psi$ measure not only generalizes earlier approaches but also extends them, delivering a more powerful, flexible, and helpful practical tool for the comparative evaluation of randomized response models. Moreover, the $\psi$ measure is a versatile measure. With the use of appropriate weights given to efficiency and privacy, researchers can choose to use the measure for purposes specific to their research. For a survey dealing with very sensitive information, for instance, more importance can be attributed to privacy but, for reliability or industrial purposes, efficiency takes precedence. This weighting flexibility guarantees that the $\psi$ measure is highly flexible and applicable in a wide range of research environments, thus guaranteeing its utility in theoretical work as well as in practice.

The proposed $\psi$ measure has several distinct advantages:

1. *Balanced consideration*: Unlike previous methods, which focused on either relative efficiency *E* in isolation or privacy *P*, the $\psi$ measure addresses them at the same time, resulting in balanced evaluation.
2. *Flexibility*: The researchers can tailor the measurement by altering $w_1$ and $w_2$. For instance, in medical questionnaires, privacy can be weighted heavily, and in industrial reliability, efficiency.
3. *Geometric mean formula* – The use of a geometric mean avoids either dimension from winning out and instead captures a natural trade-off between efficiency and privacy.
4. *Interpretability*: The measurement is simple to interpret as $\log(\psi) > 0$ indicates Model *S* dominates Model *O*. This can be seen as $\log(\psi) = 0$ suggests that the competing models are equally good and $\log(\psi) < 0$ means that the Model *S* is inferior to the Model *O*.
5. *Generalization of earlier measurements*: By setting $w_1 = 1$, $w_2 = 0$, it reduces to efficiency-only measurement (*E*) and $w_1 = 0$, $w_2 = 1$, it reduces to privacy-only measurement as that of Yan et al. [16].

The Proposed weighted $\psi$ measure thus not only integrates efficiency and privacy into a single criterion but also surpasses earlier methods through robustness, flexibility, and interpretability. It represents a clear advancement in the comparative evaluation of randomized response models and provides researchers with a powerful tool for balancing statistical accuracy with privacy protection.

### *4.1 Further Properties of $\psi$*

(i) $\psi$ (and $\log(\psi)$ is monotone in both the components $P$ and $E$. This can be verified as $\frac{\partial \log(\psi)}{\partial P} = \frac{w_1}{P}$. Since $w_1, P > 0$, we have $\frac{\partial \log(\psi)}{\partial P} > 0$. This implies $\log(\psi)$ is strictly increasing in $P$ .Similarly, $\frac{\partial \log(\psi)}{\partial E} = \frac{w_2}{E} > 0$ implies that $\log(\psi)$ is strictly increasing in *E*. Hence, we can conclude that for $w_1, w_2 > 0$, the $\psi$ is strictly increasing in both *P* and *E*. This implies that increasing both *P* and *E* improve $\psi$ and there is no paradoxical behavior of $\psi$. Consequently, $\psi$ is a consistent decision measure.

(ii) $\psi$ is scale and unit invariant. To show this, suppose $MSE_o$ is rescaled as $cMSE_0$ and $MSE_s$ is rescaled as $cMSE_s$ then $E=\frac{cMSE_o}{cMSE_s}=\frac{MSE_o}{MSE_s}$. Similarly, $\nabla_o$ is rescaled as $k\nabla_o$ and $\nabla_s$ is rescaled as $k\nabla_s$. So $\nabla_o$ is rescaled as $P=\frac{k\nabla_s}{k\nabla_o}=\frac{\nabla_s}{\nabla_o}$. Thus $\psi$ is invariant under common scaling. Now as we know that changing units (e.g., variance scale, measurement units) multiplies numerator and denominator equally, ratios remain unchanged. This means $\psi$ is a unit free measure.

(iii) $\psi$ is log- scale translation invariant. This can be shown by letting $P$ is translated to $cP$ then $\log(\psi)$ translates to $\log(\psi)+w_1\log(c)$. We see there is an absolute value shift in $\log(\psi)$. It implies ranking between models remain unchanged. Hence, ordering is invariant under multiplicative scaling.

(iv) $\psi$ is a symmetric measure. To check this, we reverse the comparison of Model *S* versus Model *O* as Model O versus Model *S* then $(\psi)^{-1}=\left(\frac{\nabla_o}{\nabla_s}\right)^{w_1}\times\left(\frac{MSE_s}{MSE_o}\right)^{w_2}$. Then $\log(\psi)^{-1}=-\log(\psi)$. Hence, we have a perfect symmetry between the competing Models *O* and *S*.

## 5. Some Existing Scrambling Models

We now present some of the important and recent scrambling models.

### *5.1 Warner [5] Model*

Using Warner [5] model, the response observed from the $i^{th}$ respondent can be written as:

$$Z_{1i}=Y_i+S_i.$$

Since $E(Z_{1i})=\mu_y$, an unbiased estimator of $\mu_y$, under the Warner [5] model is written as:

$$\hat{\mu}_W=\frac{1}{n}\sum_{i=1}^{n}Z_{1i},$$

with variance given by

$$Var(\hat{\mu}_W)=\frac{1}{n}\left(\sigma_y^2+\sigma_s^2\right). \tag{11}$$

The privacy measure $\nabla$, is calculated as

$$\nabla_W=\sigma_s^2. \tag{12}$$

### *5.2 Eichhorn and Hayre [19] Model*

The observed $i^{th}$ responses according to the Eichhorn and Hayre [19] model is given below:

$$Z_{2i}=Y_iS_i.$$

An unbiased estimator of sensitive mean $\mu_y$ based on the responses $Z_{2i}$, $i=(1,2,...,n)$ is given as:

$$\hat{\mu}_{EH}=\frac{1}{n}\sum_{i=1}^{n}Z_{2i}.$$

The variance of $\hat{\mu}_{EH}$ is given by:

$$Var(\hat{\mu}_{EH})=\frac{1}{n}\left[\sigma_y^2+\sigma_t^2\mu'_{2y}\right]. \tag{13}$$

The privacy measure $\nabla$, is given as

$$\nabla_{EH}=\sigma_t^{\,2}\mu'_{2y}. \tag{14}$$

### *5.3 The Diana and Perri [20] Model*

The $i^{th}$ response revealed under the Diana and Perri [20] model is written as:

$$Z_{3i}=Y_iT_i+S_i.$$

An unbiased estimator of $\mu_y$ using $Z_{3i}$ is given by:

$$\hat{\mu}_{DP}=\frac{1}{n}\sum_{i=1}^{n}Z_{3i}.$$

Its variance is given by:

$$Var\left(\hat{\mu}_{DP}\right)=\frac{1}{n}\left[\sigma_y^2+\sigma_s^2+\sigma_t^2\mu_{2y}'\right]. \tag{15}$$

The privacy measure $\nabla$, for this model is given by :

$$\nabla_{DP}=\sigma_y^2+\sigma_s^2+\sigma_t^2\mu_{2y}'. \tag{16}$$

***5.4 The Gjestvang ang Singh [21] Model***

The distribution of $i^{th}$ response under the Gjestvang and Singh [21] model can be written as:

$$Z_{4i}=\begin{cases} Y_i-\beta S_i & \text{with probability } p_1=\dfrac{\alpha}{\alpha+\beta} \\ Y_i+\alpha S_i & \text{with probability } p_2=\dfrac{\beta}{\alpha+\beta}, \end{cases}$$

where $\alpha$, and $\beta$ are constants determined by the interviewer. Gjestvang and Singh [21] suggested the mean estimator as

$$\hat{\mu}_{GS}=\frac{1}{n}\sum_{i=1}^{n}Z_{4i}.$$

The variance of $\hat{\mu}_{GS}$ is given by:

$$Var\left(\hat{\mu}_{GS}\right)=\frac{1}{n}\left[\sigma_y{}^2+\alpha\beta\left(\sigma_s{}^2\right)\right]. \tag{17}$$

The $\nabla$ measure for this model is given by:

$$\nabla_{GS}=\alpha\beta\left(\sigma_s{}^2\right). \tag{18}$$

***5.5 Murtaza et al. [22] Model***

The distribution of $i^{th}$ response under the Murtaza et al. [22] model may be written as:

$$Z_{5i}=\begin{cases} Y_i, & \text{with probability } (1-w) \\ Y_iT_i+\alpha S_i, & \text{with probability } (w), \end{cases}$$

where $\boldsymbol{\alpha}$ is a constant and $w$ is unknown proportion of the individuals in the population considering the study variable sensitive enough to hide. An unbiased estimator of $\mu_y$ through Murtaza et al. [22] model is given by:

$$\hat{\mu}_M=\frac{1}{n}\sum_{i=1}^{n}Z_{5i}.$$

Assuming independent variables, the variance of $\hat{\mu}_M$ is given by:

$$Var\left(\hat{\mu}_M\right)=\frac{1}{n}\left[\sigma_y{}^2+W\left\{\alpha^2\sigma_s{}^2+\sigma_t{}^2\mu_{2y}'\right\}\right]. \tag{19}$$

The $\nabla$ measure of respondent's privacy is given by :

$$\nabla_M=\left\{\sigma_t{}^2\mu_{2y}'+\alpha^2\sigma_s{}^2\right\}. \tag{20}$$

***5.6 Narjis and Shabbir [23] scrambling Model***

The distribution of $i^{th}$ response based on the Narjis and Shabbir [23] Model is as follows:

$$Z_{6i}=\begin{cases} Y_i-\beta S_i & \text{with probability } P_1=\dfrac{\alpha}{\alpha+\beta+\gamma} \\ Y_i+\alpha S_i & \text{with probability } P_2=\dfrac{\beta}{\alpha+\beta+\gamma} \\ Y_i & \text{with probability } P_3=\dfrac{\gamma}{\alpha+\beta+\gamma}, \end{cases}$$

where $\alpha$, $\beta$ and $\gamma$ denote the constants determined by the interviewer.

The estimator of $\mu_y$ using the Narjis and Shabbir [23] model is defined as :

$$\hat{\mu}_{NS} = \frac{1}{n}\sum_{i=1}^{n} Z_{6i}\,.$$

The variance of $\hat{\mu}_{NS}$ is given by :

$$Var\left(\hat{\mu}_{NS}\right) = \frac{1}{n}\left[\sigma_y{}^2 + \frac{\alpha\beta\left(\alpha+\beta\right)\sigma_s{}^2}{\alpha+\beta+\gamma}\right]. \tag{21}$$

The privacy measure $\nabla$, for this model is given by:

$$\nabla_{NS} = \frac{\alpha\beta\left(\alpha+\beta\right)\sigma_s{}^2}{\alpha+\beta+\gamma}\,. \tag{22}$$

### *5.7 Gupta et al. [24] Model*

Using the Gupta et al. [24] scrambling model, the distribution of $i^{th}$ response can be written as:

$$Z_{7i} = \begin{cases} Y_i & \text{with probability} \quad 1-W \\ Y_i + S_i & \text{with probability} \quad AW \\ Y_i T_i + S_i & \text{with probability} \quad (1-A)W. \end{cases}$$

An unbiased estimator of $\mu_y$ through the Gupta et al. [24] model is given by:

$$\hat{\mu}_G = \frac{1}{n}\sum_{i=1}^{n} Z_{7i}\,.$$

The variance of $\hat{\mu}_G$ is given below.

$$Var\left(\hat{\mu}_G\right) = \frac{1}{n}\left[\sigma_y^2 + W\sigma_s^2 + W\left(1-A\right)\sigma_t^2\mu_{2y}'\right] \tag{23}$$

Finally, the privacy measure $\nabla$, for Gupta et al. [24] model is given by :

$$\nabla_G = \frac{1}{n}\left[\sigma_s{}^2 + \left(1-A\right)\sigma_t{}^2\mu_{2y}'\right]. \tag{24}$$

## 6. Proposed Models as Special Cases of Generalized Framework

In this section, we present three different proposed randomized response models which are special cases of the generalized framework.

### *6.1 Proposed Model 1* $\left(M_1\right)$

The $i^{th}$ response under $M_1$ is written as

$$R_{M_{1i}} = D_{1i}Y_i\left(1+S_i\right) + \left(1-D_{1i}\right)Y_i\left(1-S_i\right),$$

where $D_{1i} \sim Bernoulli\left(P_1\right)$. The distribution of $i^{th}$ respondent response under the $M_1$ is given as:

$$R_{M_{1i}} = \begin{cases} Y_i\left(1+S_i\right) & \text{with probability} \quad P_1 \\ Y_i\left(1-S_i\right) & \text{with probability}\left(1-P_1\right) \end{cases}$$

An unbiased mean estimator $\mu_y$ of the sensitive variable is presented by $M_1$ can be expressed as:

$$\hat{\mu}_{M_1} = \frac{1}{n}\sum_{i=1}^{n} R_{M_{1i}}$$

The variance of $\hat{\mu}_{M_1}$ is given by:

$$Var\left(\hat{\mu}_{M_1}\right) = \frac{1}{n}\left[\sigma_y{}^2 + \sigma_s{}^2\mu_{2y}'\right] \tag{25}$$

For the $M_1$, respondent privacy can be written as:

$$\nabla_{M_1} = \sigma_s{}^2\mu_{2y}' \tag{26}$$

### *6.2 Proposed Model 2* $(M_2)$

The $i^{th}$ response under $M_2$ is written as

$$R_{M_{2i}} = D_{2i}Y_i(1+\alpha S_i)+(1-D_{2i})Y_i(1-\beta S_i),$$

where $D_{1i} \sim Bernoulli\left(P_1 = \frac{\alpha}{\alpha+\beta}\right)$. The distribution of $i^{th}$ respondent response under the $M_2$ is given as:

$$R_{M_{2i}} = \begin{cases} Y_i(1+\alpha S_i) & \text{with probability} \quad P_2 = \frac{\alpha}{\alpha+\beta} \\ Y(1-\beta S_i) & \text{with probability } (1-P_2) = \frac{\beta}{\alpha+\beta} \end{cases}$$

where $\alpha$ and $\beta$ denote the constants which are determined by the interviewer.
An unbiased mean estimator $\mu_y$ of the sensitive variable $M_2$ is written as:

$$\hat{\mu}_{M_2} = \frac{1}{n}\sum_{i=1}^{n} R_{M_{2i}}$$

The variance of $\hat{\mu}_{M_2}$ is given by:

$$Var\left(\hat{\mu}_{M_2}\right) = \frac{1}{n}\left[\sigma_y{}^2 + \sigma_s{}^2\mu'_{2y}\left(P_2\alpha^2 + (1-P_2)\beta^2\right)\right]. \tag{27}$$

For the $M_2$, the measure of respondent privacy is written as:

$$\nabla_{M_2} = \sigma_s{}^2\mu'_{2y}\left(P_2\alpha^2 + (1-P_2)\beta^2\right). \tag{28}$$

### *6.3 Proposed Model 3* $(M_3)$

The $i^{th}$ response under $M_3$ is written as

$$R_{M_{3i}} = D_{3i}Y_i(1+\alpha S_i)+D_{4i}Y_i(1-\beta S_i)+(1-D_{3i}-D_{4i})Y_i,$$

where $D_{3i} \sim Bernoulli\left(p_1 = \frac{\alpha}{\alpha+\beta+\gamma}\right)$ and $D_{4i} \sim Bernoulli\left(p_2 = \frac{\beta}{\alpha+\beta+\gamma}\right)$

The distribution of $i^{th}$ response under the $M_3$ is given by:

$$R_{M_{3i}} = \begin{cases} Y_i(1+\alpha S_i) & \text{with probability} \quad p_1 = \frac{\alpha}{\alpha+\beta+\gamma} \\ Y_i(1-\beta S_i) & \text{with probability } p_2 = \frac{\beta}{\alpha+\beta+\gamma} \\ Y_i & \text{with probability } p_3 = \frac{\gamma}{\alpha+\beta+\gamma}, \end{cases}$$

where $\alpha$, $\beta$ and $\gamma$ denote the constants determined by the interviewer.
An unbiased mean estimator $\mu_y$ of the sensitive variable of $M_3$ is given as:

$$\hat{\mu}_{M_3} = \frac{1}{n}\sum_{i=1}^{n} R_{M_{3i}}$$

The variance of $\hat{\mu}_{M_3}$ is given by:

$$Var\left(\hat{\mu}_{M_3}\right) = \frac{1}{n}\left[\sigma_y{}^2 + \frac{\alpha\beta(\alpha+\beta)\left(\sigma_y{}^2+\mu_y{}^2\right)\sigma_s{}^2}{\alpha+\beta+\gamma}\right] \tag{29}$$

For the $M_3$, the measure of respondent privacy is given by:

$$\nabla_{M_3} = \frac{\alpha\beta(\alpha+\beta)(\sigma_y^{\,2}+\mu_y^{\,2})\sigma_s^{\,2}}{\alpha+\beta+\gamma} \tag{30}$$

### *6.4 Proposed Models as Special Cases of Generalized Framework*

The proposed models are not entirely new in isolation, but their formulation under a unified random coefficient interaction framework provides new theoretical insights, including a closed-form variance structure and a clear efficiency–privacy trade-off governed by the second moment of the random coefficient. The proposed models are formulated within a unified random coefficient interaction framework $R_i = Y_i + C_i Y_i S_i$, where the random coefficient $C_i$ governs the level and nature of perturbation. This formulation generalizes existing additive and multiplicative models while introducing an additional layer of randomness through the coefficient structure. In particular, $M_3$ incorporates a zero-distortion state, allowing truthful responses with positive probability, which is not commonly accommodated in traditional randomized response designs. Furthermore, the proposed framework yields a unified variance expression and a clear efficiency–privacy trade-off governed by $E(C_i^2)$, providing theoretical insights that are not explicitly available in existing literature. The proposed models can be expressed within the random coefficient class as given in the Table 1 below.

**Table 1:** The proposed models and their random coefficient $C_i$.

| **Model** | **Expression** | **Random Coefficient $C_i$** |
|---|---|---|
| $M_1$ | $R_{M_{1i}} = D_{1i}(Y_i + Y_i S_i)_i + (1-D_{1i})(Y_i - Y_i S_i)$ | $2D_{1i}-1$ |
| $M_2$ | $R_{M_{2i}} = D_{2i}(Y_i + \alpha Y_i S_i) + (1-D_{2i})(Y_i - \beta Y_i S_i)$ | $D_{2i}\alpha - (1-D_{2i})\beta$ |
| $M_3$ | $R_{M_{3i}} = D_{3i}Y_i(1+\alpha S_i) + D_{4i}Y_i(1-\beta S_i) + (1-D_{3i}-D_{4i})(Y_i)$ | $D_{3i}\alpha - D_{4i}\beta$ |

Now, we see that how $M_1, M_2$ and $M_3$ differ in structure**.** Recall the common structure $R_i = Y_i + C_i Y_i S_i$**.** So, the only thing that differentiates models is the structure of $C_i$**.** Model $M_1$ has a binary symmetric coefficient such as $C_i = 2D_{1i} - 1$ and $D_{1i} \sim Bernoulli(P)$**.** The random coefficient $C_i \in \{-1,1\}$. Model $M_2$ is identified as asymmetric two-point coefficient model where $C_i \in \{\alpha, -\beta\}$**.** The Model $M_3$ is seen to be a multinomial coefficient model as in this case $C_i = D_{3i}\alpha - D_{4i}\beta$ and $C_i \in \{\alpha, -\beta, 0\}$**.** This establishes that the proposed models are structurally different with different level of flexibility. It is important to note that there exist mutual relationships between the models**.**

(i) $M_1$ as a Special Case of $M_2$

Set $\alpha = 1, \beta = 1$ Then**:** $C_i = D_{2i}\alpha - (1-D_{2i})\beta = D_{2i}(1) - (1-D_{2i})(1) = 2D_{2i} - 1$

Hence: $M_1 \subset M_2$

(ii) $M_2$ as a Special Case of $M_3$

If we set **,** $D_{3i} = D_{2i}$, and $D_{4i} = 1 - D_{2i}$ Then: $C_i = D_{3i}\alpha - D_{4i}\beta = D_{2i}\alpha - (1-D_{2i})\beta$.

Hence $M_2 \subset M_3$. So, we have the hierarchy as $M_1 \subset M_2 \subset M_3$.

### **6.5 Comparison of $M_1, M_2$ and $M_3$**

In this section, comparison of $M_1$**,** $M_2$ and $M_3$ using the new $\psi$ measure is presented. For clarity in presenting the comparison results, each of the proposed models is considered separately in comparison with the same set of seven existing randomized response models considered in section 5.

Let $E_{(j/M_1)} = \frac{Var(\hat{\mu}_j)}{Var(\hat{\mu}_{M_t})}$ and $P_{(j/M_t)} = \frac{\nabla_{M_t}}{\nabla_j}$, $j = W, EH, DP, GS, M, NS, G$ and $t = 1,2,3$ be the relative efficiency and relative privacy, respectively, of the model $M_t$ relative to $j^{th}$ existing models considered in section 5. Then we have the $\log(\psi)$ measure for the competing models as given below.

$$\log(\psi_{(j/M_t)}) = w_1 \log(E_{(j/M_t)}) + w_2 \log(P_{(j/M_t)}). \tag{31}$$

Using (11)-(30) in (31) with $t=1$, we get the $\log(\psi)$measure for $M_1$ relative to existing models. Graphical presentation of the $\log(\psi)$for different parametric settings is displayed in Figure 1 below.

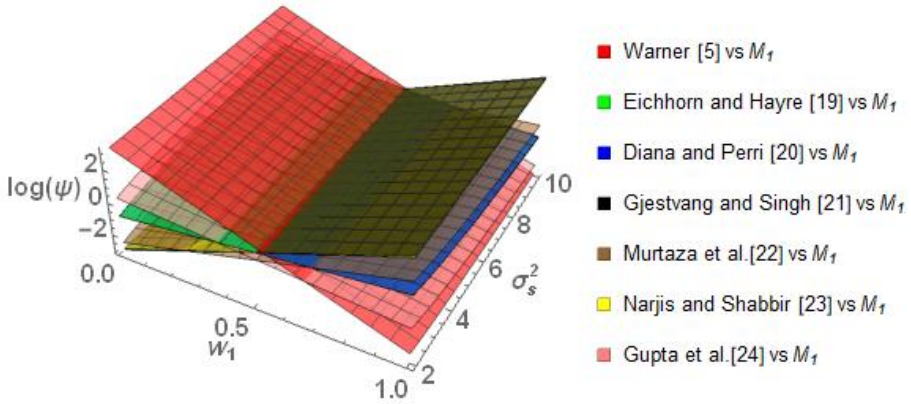


**Figure 1:** Behavior of $\log(\psi))$ for $M_1$ relative to existing models for $\sigma_y^2=2,\ \mu_y=5,\ \sigma_t^2=5,\ \alpha=25, \beta=25,\ \gamma=5,$ $P=0.4,\ P_1=0.4,\ P_2=0.6,\ A=0.7,\ W=0.2.$

The 3D plots in Figure 1 compare the performance of the proposed model $M_1$ with several existing randomized response models, including those of Warner [5], Eichhorn and Hayre [19], Diana and Perri [20], Gjestvang and Singh [21], Murtaza et al. [22], Narjis and Shabbir [23], and Gupta et al. [24]. The comparison is based on the measure $\log(\psi)$, which combines both efficiency and privacy into a single criterion. In this setup, positive values of $\log(\psi)$ indicate that the proposed model performs better than the competing model.

A careful look at the figure shows a clear pattern. The proposed model $M_1$ tends to perform better as the value of $w_1$ increases. For smaller values of $w_1$, some of the existing models remain competitive, and in a few cases perform slightly better. However, once $w_1$ moves into the moderate range, the advantage gradually shifts toward $M_1$. For example, the proposed model begins to outperform the Warner [5] model roughly beyond the mid-range of $w_1$, and a similar trend can be seen when compared with the models of Eichhorn and Hayre [19], and Diana and Perri [20], particularly when the variance parameter $\sigma_s^2$ is not too small.

For the models of Gjestvang and Singh [21] and Narjis and Shabbir [23], the dominance of $M_1$ is even more noticeable, as the corresponding surfaces lie mostly in the positive region across a wide span of the parameter space. A similar, though slightly less uniform, pattern is observed for the model of Murtaza et al.[22], where the proposed model again performs better for moderate to higher values of $w_1$. In the case of Gupta et al. 24], the proposed model appears to dominate almost entirely, except possibly when $w_1$ is very close to zero.

Overall, the figure suggests that the proposed model provides a more balanced performance, particularly when greater weight is given to the combined efficiency–privacy criterion. In practical terms, this means that $M_1$ becomes increasingly preferable as the emphasis shifts toward joint performance rather than relying on either efficiency or privacy alone.

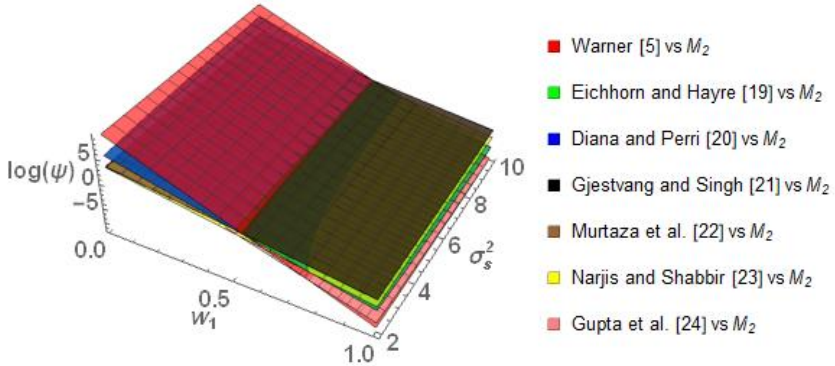

**Figure 2:** Behavior of $\log(\psi)$) for $M_2$ relative to existing models for $\sigma_y^2 = 2,\ \mu_y = 5,\ \sigma_t^2 = 5,\ \alpha = 25, \beta = 25,\ \gamma = 5,$ $P = 0.4,\ P_1 = 0.4,\ P_2 = 0.6,\ A = 0.7,\ W = 0.2.$

The 3D plots in Figure 2 compare the performance of the proposed model $M_2$ with several well-known randomized response models using the measure $\psi$. Here, positive values of $\log(\psi)$ indicate that $M_2$ performs better than the competing model.

One clear pattern that emerges from the figure is that the value of $\log(\psi)$ generally decreases as $w_1$ increases. This suggests that the relative advantage of $M_2$ becomes weaker when more weight is assigned to the combined efficiency–privacy component. In fact, for most of the competing models, the surfaces start in the positive region for small values of $w_1$, but gradually move toward zero and then into the negative region as $w_1$ increases.

For smaller values of $w_1$, the proposed model performs quite well. In this range, the surfaces corresponding to all competing models remain above zero, indicating that $w_1$ consistently outperforms them. This advantage is fairly strong against the Warner model and is also clearly visible for the models of Diana and Perri [20], Eichhorn and Hayre [19], and Gjestvang and Singh [21].

As $w_1$ increases to moderate levels, the situation begins to change. The gap between $M_2$ and some competing models narrows, and in certain cases the competing models start to perform better. This is particularly noticeable for the models of Narjis and Shabbir [23], Murtaza et al.[22], and Gupta et al. [24], where the corresponding surfaces approach or fall below zero.

The role of $\sigma_s^2$ appears to be less pronounced. While it slightly affects the height and slope of the surfaces, it does not seem to change the overall pattern of dominance. The main shift is clearly driven by $w_1$.

Overall, the Figure 2 suggests that $M_2$ is more suitable when the value of $w_1$ is relatively small. As $w_1$ increases, its advantage reduces, and other models may become preferable. In this sense, the performance of $M_2$ appears to be more sensitive to the choice of the weighting parameter compared to what was observed for $M_1$.

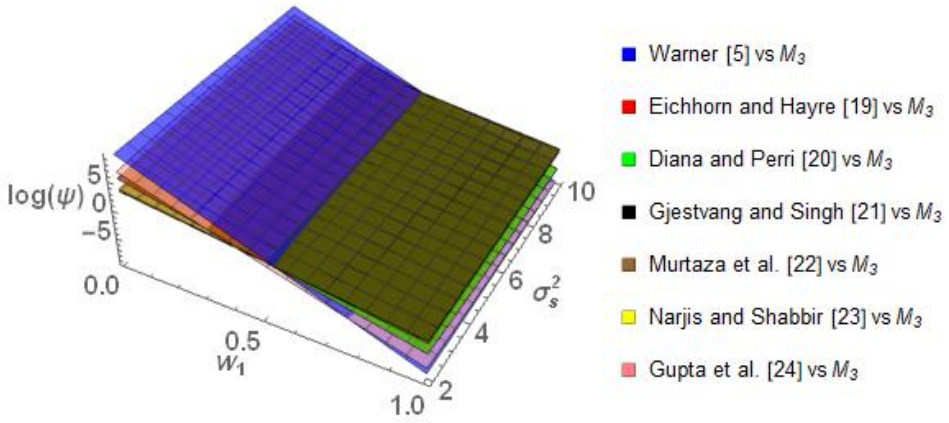


**Figure 3**: Behavior of $\log(\psi)$) for $M_3$ relative to existing models for $\sigma_y^2 = 2,\ \mu_y = 5,\ \sigma_t^2 = 5,\ \alpha = 25, \beta = 25,\ \gamma = 5,$ $P = 0.4,\ P_1 = 0.4,\ P_2 = 0.6,\ A = 0.7,\ W = 0.2.$

The 3D plots in Figure 3 present a comparison of the proposed model $M_3$ with several existing randomized response models using the combined performance measure $\psi$. As in the previous figures, positive values of $\log(\psi)$ indicate that $M_3$ performs better than the competing model.

A noticeable feature of the figure is that $\log(\psi)$ tends to decrease as $w_1$ increases. For smaller values of $w_1$, most of the surfaces lie above zero, indicating that $M_3$ performs better than the competing models in this region. This advantage is quite visible across nearly all models, including Warner [5], Eichhorn and Hayre [19], and Diana and Perri [20].

However, as $w_1$ increases, the surfaces gradually move downward and, in several cases, cross into the negative region. This suggests that the relative performance of $M_3$ weakens for moderate to larger values of $w_1$. In particular, for models such

as Gjestvang and Singh [21], Murtaza et al. [22], Narjis and Shabbir [23], and Gupta et al. [24], the advantage of $M_3$ reduces noticeably and may even reverse in parts of the parameter space.

The effect of $\sigma_s^2$ appears to be relatively mild compared to $w_1$. While variations in $\sigma_s^2$ slightly adjust the level of the surfaces, they do not significantly alter the overall pattern. The dominant factor influencing performance remains the weight parameter $w_1$.

Overall, the Figure 3 suggests that $M_3$ performs well when $w_1$ is small, but its advantage diminishes as $w_1$ increases. This indicates that, similar to $M_2$, the effectiveness of $M_3$ is more pronounced under lower weighting of the combined efficiency–privacy component.

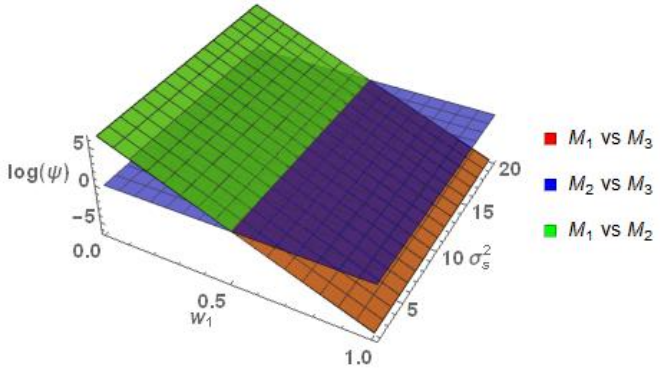


**Figure 4:** Behavior of $\log(\psi)$) for $M_1$ vs $M_2$, $M_1$ vs $M_3$ and $M_2$ vs $M_3$ for $\sigma_y^2 = 2$, $\mu_y = 5$, $\sigma_t^2 = 5$, $\alpha = 25$, $\beta = 25$, $\gamma = 5$, $P = 0.4$, $P_1 = 0.4$, $P_2 = 0.6$, $A = 0.7$, $W = 0.2$.

The 3D surfaces in the figure 4 present a direct comparison among the three proposed models $M_1, M_2$ and $M_3$ using the performance measure $\psi$. As before, positive values of $\log(\psi)$ indicate that the model in the numerator of the comparison outperforms the one in the denominator.

A clear pattern emerges when observing how the surfaces behave with respect to $w_1$. The surface corresponding to the comparison between $M_1$ and $M_2$ (shown in green) remains largely above zero across most of the parameter space. This suggests that $M_1$ consistently outperforms $M_2$ and this advantage becomes more pronounced as $w_1$ increases.

A similar trend can be seen when comparing $M_1$ with $M_3$ (red surface). For moderate to higher values of $w_1$, the surface lies above zero, indicating that $M_1$ dominates $M_3$ in this region. However, for smaller values of $w_1$, the surface approaches zero and may dip slightly below it, suggesting that the advantage of $M_1$ is less clear and that $M_3$ can be competitive in this range.

In contrast, the comparison between $M_2$ and $M_3$ (blue surface) shows a more balanced behavior. For smaller values of $w_1$, the surface tends to lie above zero, indicating that $M_2$ performs better than $M_3$ in this region. As $w_1$ increases, however, the surface declines and approaches or crosses zero, suggesting that the relative advantage between these two models diminishes and may even reverse.

The role of $\sigma_s^2$ appears to be secondary in all three comparisons. While it influences the magnitude of $\log(\psi)$, it does not significantly change the overall ordering of the models. The primary driver of the relative performance remains the weight parameter $w_1$.

Taken together, the pairwise comparisons reveal a clear ordering among the proposed models that depends primarily on the value of the weight parameter $w_1$. For smaller values of $w_1$, the models $M_2$ and $M_3$ tend to perform better, with $M_2$ showing a slight advantage over $M_3$. However, as $w_1$ increases, the performance of $M_1$ improves steadily and eventually surpasses both $M_2$ and $M_3$.

This indicates that no single model is uniformly superior across all parameter settings. Instead, each model performs best within a specific region of the design space. In particular, $M_2$ and $M_3$ are more suitable for lower values of $w_1$, whereas

$M_1$ becomes the preferred choice when $w_1$ is moderate to large. This structured behavior provides useful guidance for model selection based on the desired balance between efficiency and privacy.

In order to have more clarity of the behavior of $\log(\psi)$, different design parameters' settings are considered. The values of $\log(\psi)$ for $M_1$, $M_2$ and $M_3$ are arranged in the Tables 2-9, given below.

**Table 2**: Values of $\log(\psi)$ for $M_1$ when $\sigma_y^2 = 2$, $\mu_y = 2$, $\sigma_t^2 = 3$, $\alpha = 15$, $\beta = 15$, $\gamma = 10$, $A = 0.7$, $P = 0.6$, $W = 0.2$, $P_1 = 0.6$, $P_2 = 0.4$.

| $\sigma_s^2$ | $w_1$ | $\log\left(\psi_{(W/M_1)}\right)$ | $\log\left(\psi_{(EH/M_1)}\right)$ | $\log\left(\psi_{(DP/M_1)}\right)$ | $\log\left(\psi_{(GS/M_1)}\right)$ | $\log\left(\psi_{(M/M_1)}\right)$ | $\log\left(\psi_{(NS/M_1)}\right)$ | $\log\left(\psi_{(G/M_1)}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.05 | 1.639533 | -0.46269 | -0.36736 | -3.26939 | -3.01041 | -2.74729 | 0.420429 |
| 2 | 0.15 | 1.335081 | -0.3664 | -0.29114 | -2.5595 | -2.3579 | -2.20917 | 0.294433 |
| 2 | 0.3 | 0.878403 | -0.22198 | -0.17682 | -1.49465 | -1.37914 | -1.40198 | 0.10544 |
| 2 | 0.45 | 0.421724 | -0.07756 | -0.0625 | -0.42981 | -0.40037 | -0.59479 | -0.08355 |
| 2 | 0.6 | -0.03495 | 0.066861 | 0.051819 | 0.635039 | 0.578386 | 0.212393 | -0.27255 |
| 2 | 0.75 | -0.49163 | 0.211282 | 0.16614 | 1.699883 | 1.557148 | 1.019579 | -0.46154 |
| 2 | 0.9 | -0.94831 | 0.355704 | 0.280461 | 2.764728 | 2.535909 | 1.826765 | -0.65053 |
| 4 | 0.05 | 1.628855 | 0.078659 | 0.26018 | -3.2658 | -3.00685 | -2.12707 | 0.829242 |
| 4 | 0.15 | 1.303045 | 0.061953 | 0.205175 | -2.54871 | -2.34722 | -1.71855 | 0.613038 |
| 4 | 0.3 | 0.814331 | 0.036895 | 0.122668 | -1.47308 | -1.35779 | -1.10577 | 0.288731 |
| 4 | 0.45 | 0.325616 | 0.011837 | 0.040161 | -0.39745 | -0.36835 | -0.49299 | -0.03558 |
| 4 | 0.6 | -0.1631 | -0.01322 | -0.04235 | 0.678174 | 0.621081 | 0.119789 | -0.35988 |
| 4 | 0.75 | -0.65181 | -0.03828 | -0.12485 | 1.753803 | 1.610516 | 0.73257 | -0.68419 |
| 4 | 0.9 | -1.14053 | -0.06334 | -0.20736 | 2.829432 | 2.599951 | 1.345351 | -1.00849 |
| 6 | 0.05 | 1.624264 | 0.366217 | 0.626397 | -3.26454 | -3.0056 | -1.76812 | 1.019495 |
| 6 | 0.15 | 1.289274 | 0.287722 | 0.492897 | -2.54493 | -2.34348 | -1.43649 | 0.758675 |
| 6 | 0.3 | 0.786788 | 0.169979 | 0.292647 | -1.46551 | -1.35029 | -0.93905 | 0.367444 |
| 6 | 0.45 | 0.284303 | 0.052235 | 0.092397 | -0.3861 | -0.35711 | -0.44161 | -0.02379 |
| 6 | 0.6 | -0.21818 | -0.06551 | -0.10785 | 0.693316 | 0.636075 | 0.05583 | -0.41502 |
| 6 | 0.75 | -0.72067 | -0.18325 | -0.3081 | 1.77273 | 1.629259 | 0.553271 | -0.80625 |
| 6 | 0.9 | -1.22315 | -0.30099 | -0.50835 | 2.852145 | 2.622443 | 1.050712 | -1.19748 |
| 8 | 0.05 | 1.6217 | 0.553458 | 0.885973 | -3.26389 | -3.00496 | -1.51575 | 1.131877 |
| 8 | 0.15 | 1.28158 | 0.434166 | 0.696261 | -2.543 | -2.34156 | -1.23874 | 0.843737 |
| 8 | 0.3 | 0.7714 | 0.255228 | 0.411693 | -1.46165 | -1.34647 | -0.82324 | 0.411529 |
| 8 | 0.45 | 0.261221 | 0.076289 | 0.127125 | -0.3803 | -0.35137 | -0.40773 | -0.02068 |
| 8 | 0.6 | -0.24896 | -0.10265 | -0.15744 | 0.701041 | 0.643727 | 0.007777 | -0.45289 |
| 8 | 0.75 | -0.75914 | -0.28159 | -0.44201 | 1.782387 | 1.638823 | 0.423285 | -0.8851 |
| 8 | 0.9 | -1.26932 | -0.46053 | -0.72658 | 2.863732 | 2.63392 | 0.838792 | -1.31731 |
| 10 | 0.05 | 1.62006 | 0.687736 | 1.087204 | -3.2635 | -3.00457 | -1.32167 | 1.206595 |
| 10 | 0.15 | 1.276661 | 0.538928 | 0.853667 | -2.54182 | -2.3404 | -1.08691 | 0.899832 |
| 10 | 0.3 | 0.761563 | 0.315717 | 0.50336 | -1.45931 | -1.34415 | -0.73478 | 0.439687 |
| 10 | 0.45 | 0.246465 | 0.092505 | 0.153054 | -0.37679 | -0.34789 | -0.38264 | -0.02046 |
| 10 | 0.6 | -0.26863 | -0.13071 | -0.19725 | 0.705727 | 0.648369 | -0.03051 | -0.4806 |
| 10 | 0.75 | -0.78373 | -0.35392 | -0.54756 | 1.788244 | 1.644625 | 0.321628 | -0.94075 |
| 10 | 0.9 | -1.29883 | -0.57713 | -0.89786 | 2.870762 | 2.640882 | 0.673762 | -1.40089 |

**Table 3:** Values of $\log(\psi)$ for $M_1$ when $\sigma_y^2 = 2$, $\mu_y = 5$, $\sigma_t^2 = 5$, $\alpha = 25$, $\beta = 25$, $\gamma = 5$, $A = 0.7$, $P = 0.4$, $W = 0.2$, $P_1 = 0.4$, $P_2 = 0.6$.

| $\sigma_s^2$ | $w_1$ | $\log\left(\psi_{(W/M_1)}\right)$ | $\log\left(\psi_{(EH/M_1)}\right)$ | $\log\left(\psi_{(DP/M_1)}\right)$ | $\log\left(\psi_{(GS/M_1)}\right)$ | $\log\left(\psi_{(M/M_1)}\right)$ | $\log\left(\psi_{(NS/M_1)}\right)$ | $\log\left(\psi_{(G/M_1)}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.05 | 2.999092 | -0.83899 | -0.82574 | -2.82946 | -2.74367 | -2.42957 | 0.17346 |
| 2 | 0.15 | 2.405603 | -0.65498 | -0.64465 | -2.20456 | -2.13781 | -1.99479 | 0.041421 |
| 2 | 0.3 | 1.515369 | -0.37896 | -0.37301 | -1.2672 | -1.22902 | -1.34262 | -0.15664 |
| 2 | 0.45 | 0.625134 | -0.10294 | -0.10138 | -0.32984 | -0.32023 | -0.69045 | -0.3547 |
| 2 | 0.6 | -0.2651 | 0.17307 | 0.170261 | 0.607522 | 0.588555 | -0.03828 | -0.55276 |
| 2 | 0.75 | -1.15533 | 0.44909 | 0.441899 | 1.544881 | 1.497345 | 0.613893 | -0.75081 |
| 2 | 0.9 | -2.04557 | 0.7251 | 0.713537 | 2.48224 | 2.406135 | 1.266064 | -0.94887 |
| 4 | 0.05 | 2.985609 | -0.22731 | -0.20101 | -2.8286 | -2.74282 | -1.8072 | 0.75657 |
| 4 | 0.15 | 2.365153 | -0.17725 | -0.15675 | -2.20197 | -2.13524 | -1.50872 | 0.496426 |
| 4 | 0.3 | 1.43447 | -0.10216 | -0.09035 | -1.26203 | -1.22388 | -1.06101 | 0.10621 |
| 4 | 0.45 | 0.503786 | -0.02706 | -0.02395 | -0.32209 | -0.31252 | -0.6133 | -0.28401 |
| 4 | 0.6 | -0.4269 | 0.04803 | 0.042443 | 0.617853 | 0.598839 | -0.16558 | -0.67422 |
| 4 | 0.75 | -1.35758 | 0.12312 | 0.10884 | 1.557795 | 1.5102 | 0.282131 | -1.06444 |
| 4 | 0.9 | -2.28827 | 0.19821 | 0.175236 | 2.497737 | 2.421561 | 0.729844 | -1.45465 |
| 6 | 0.05 | 2.980024 | 0.12504 | 0.164211 | -2.82831 | -2.74253 | -1.44433 | 1.082008 |
| 6 | 0.15 | 2.348398 | 0.09745 | 0.12799 | -2.2011 | -2.13438 | -1.22582 | 0.749736 |
| 6 | 0.3 | 1.400958 | 0.05607 | 0.073659 | -1.26029 | -1.22215 | -0.89805 | 0.251328 |
| 6 | 0.45 | 0.453519 | 0.0147 | 0.019328 | -0.31947 | -0.30992 | -0.57028 | -0.24708 |
| 6 | 0.6 | -0.49392 | -0.02668 | -0.035 | 0.621341 | 0.602311 | -0.24252 | -0.74549 |
| 6 | 0.75 | -1.44136 | -0.06806 | -0.08933 | 1.562155 | 1.51454 | 0.085252 | -1.2439 |
| 6 | 0.9 | -2.3888 | -0.10944 | -0.14366 | 2.502969 | 2.426768 | 0.41302 | -1.7423 |
| 8 | 0.05 | 2.97695 | 0.371424 | 0.423278 | -2.82816 | -2.74238 | -1.18761 | 1.302974 |
| 8 | 0.15 | 2.339175 | 0.289405 | 0.329826 | -2.20066 | -2.13394 | -1.02581 | 0.921492 |
| 8 | 0.3 | 1.382513 | 0.166375 | 0.189648 | -1.25941 | -1.22127 | -0.78311 | 0.349269 |
| 8 | 0.45 | 0.425851 | 0.04336 | 0.049471 | -0.31816 | -0.30861 | -0.54041 | -0.22295 |
| 8 | 0.6 | -0.53081 | -0.07968 | -0.09071 | 0.623093 | 0.604055 | -0.29771 | -0.79518 |
| 8 | 0.75 | -1.48747 | -0.20271 | -0.23088 | 1.564345 | 1.51672 | -0.05501 | -1.3674 |
| 8 | 0.9 | -2.44414 | -0.32574 | -0.37106 | 2.505597 | 2.429385 | 0.187693 | -1.93962 |
| 10 | 0.05 | 2.975 | 0.559835 | 0.624199 | -2.82808 | -2.7423 | -0.98903 | 1.467341 |
| 10 | 0.15 | 2.333327 | 0.43613 | 0.486302 | -2.2004 | -2.13368 | -0.87116 | 1.049125 |
| 10 | 0.3 | 1.370817 | 0.250571 | 0.279457 | -1.25888 | -1.22075 | -0.69435 | 0.421801 |
| 10 | 0.45 | 0.408307 | 0.065012 | 0.072611 | -0.31737 | -0.30782 | -0.51754 | -0.20552 |
| 10 | 0.6 | -0.5542 | -0.12055 | -0.13423 | 0.624147 | 0.605104 | -0.34073 | -0.83285 |
| 10 | 0.75 | -1.51671 | -0.30611 | -0.34108 | 1.565662 | 1.518031 | -0.16391 | -1.46017 |
| 10 | 0.9 | -2.47922 | -0.49166 | -0.54792 | 2.507178 | 2.430958 | 0.012896 | -2.08749 |

**Table 4:** Values of $\log(\psi)$ for $M_2$ when $\sigma_y^2 = 2$, $\mu_y = 2$, $\sigma_t^2 = 3$, $\alpha = 15$, $\beta = 15$, $\gamma = 10$, $A = 0.7$, $P = 0.6$, $W = 0.2$, $P_1 = 0.6$, $P_2 = 0.4$.

| $\sigma_s^2$ | $w_1$ | $\log\left(\psi_{(W/M_2)}\right)$ | $\log\left(\psi_{(EH/M_2)}\right)$ | $\log\left(\psi_{(DP/M_2)}\right)$ | $\log\left(\psi_{(GS/M_2)}\right)$ | $\log\left(\psi_{(M/M_2)}\right)$ | $\log\left(\psi_{(NS/M2)}\right)$ | $\log\left(\psi_{(G/M_2)}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.05 | 6.521694 | 4.514803 | 4.419476 | 1.612768 | 1.871756 | 1.497833 | 5.271815 |
| 2 | 0.15 | 5.149363 | 3.523138 | 3.447878 | 1.254786 | 1.456384 | 0.988422 | 4.016391 |
| 2 | 0.3 | 3.090866 | 2.03564 | 1.990481 | 0.717812 | 0.833327 | 0.224305 | 2.133254 |
| 2 | 0.45 | 1.032369 | 0.548143 | 0.533084 | 0.180838 | 0.21027 | -0.53981 | 0.250118 |
| 2 | 0.6 | -1.02613 | -0.93936 | -0.92431 | -0.35614 | -0.41279 | -1.30393 | -1.63302 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.75 | -3.08462 | -2.42685 | -2.38171 | -0.89311 | -1.03584 | -2.06805 | -3.51615 |
| 2 | 0.9 | -5.14312 | -3.91435 | -3.83911 | -1.43008 | -1.6589 | -2.83216 | -5.39929 |
| 4 | 0.05 | 6.507329 | 5.138654 | 4.957133 | 1.612676 | 1.871627 | 1.514812 | 5.673838 |
| 4 | 0.15 | 5.106266 | 4.008396 | 3.865174 | 1.254509 | 1.455997 | 1.000524 | 4.314624 |
| 4 | 0.3 | 3.004672 | 2.31301 | 2.227237 | 0.717259 | 0.832553 | 0.229091 | 2.275804 |
| 4 | 0.45 | 0.903079 | 0.617624 | 0.5893 | 0.180008 | 0.209109 | -0.54234 | 0.236984 |
| 4 | 0.6 | -1.19852 | -1.07776 | -1.04864 | -0.35724 | -0.41434 | -1.31378 | -1.80184 |
| 4 | 0.75 | -3.3001 | -2.77315 | -2.68658 | -0.89449 | -1.0377 | -2.08521 | -3.84066 |
| 4 | 0.9 | -5.4017 | -4.46853 | -4.32451 | -1.43174 | -1.6612 | -2.85664 | -5.87948 |
| 6 | 0.05 | 6.50144 | 5.50357 | 5.243399 | 1.612645 | 1.87158 | 1.520543 | 5.86041 |
| 6 | 0.15 | 5.08861 | 4.2922 | 4.087065 | 1.254417 | 1.45586 | 1.0046 | 4.449218 |
| 6 | 0.3 | 2.96937 | 2.47523 | 2.352565 | 0.717074 | 0.83229 | 0.230685 | 2.33243 |
| 6 | 0.45 | 0.85013 | 0.65822 | 0.618065 | 0.179731 | 0.20872 | -0.54323 | 0.215642 |
| 6 | 0.6 | -1.2691 | -1.15878 | -1.11644 | -0.35761 | -0.4148 | -1.31714 | -1.90115 |
| 6 | 0.75 | -3.3883 | -2.97579 | -2.85094 | -0.89495 | -1.0384 | -2.09106 | -4.01793 |
| 6 | 0.9 | -5.5076 | -4.7928 | -4.58544 | -1.4323 | -1.662 | -2.86497 | -6.13472 |
| 8 | 0.05 | 6.49822 | 5.76249 | 5.429981 | 1.61263 | 1.87156 | 1.523421 | 5.970235 |
| 8 | 0.15 | 5.07894 | 4.49362 | 4.231532 | 1.25437 | 1.45580 | 1.006646 | 4.526611 |
| 8 | 0.3 | 2.95003 | 2.59032 | 2.433859 | 0.716981 | 0.83216 | 0.231482 | 2.361176 |
| 8 | 0.45 | 0.82111 | 0.68702 | 0.636186 | 0.179592 | 0.20852 | -0.54368 | 0.195741 |
| 8 | 0.6 | -1.3078 | -1.21628 | -1.16149 | -0.3578 | -0.4151 | -1.31884 | -1.96969 |
| 8 | 0.75 | -3.4367 | -3.11958 | -2.95916 | -0.89519 | -1.0387 | -2.09401 | -4.13513 |
| 8 | 0.9 | -5.5656 | -5.02289 | -4.75683 | -1.43257 | -1.6623 | -2.86917 | -6.30056 |
| 10 | 0.05 | 6.49618 | 5.96332 | 5.563859 | 1.612621 | 1.87154 | 1.525153 | 6.043009 |
| 10 | 0.15 | 5.07282 | 4.64983 | 4.335095 | 1.254343 | 1.45576 | 1.007876 | 4.576874 |
| 10 | 0.3 | 2.93779 | 2.67959 | 2.49195 | 0.716926 | 0.83208 | 0.231961 | 2.37767 |
| 10 | 0.45 | 0.80276 | 0.70935 | 0.648804 | 0.179509 | 0.20841 | -0.54395 | 0.178466 |
| 10 | 0.6 | -1.3322 | -1.26089 | -1.19434 | -0.35791 | -0.4152 | -1.31987 | -2.02074 |
| 10 | 0.75 | -3.4673 | -3.23113 | -3.03749 | -0.89532 | -1.0389 | -2.09579 | -4.21994 |
| 10 | 0.9 | -5.6023 | -5.20137 | -4.88063 | -1.43274 | -1.6626 | -2.8717 | -6.41915 |

**Table 5:** Values of $\log(\psi)$ for $M_2$ when $\sigma_y^2 = 2$, $\mu_y = 5$, $\sigma_t^2 = 5$, $\alpha = 25$, $\beta = 25$, $\gamma = 5$, $A = 0.7$, $P = 0.4$, $W = 0.2$, $P_1 = 0.4$, $P_2 = 0.6$

| $\sigma_s^2$ | $w_1$ | $\log\left(\psi_{(W/M_2)}\right)$ | $\log\left(\psi_{(EH/M_2)}\right)$ | $\log\left(\psi_{(DP/M_2)}\right)$ | $\log\left(\psi_{(GS/M_2)}\right)$ | $\log\left(\psi_{(M/M_2)}\right)$ | $\log\left(\psi_{(NS/M_2)}\right)$ | $\log\left(\psi_{(G/M_2)}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.05 | 8.794884 | 4.970047 | 4.956801 | 2.96633 | 3.202482 | 2.793837 | 5.939599 |
| 2 | 0.15 | 6.917475 | 3.86722 | 3.856893 | 2.307317 | 2.491044 | 1.994951 | 4.464334 |
| 2 | 0.3 | 4.101362 | 2.212978 | 2.207032 | 1.318797 | 1.423886 | 0.796621 | 2.251436 |
| 2 | 0.45 | 1.285248 | 0.558737 | 0.557171 | 0.330276 | 0.356728 | -0.40171 | 0.038539 |
| 2 | 0.6 | -1.53086 | -1.0955 | -1.09269 | -0.65824 | -0.71043 | -1.60004 | -2.17436 |
| 2 | 0.75 | -4.34698 | -2.74975 | -2.74255 | -1.64676 | -1.77759 | -2.79837 | -4.38726 |
| 2 | 0.9 | -7.16309 | -4.40399 | -4.39241 | -2.63528 | -2.84474 | -3.9967 | -6.60015 |
| 4 | 0.05 | 8.780501 | 5.593881 | 5.567581 | 2.966292 | 3.202432 | 2.838636 | 6.521617 |
| 4 | 0.15 | 6.874327 | 4.352427 | 4.331925 | 2.307201 | 2.490892 | 2.029419 | 4.916062 |
| 4 | 0.3 | 4.015066 | 2.490246 | 2.47844 | 1.318566 | 1.423583 | 0.815594 | 2.507731 |
| 4 | 0.45 | 1.155805 | 0.628065 | 0.624956 | 0.32993 | 0.356275 | -0.39823 | 0.099399 |
| 4 | 0.6 | -1.70346 | -1.23412 | -1.22853 | -0.65871 | -0.71103 | -1.61206 | -2.30893 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4 | 0.75 | -4.56272 | -3.0963 | -3.08201 | -1.64734 | -1.77834 | -2.82588 | -4.71726 |
| 4 | 0.9 | -7.42198 | -4.95848 | -4.9355 | -2.63598 | -2.84565 | -4.03971 | -7.1256 |
| 6 | 0.05 | 8.774613 | 5.9588 | 5.919633 | 2.966279 | 3.202415 | 2.854078 | 6.846573 |
| 6 | 0.15 | 6.856661 | 4.636254 | 4.605721 | 2.307163 | 2.490842 | 2.041296 | 5.167927 |
| 6 | 0.3 | 3.979734 | 2.652435 | 2.634855 | 1.318489 | 1.423482 | 0.822122 | 2.649957 |
| 6 | 0.45 | 1.102807 | 0.668616 | 0.663988 | 0.329815 | 0.356123 | -0.39705 | 0.131988 |
| 6 | 0.6 | -1.77412 | -1.3152 | -1.30688 | -0.65886 | -0.71124 | -1.61623 | -2.38598 |
| 8 | 0.75 | -4.65105 | -3.29902 | -3.27775 | -1.6475 | -1.7786 | -2.8354 | -4.90395 |
| 8 | 0.9 | -7.52797 | -5.28284 | -5.24861 | -2.6362 | -2.8459 | -4.05457 | -7.42192 |
| 8 | 0.05 | 8.771386 | 6.21771 | 6.165861 | 2.96627 | 3.20240 | 2.8619 | 7.06722 |
| 8 | 0.15 | 6.846981 | 4.83763 | 4.797211 | 2.30714 | 2.49081 | 2.04731 | 5.338728 |
| 8 | 0.3 | 3.960374 | 2.7675 | 2.744236 | 1.31845 | 1.42343 | 0.825426 | 2.74599 |
| 8 | 0.45 | 1.073767 | 0.69738 | 0.691262 | 0.32975 | 0.35604 | -0.39646 | 0.153251 |
| 8 | 0.6 | -1.81284 | -1.37274 | -1.36171 | -0.6589 | -0.7113 | -1.61834 | -2.43949 |
| 8 | 0.75 | -4.69945 | -3.44286 | -3.41469 | -1.6476 | -1.7787 | -2.84023 | -5.03223 |
| 8 | 0.9 | -7.58605 | -5.51298 | -5.46766 | -2.6363 | -2.8461 | -4.06211 | -7.62496 |
| 10 | 0.05 | 8.769345 | 6.41854 | 6.35418 | 2.96626 | 3.20241 | 2.866626 | 7.231341 |
| 10 | 0.15 | 6.840858 | 4.99383 | 4.943661 | 2.30713 | 2.49080 | 2.050944 | 5.465623 |
| 10 | 0.3 | 3.948128 | 2.85676 | 2.827882 | 1.31842 | 1.42340 | 0.827421 | 2.817046 |
| 10 | 0.45 | 1.055398 | 0.71970 | 0.712103 | 0.32972 | 0.35600 | -0.3961 | 0.168469 |
| 10 | 0.6 | -1.83733 | -1.41736 | -1.40368 | -0.6589 | -0.7114 | -1.61962 | -2.48011 |
| 10 | 0.75 | -4.73006 | -3.55443 | -3.51945 | -1.6476 | -1.7788 | -2.84315 | -5.12868 |
| 10 | 0.9 | -7.62279 | -5.69149 | -5.63523 | -2.6363 | -2.8462 | -4.06667 | -7.77726 |

**Table 6:** Values of $\log(\psi)$ for $M_3$ when $\sigma_y^2 = 2,\ \mu_y = 2,\ \sigma_t^2 = 3,\ \alpha = 15,\ \beta = 15,\ \gamma = 10,\ A = 0.7,\ P = 0.6,\ W = 0.2,\ P_1 = 0.6,\ P_2 = 0.4.$

| $\sigma_s^2$ | $w_1$ | $\log\left(\psi_{(W/M_3)}\right)$ | $\log\left(\psi_{(EH/M_3)}\right)$ | $\log\left(\psi_{(DP/M_3)}\right)$ | $\log\left(\psi_{(GS/M_3)}\right)$ | $\log\left(\psi_{(M/M_3)}\right)$ | $\log\left(\psi_{(NS/M_3)}\right)$ | $\log\left(\psi_{(G/M_3)}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.1 | 5.605358 | 3.7888 | 3.703506 | 1.203607 | 1.4339 | 1.905882 | 4.413932 |
| 2 | 0.3 | 2.975719 | 1.920493 | 1.875334 | 0.602665 | 0.71818 | 0.955351 | 2.018108 |
| 2 | 0.5 | 0.34608 | 0.052187 | 0.047162 | 0.001724 | 0.002461 | 0.00482 | -0.37772 |
| 2 | 0.7 | -2.28356 | -1.81612 | -1.78101 | -0.59922 | -0.71326 | -0.94571 | -2.77354 |
| 2 | 0.9 | -4.9132 | -3.68443 | -3.60918 | -1.20016 | -1.42898 | -1.89624 | -5.16937 |
| 4 | 0.1 | 5.576639 | 4.343367 | 4.180996 | 1.203435 | 1.433654 | 1.920946 | 4.764073 |
| 4 | 0.3 | 2.889563 | 2.1979 | 2.112127 | 0.602149 | 0.717443 | 0.961708 | 2.160695 |
| 4 | 0.5 | 0.202486 | 0.052433 | 0.043259 | 0.000863 | 0.001232 | 0.002469 | -0.44268 |
| 4 | 0.7 | -2.48459 | -2.09303 | -2.02561 | -0.60042 | -0.71498 | -0.95677 | -3.04606 |
| 4 | 0.9 | -5.17167 | -4.2385 | -4.09448 | -1.20171 | -1.43119 | -1.91601 | -5.64944 |
| 6 | 0.1 | 5.564878 | 4.667756 | 4.435078 | 1.203377 | 1.433572 | 1.92603 | 4.92466 |
| 6 | 0.3 | 2.854277 | 2.360136 | 2.237467 | 0.601976 | 0.717197 | 0.963845 | 2.217333 |
| 6 | 0.5 | 0.143676 | 0.052516 | 0.039857 | 0.000576 | 0.000822 | 0.00166 | -0.48999 |
| 6 | 0.7 | -2.56692 | -2.2551 | -2.15775 | -0.60083 | -0.71555 | -0.96053 | -3.19732 |
| 6 | 0.9 | -5.27752 | -4.56272 | -4.35536 | -1.20223 | -1.43193 | -1.92271 | -5.90465 |
| 8 | 0.1 | 5.558432 | 4.897909 | 4.600604 | 1.203348 | 1.433531 | 1.928584 | 5.018271 |
| 8 | 0.3 | 2.83494 | 2.475233 | 2.318767 | 0.60189 | 0.717074 | 0.964917 | 2.246085 |
| 8 | 0.5 | 0.111448 | 0.052557 | 0.036931 | 0.000432 | 0.000617 | 0.00125 | -0.5261 |
| 8 | 0.7 | -2.61204 | -2.37012 | -2.24491 | -0.60103 | -0.71584 | -0.96242 | -3.29829 |
| 8 | 0.9 | -5.33554 | -4.7928 | -4.52674 | -1.20248 | -1.4323 | -1.92608 | -6.07047 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 10 | 0.1 | 5.554355 | 5.076429 | 4.719326 | 1.203331 | 1.433506 | 1.93012 | 5.079791 |
| 10 | 0.3 | 2.822708 | 2.564505 | 2.376862 | 0.601838 | 0.717 | 0.965561 | 2.262582 |
| 10 | 0.5 | 0.091062 | 0.052582 | 0.034398 | 0.000345 | 0.000493 | 0.001002 | -0.55463 |
| 10 | 0.7 | -2.64058 | -2.45934 | -2.30807 | -0.60115 | -0.71601 | -0.96356 | -3.37184 |
| 10 | 0.9 | -5.37223 | -4.97127 | -4.65053 | -1.20264 | -1.43252 | -1.92812 | -6.18904 |

**Table 7:** Values of $\log(\psi)$ for $M_3$ when $\sigma_y^2 = 2$, $\mu_y = 5$, $\sigma_t^2 = 5$, $\alpha = 25$, $\beta = 25$, $\gamma = 5$, $A = 0.7$, $P = 0.4$, $W = 0.2$, $P_1 = 0.4$, $P_2 = 0.6$.

| $\sigma_s^2$ | $w_1$ | $\log\left(\psi_{(W/M_3)}\right)$ | $\log\left(\psi_{(EH/M_3)}\right)$ | $\log\left(\psi_{(DP/M_3)}\right)$ | $\log\left(\psi_{(GS/M_3)}\right)$ | $\log\left(\psi_{(M/M_3)}\right)$ | $\log\left(\psi_{(NS/M_3)}\right)$ | $\log\left(\psi_{(G/M_3)}\right)$ |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.1 | 7.77993 | 4.34238 | 4.330598 | 2.560575 | 2.636839 | 2.88727 | 5.125718 |
| 2 | 0.3 | 4.06323 | 2.17485 | 2.168906 | 1.280671 | 1.318843 | 1.44422 | 2.21331 |
| 2 | 0.5 | 0.34654 | 0.00732 | 0.007214 | 0.000767 | 0.000847 | 0.001169 | -0.6991 |
| 2 | 0.7 | -3.37015 | -2.16021 | -2.15448 | -1.27914 | -1.31715 | -1.44188 | -3.6115 |
| 2 | 0.9 | -7.08685 | -4.32774 | -4.31617 | -2.55904 | -2.63515 | -2.88493 | -6.52391 |
| 4 | 0.1 | 7.75116 | 4.89690 | 4.873504 | 2.560498 | 2.636754 | 2.927131 | 5.642591 |
| 4 | 0.3 | 3.97694 | 2.45212 | 2.440316 | 1.280441 | 1.318589 | 1.463873 | 2.469606 |
| 4 | 0.5 | 0.20271 | 0.00733 | 0.007127 | 0.000384 | 0.000424 | 0.000616 | -0.70338 |
| 4 | 0.7 | -3.57151 | -2.43745 | -2.42606 | -1.27967 | -1.31774 | -1.46264 | -3.87636 |
| 4 | 0.9 | -7.34573 | -4.88223 | -4.85925 | -2.55973 | -2.63591 | -2.9259 | -7.04935 |
| 6 | 0.1 | 7.73938 | 5.22127 | 5.186429 | 2.560472 | 2.636726 | 2.940872 | 5.931001 |
| 6 | 0.3 | 3.94160 | 2.61431 | 2.59673 | 1.280364 | 1.318504 | 1.470645 | 2.611833 |
| 6 | 0.5 | 0.14383 | 0.00734 | 0.007032 | 0.000256 | 0.000282 | 0.000418 | -0.70734 |
| 6 | 0.7 | -3.65395 | -2.59963 | -2.58267 | -1.27985 | -1.31794 | -1.46981 | -4.0265 |
| 6 | 0.9 | -7.45173 | -5.20659 | -5.17237 | -2.55996 | -2.63616 | -2.94004 | -7.34567 |
| 8 | 0.1 | 7.73293 | 5.45142 | 5.405287 | 2.56046 | 2.636712 | 2.947832 | 6.126726 |
| 8 | 0.3 | 3.92225 | 2.72938 | 2.706112 | 1.280326 | 1.318462 | 1.474074 | 2.707865 |
| 8 | 0.5 | 0.11156 | 0.00734 | 0.006936 | 0.000192 | 0.000212 | 0.000316 | -0.711 |
| 8 | 0.7 | -3.69912 | -2.7147 | -2.69224 | -1.27994 | -1.31804 | -1.47344 | -4.12986 |
| 8 | 0.9 | -7.50981 | -5.43674 | -5.39141 | -2.56008 | -2.63629 | -2.9472 | -7.54872 |
| 10 | 0.1 | 7.72885 | 5.62994 | 5.572672 | 2.560452 | 2.636703 | 2.952037 | 6.272234 |
| 10 | 0.3 | 3.91000 | 2.81864 | 2.789758 | 1.280303 | 1.318436 | 1.476146 | 2.778922 |
| 10 | 0.5 | 0.09115 | 0.00734 | 0.006843 | 0.000153 | 0.000169 | 0.000254 | -0.71439 |
| 10 | 0.7 | -3.7277 | -2.80395 | -2.77607 | -1.28 | -1.3181 | -1.47564 | -4.2077 |
| 10 | 0.9 | -7.54655 | -5.61525 | -5.55899 | -2.56015 | -2.63636 | -2.95153 | -7.70101 |

**Table 8:** Values of $\log(\psi)$ for $M_1$ *vs* $M_3$, $M_2$ *vs* $M_3$ and $M_1$ *vs* $M_2$ when $\sigma_y^2 = 2$, $\mu_y = 2$, $\sigma_t^2 = 3$, $\alpha = 15$, $\beta = 15$, $\gamma = 510$ $A = 0.7$, $P = 0.6$, $W = 0.2$.

| $\sigma_s^2$ | $w_1$ | $M_1$ *vs* $M_3$ | $M_2$ *vs* $M_3$ | $M_1$ *vs* $M_2$ |
|---|---|---|---|---|
| 2 | 0.05 | 4.623235 | -0.25893 | 4.882161 |
| 2 | 0.15 | 3.612867 | -0.20141 | 3.814282 |
| 2 | 0.3 | 2.097316 | -0.11515 | 2.212463 |
| 2 | 0.45 | 0.581765 | -0.02888 | 0.610645 |
| 2 | 0.6 | -0.93379 | 0.057388 | -0.99117 |
| 2 | 0.75 | -2.44934 | 0.143656 | -2.59299 |
| 2 | 0.9 | -3.96489 | 0.229924 | -4.19481 |

| | | | | |
|---|---|---|---|---|
| 4 | 0.05 | 4.619554 | -0.25892 | 4.878474 |
| 4 | 0.15 | 3.601825 | -0.2014 | 3.803221 |
| 4 | 0.3 | 2.075232 | -0.11511 | 2.190342 |
| 4 | 0.45 | 0.548639 | -0.02882 | 0.577463 |
| 4 | 0.6 | -0.97795 | 0.057462 | -1.03542 |
| 4 | 0.75 | -2.50455 | 0.143748 | -2.6483 |
| 4 | 0.9 | -4.03114 | 0.230035 | -4.26118 |
| 6 | 0.05 | 4.618263 | -0.25892 | 4.877181 |
| 6 | 0.15 | 3.597954 | -0.20139 | 3.799343 |
| 6 | 0.3 | 2.067489 | -0.1151 | 2.182586 |
| 6 | 0.45 | 0.537024 | -0.02881 | 0.565829 |
| 6 | 0.6 | -0.99344 | 0.057487 | -1.05093 |
| 6 | 0.75 | -2.52391 | 0.143779 | -2.66768 |
| 6 | 0.9 | -4.05437 | 0.230072 | -4.28444 |
| 8 | 0.05 | 4.617605 | -0.25892 | 4.876522 |
| 8 | 0.15 | 3.595979 | -0.20139 | 3.797366 |
| 8 | 0.3 | 2.06354 | -0.11509 | 2.178631 |
| 8 | 0.45 | 0.531101 | -0.0288 | 0.559897 |
| 8 | 0.6 | -1.00134 | 0.057499 | -1.05884 |
| 8 | 0.75 | -2.53378 | 0.143795 | -2.67757 |
| 8 | 0.9 | -4.06622 | 0.23009 | -4.29631 |
| 10 | 0.05 | 4.617206 | -0.25892 | 4.876122 |
| 10 | 0.15 | 3.594782 | -0.20138 | 3.796167 |
| 10 | 0.3 | 2.061145 | -0.11509 | 2.176233 |
| 10 | 0.45 | 0.527508 | -0.02879 | 0.556299 |
| 10 | 0.6 | -1.00613 | 0.057507 | -1.06364 |
| 10 | 0.75 | -2.53976 | 0.143804 | -2.68357 |
| 10 | 0.9 | -4.0734 | 0.230101 | -4.3035 |

**Table 9:** Values of $\log(\psi)$ for $M_1$ *vs* $M_3$ , $M_2$ *vs* $M_3$ and $M_1$ *vs* $M_2$ when $\sigma_y^2 = 2$ , $\mu_y = 5$ , $\sigma_t^2 = 5$ , $\alpha = 25,\ \beta = 25,\ \gamma = 5$ $A = 0.7$ , $P = 0.4$ , $W = 0.2$ .

| $\sigma_s^2$ | $w_1$ | $M_1$ *vs* $M_3$ | $M_2$ *vs* $M_3$ | $M_1$ *vs* $M_2$ |
|---|---|---|---|---|
| 2 | 0.05 | 5.710012 | -0.08578 | 5.795792 |
| 2 | 0.15 | 4.445154 | -0.06672 | 4.511872 |
| 2 | 0.3 | 2.547867 | -0.03813 | 2.585993 |
| 2 | 0.45 | 0.65058 | -0.00953 | 0.660114 |
| 2 | 0.6 | -1.24671 | 0.019058 | -1.26577 |
| 2 | 0.75 | -3.14399 | 0.047651 | -3.19164 |
| 2 | 0.9 | -5.04128 | 0.076243 | -5.11752 |
| 4 | 0.05 | 5.709113 | -0.08578 | 5.794892 |
| 4 | 0.15 | 4.442457 | -0.06672 | 4.509174 |
| 4 | 0.3 | 2.542472 | -0.03812 | 2.580597 |
| 4 | 0.45 | 0.642487 | -0.00953 | 0.652019 |
| 4 | 0.6 | -1.2575 | 0.01906 | -1.27656 |
| 4 | 0.75 | -3.15748 | 0.047653 | -3.20514 |
| 4 | 0.9 | -5.05747 | 0.076245 | -5.13371 |
| 6 | 0.05 | 5.70881 | -0.08578 | 5.794589 |
| 6 | 0.15 | 4.441546 | -0.06672 | 4.508264 |
| 6 | 0.3 | 2.540651 | -0.03812 | 2.578776 |

| | | | | |
|---|---|---|---|---|
| 6 | 0.45 | 0.639756 | -0.00953 | 0.649288 |
| 6 | 0.6 | -1.26114 | 0.019061 | -1.2802 |
| 6 | 0.75 | -3.16203 | 0.047654 | -3.20969 |
| 6 | 0.9 | -5.06293 | 0.076246 | -5.13918 |
| 8 | 0.05 | 5.708657 | -0.08578 | 5.794437 |
| 8 | 0.15 | 4.441089 | -0.06672 | 4.507806 |
| 8 | 0.3 | 2.539737 | -0.03812 | 2.577861 |
| 8 | 0.45 | 0.638384 | -0.00953 | 0.647916 |
| 8 | 0.6 | -1.26297 | 0.019061 | -1.28203 |
| 8 | 0.75 | -3.16432 | 0.047654 | -3.21197 |
| 8 | 0.9 | -5.06567 | 0.076247 | -5.14192 |
| 10 | 0.05 | 5.708566 | -0.08578 | 5.794345 |
| 10 | 0.15 | 4.440814 | -0.06672 | 4.507531 |
| 10 | 0.3 | 2.539187 | -0.03812 | 2.577311 |
| 10 | 0.45 | 0.637559 | -0.00953 | 0.647091 |
| 10 | 0.6 | -1.26407 | 0.019061 | -1.28313 |
| 10 | 0.75 | -3.1657 | 0.047654 | -3.21335 |
| 10 | 0.9 | -5.06732 | 0.076247 | -5.14357 |

The numerical results presented in Tables 2–9 provide a comprehensive comparison of the proposed models with existing randomized response techniques using the logarithmic performance measure $\log(\psi)$. A consistent pattern emerges across all configurations: for smaller values of the weighting parameter, the proposed models yield positive values of $\log(\psi)$, indicating clear superiority over competing methods. As the weight increases, the performance advantage gradually diminishes, with a transition occurring in the mid-range where competing models begin to perform comparably or slightly better. This transition reflects the inherent trade-off between efficiency and privacy embedded in the design. The effect of the variance parameter is primarily to scale the magnitude of comparison without altering the overall ranking structure.

A more detailed examination reveals that while certain existing models remain competitive over limited regions, the proposed framework offers broader and more stable dominance across practical parameter ranges. Importantly, the pairwise comparisons among the proposed models (Tables 8 and 9) demonstrate a clear and consistent hierarchy, with $M_1$ emerging as the most robust performer, followed by $M_3$, while $M_2$ exhibits relatively weaker performance. This internal consistency further validates the flexibility and effectiveness of the generalized framework, highlighting its ability to generate models with varying strengths suited to different design priorities.

## 7. Conclusion and future directions

This study develops a generalized framework for randomized response scrambling models, within which a broad class of existing designs can be expressed and systematically compared. As special cases of this unified structure, three new models, $M_1$, $M_2$, and $M_3$, are proposed and examined in detail. To facilitate a meaningful comparison that accounts for both efficiency and privacy, a logarithmic performance measure, $\log(\psi)$, is introduced. The graphical and analytical results show that the proposed models exhibit distinct but complementary strengths: $M_2$ and $M_3$ tend to perform better for smaller values of the weighting parameter, while $M_1$ becomes increasingly preferable as the weight assigned to the combined performance criterion grows. This behavior highlights that no single model is uniformly optimal; rather, the choice of model should depend on the underlying design priorities. Overall, the proposed framework not only unifies several existing approaches but also provides flexibility, improved interpretability, and practical guidance for selecting an appropriate scrambling mechanism under varying conditions.

There are several directions in which this work can be extended. First, the proposed generalized framework can be further explored by identifying additional classes of scrambling models as special cases, which may offer improved performance under specific design settings. Second, a more rigorous theoretical investigation of the proposed performance measure $\log(\psi)$, particularly its properties as a divergence-type criterion, could provide deeper insight into model comparison and optimality. Third, the framework may be extended to more complex survey settings, such as multi-sensitive attributes, stratified sampling, or adaptive designs, where the balance between efficiency and privacy becomes even more critical. In addition, simulation studies under a wider range of distributions and real data applications would help assess the practical

robustness of the proposed models. Finally, developing formal decision rules or optimization strategies for selecting the most appropriate model based on design parameters remains an important area for future research.

## References


1. Miller, J. D. (1984). *A new survey technique for studying deviant behavior*. Unpublished Ph.D. dissertation, George Washington University. Department of Sociology.
2. Sirken, M. G. (1970). Household surveys with multiplicity. *Journal of the American statistical Association*, *65*(329), 257-266.
3. Raghavarao, D., & Federer, W. T. (1979). Block total response as an alternative to the randomized response method in surveys. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *41*(1), 40-45.
4. Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American statistical association*, 60(309), 63-69.
5. Warner, S. L. (1971). The linear randomized response model. *Journal of the American Statistical Association*, 66(336), 884-888.
6. Greenberg, B. G., Abul-Ela, A. L. A., Simmons, W. R., & Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, *64*(326), 520-539.
7. Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The unrelated question randomized response model with unequal probabilities. *Journal of the American Statistical Association*, *71*(354), 311–320.
8. Kuk, A. Y. (1990). Asking sensitive questions indirectly. *Biometrika*, 77(2), 436-438.
9. Mangat, N. S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1), 93-95.
10. Christofides, T. C. (2003). A generalized randomized response technique. *Metrika*, 57, 195-200.
11. Kim, J. J., & Warde, W. D. (2005). A stratified randomized response model. *Journal of Statistical Planning and Inference*, *130* (2), 341–349.
12. Arnab, R., & Singh, S**.** (2006). An improved quantitative randomized response model. *Journal of Statistical Planning and Inference*, *136* (8), 2820–2830
13. Gupta, S., Shabbir, J., & Sehra, S. K. (2002). A new randomized response model for quantitative data. *Journal of Statistical Theory and Practice*, *2* (4), 449–458.
14. Eichhorn, B. H., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and inference*, 7 (4), 307-316.
15. Le, T. N., Lee, S. M., Tran, P. L., & Li, C. S. (2023). Randomized response techniques: a systematic review from the pioneering work of warner (1965) to the present. *Mathematics*, 11 (7), 1718.
16. Yan, Z., Wang, J., & Lai, J. (2008). An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics-Theory and Methods*, 38 (3), 400-408.
17. Gupta, S., Mehta, S., Shabbir, J., & Khalil, S. (2018). A unified measure of respondent privacy and model efficiency in quantitative RRT models. *Journal of Statistical Theory and Practice*, 12, 506-511.
18. Azeem, M. (2023). Introducing a weighted measure of privacy and efficiency for comparison of quantitative randomized response models. *Pakistan Journal of Statistics*, 39 (3), 377-385.
19. Eichhorn, B. H., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and inference*, 7 (4), 307-316.
20. Diana, G., & Perri, P. F. (2011). A class of estimators for quantitative sensitive data. *Statistical Papers*, 52, 633-650.
21. Gjestvang, C. R., & Singh, S. (2009). An improved randomized response model: Estimation of mean. *Journal of Applied Statistics*, 36 (12), 1361-1367.
22. Murtaza, M., Singh, S., & Hussain, Z. (2020). An innovative optimal randomized response model using correlated scrambling variables. *Journal of Statistical Computation and Simulation*, 90 (15), 2823-2839.
23. Narjis, G., & Shabbir, J. (2021). An efficient new scrambled response model for estimating sensitive population mean in successive sampling. *Communications in Statistics-Simulation and Computation*, 52 (11), 5327-5344.
24. Gupta, S., Zhang, J., Khalil, S., & Sapra, P. (2024). Mitigating lack of trust in quantitative randomized response technique models. *Communications in Statistics-Simulation and Computation*, 53 (6), 2624-2632.


## Declaration


**Conflict of Study:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Funding Statement:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author Contribution Statement**: Both the authors worked together as a cell group and designed the research; Analyzed and interpreted the data, and wrote the paper.

**Availability of Data and Material**: Data will be made available by the corresponding author on reasonable request.


## Author(s) Bio / Authors' Note

***Shoaib Iqbal*** *is from the Department of Statistics, Faculty of Computing, The Islamia University of Bahawalpur, 63100, Pakistan. Email:* shoaibiqbalqau654@gmail.com

***Zawar Hussain*** *is from the Department of Statistics, Faculty of Computing, The Islamia University of Bahawalpur, 63100, Pakistan. Email:* zhlangah@yahoo.com

## Appendix

***Derivation of estimator and its variance under*** $M_1$

The motivation for $M_1$ comes from the randomized response techniques developed by Warner [5] and Gjestvang and Singh [38]. The Model 1 we are proposing here uses a more straightforward one-stage randomized response approach using multiplicative scrambling.

Let us assume that there are n number of respondents in a sample selected from the population. Each respondent is given a randomization device to generate two outcomes, each occurring with known probability *P* and *(1-P)*. If the outcome 1 appears, then $Y(1+S)$ will be reported by the respondent, while if the second outcome, then $Y(1-S)$ will be reported by the respondent. The response of the $i^{th}$ respondent through $M_1$ is:

$$R_{M_{1i}} = D_{1i}\left(Y_i + Y_i S_i\right) + \left(1 - D_{2i}\right)\left(Y_i - Y_i S_i\right).$$

The distribution of this response may be written as

$$R_{M_{1i}} = \begin{cases} Y_i\left(1+S_i\right) & \text{with probability } P_1 \\ Y_i\left(1-S_i\right) & \text{with probability } 1-P_1 \end{cases}$$

Now, the expected response is given by

$$E\left(R_{M_{1i}}\right) = E\left(D_{1i}\right)\left\{E\left(Y_i\right) + E\left(Y_i\right)E\left(S_i\right)\right\} + E\left(1-D_{1i}\right)\left\{E\left(Y_i\right) - E\left(Y_i\right)E\left(S_i\right)\right\}$$

$$E\left(R_{M_{1i}}\right) = \left(P_1\right)E\left(Y_i\right) + \left(1-P_1\right)E\left(Y_i\right)$$

$$E\left(R_{M_{1i}}\right) = E\left(Y_i\right) = \mu_y$$

So, by methods of moment estimation, an unbiased mean estimator of $\mu_y$ is given by:

$$\hat{\mu}_{M_1} = \frac{1}{n}\sum_{i=1}^{n} R_{M_{1i}}.$$

Now, by definition, variance of the above estimator is given by

$$Var\left(\hat{\mu}_{M_1}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} R_{M_{1i}}\right) = \frac{Var\left(R_{M_{1i}}\right)}{n}.$$

where $Var\left(R_{M_{1i}}\right)$ is given by

$$Var\left(R_{M_{1i}}\right) = E\left(R_{M_{1i}}^2\right) - \left[E\left(R_{M_{1i}}\right)\right]^2.$$

Now, consider

$$\left(R_{M_{1i}}^2\right) = \left[D_{1i}\left(Y_i + Y_i S_i\right)_i + \left(1-D_{1i}\right)\left(Y_i - Y_i S_i\right)\right]^2$$

After applying expectation,

$$E\left(R_{M_{1i}}^2\right) = E\left[D_{1i}\left(Y_i + Y_i S_i\right) + \left(1-D_{1i}\right)\left(Y_i - Y_i S_i\right)\right]^2$$

$$E\left(R_{M_{1i}}^2\right) = E\left(D_{1i}^2\right)\left\{E\left(Y_i^2\right) + E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\} + E\left(1-D_{1i}\right)^2\left[\left\{E\left(Y_i^2\right) + E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\}\right.$$

$$E\left(R_{M_{1i}}^2\right) = \left(P_1\right)\left\{E\left(Y_i^2\right) + E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\} + \left(1-P_1\right)\left[\left\{E\left(Y_i^2\right) + E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\}\right.$$

$$E\left(R_{M_{1i}}^2\right) = \left(\sigma_y^2 + \mu_y^2\right) + \sigma_s^2\left(\sigma_y^2 + \mu_y^2\right)$$

Thus, the variance of $R_{M_{1i}}$ is given by

$$Var\left(R_{M_{1i}}\right)=E\left(R_{M_{1i}}^2\right)-\left(E\left(R_{M_{1i}}\right)\right)^2$$
$$=\left(\sigma_y^2+\mu_y^2\right)+\sigma_s^2\left(\sigma_y^2+\mu_y^2\right)-\mu_y^2$$
$$=\sigma_y^2+\sigma_s^2\left(\sigma_y^2+\mu_y^2\right)$$

Hence, the variance of $\hat{\mu}_{M_1}$ is given by:

$$Var\left(\hat{\mu}_{M_1}\right)=\frac{1}{n}\left[\ \sigma_y^2+\sigma_s^2\mu_{2y}'\ \right]$$

***Derivation of estimator and its variance under*** $M_2$

The response of the $i^{th}$ respondent through $M_2$ is:

$$R_{M_{2i}}=D_{2i}\left(Y_i+\alpha Y_iS_i\right)+\left(1-D_{2i}\right)\left(Y_i-\beta Y_iS_i\right)$$

The distribution of $i^{th}$ respondent response under the $M_2$ is given as:

$$R_{M_{2i}}=\begin{cases} Y_i\left(1+\alpha S_i\right) & \text{with probability } P_2=\dfrac{\alpha}{\alpha+\beta} \\ Y_i\left(1-\beta S_i\right) & \text{with probability } \left(1-P_2\right)=\dfrac{\beta}{\alpha+\beta} \end{cases}$$

Now, the expected response is given by

$$E\left(R_{M_{2i}}\right)=E\left(D_{2i}\right)\left\{E\left(Y_i\right)+\alpha E\left(Y_i\right)E\left(S_i\right)\right\}+E\left(1-D_{2i}\right)\left\{E\left(Y_i\right)-\beta E\left(Y_i\right)E\left(S_i\right)\right\}$$
$$E\left(R_{M_{2i}}\right)=\left(P_2\right)E\left(Y_i\right)+\left(1-P_2\right)E\left(Y_i\right)$$
$$E\left(R_{M_{2i}}\right)=E\left(Y_i\right)=\mu_y$$

So, by methods of moment estimation, an unbiased mean estimator of $\mu_y$ is given by:

$$\hat{\mu}_{M_2}=\frac{1}{n}\sum_{i=1}^{n}R_{M_{2i}}\ .$$

Now, by definition, variance of the above estimator is given by

$$Var\left(\hat{\mu}_{M_2}\right)=Var\left(\frac{1}{n}\sum_{i=1}^{n}R_{M_{2i}}\right)=\frac{Var\left(R_{M_{2i}}\right)}{n}.$$

where $Var\left(R_{M_{2i}}\right)$ is given by

$$Var\left(R_{M_{2i}}\right)=E\left(R_{M_{2i}}^2\right)-\left[E\left(R_{M_{2i}}\right)\right]^2.$$

Now, consider

$$\left(R_{M_{2i}}^2\right)=\left[D_{2i}\left(Y_i+\alpha Y_iS_i\right)+\left(1-D_{2i}\right)\left(Y_i-\beta Y_iS_i\right)\right]^2$$

After applying expectation,

$$E\left(R_{M_{2i}}^2\right)=E\left[D_{2i}\left(Y_i+\alpha Y_iS_i\right)+\left(1-D_{2i}\right)\left(Y_i-\beta Y_iS_i\right)\right]^2$$
$$E\left(R_{M_{2i}}^2\right)=E\left(D_{2i}^2\right)\left\{E\left(Y_i^2\right)+\alpha^2E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\}+E\left(1-D_{2i}\right)^2\left[\left\{E\left(Y_i^2\right)+\beta^2E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\}\right.$$
$$E\left(R_{M_{2i}}^2\right)=\left(P_2\right)\left\{E\left(Y_i^2\right)+\alpha^2E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\}+\left(1-P_2\right)\left[\left\{E\left(Y_i^2\right)+\beta^2E\left(Y_i^2\right)E\left(S_i^{\,2}\right)\right\}\right.$$
$$E\left(R_{M_{2i}}^2\right)=\left(\sigma_y^2+\mu_y^2\right)+\left(\sigma_y^2+\mu_y^2\right)\left(P_2\alpha^2+\left(1-P_2\right)\beta^2\right)\sigma_s^2$$

Thus, the variance of $R_{M_{2i}}$ is given by

$$Var\left(R_{M_{2i}}\right)=E\left(R_{M_{2i}}^2\right)-\left(E\left(R_{M_{2i}}\right)\right)^2$$
$$=\left(\sigma_y^2+\mu_y^2\right)+\left(\sigma_y^2+\mu_y^2\right)\left(P_2\alpha^2+\left(1-P_2\right)\beta^2\right)\sigma_s^2-\mu_y^2$$
$$=\sigma_y^2+\mu_{2y}'\left(P_2\alpha^2+\left(1-P_2\right)\beta^2\right)\sigma_s^2$$

Hence, the variance of $\hat{\mu}_{M_2}$ is given by:

$$Var\left(\hat{\mu}_{M_2}\right)=\frac{1}{n}\left[\ \sigma_y^2+\mu_{2y}'\left(P_2\alpha^2+\left(1-P_2\right)\beta^2\right)\sigma_s^2\right]$$

***Derivation of estimator and its variance under*** $M_3$

The $i^{th}$ respondent's response may be written as

$$R_{M_{3i}}=D_{3i}Y_i\left(1+\alpha S_i\right)+D_{4i}Y_i\left(1-\beta S_i\right)+\left(1-D_{2i}-D_{3i}\right)\left(Y_i\right)$$

And its distribution is given by

$$R_{M_{3i}}=\begin{cases}Y_i\left(1+\alpha S_i\right) & \text{with probability } p_1=\dfrac{\alpha}{\alpha+\beta+\gamma}\\ Y_i\left(1-\beta S_i\right) & \text{with probability } p_2=\dfrac{\beta}{\alpha+\beta+\gamma}\\ Y_i & \text{with probability } p_3=\dfrac{\gamma}{\alpha+\beta+\gamma}\end{cases}$$

Now, the expected response is given by

$$E\left(R_{M_{3i}}\right)=E\left(D_{3i}\right)\left\{E\left(Y_i\right)+\alpha E\left(Y_i\right)E\left(S_i\right)\right\}+E\left(D_{4i}\right)\left\{E\left(Y_i\right)-\beta E\left(Y_i\right)E\left(S_i\right)\right\}+E\left(1-D_{3i}-D_{4i}\right)E\left(Y_i\right)$$
$$E\left(R_{M_{3i}}\right)=\left(p_1\right)E\left(Y_i\right)+\left(p_2\right)E\left(Y_i\right)+\left(p_3\right)E\left(Y_i\right)$$
$$E\left(R_{M_{3i}}\right)=E\left(Y_i\right)=\mu_y$$

So, by methods of moment estimation, an unbiased mean estimator of $\mu_y$ is given by:

$$\hat{\mu}_{M_3}=\frac{1}{n}\sum_{i=1}^{n}R_{M_{3i}}\ .$$

Now, by definition, variance of the above estimator is given by

$$Var\left(\hat{\mu}_{M_3}\right)=Var\left(\frac{1}{n}\sum_{i=1}^{n}R_{M_{3i}}\right)=\frac{Var\left(R_{M_{3i}}\right)}{n}\ .$$

where $Var\left(R_{M_{3i}}\right)$ is given by

$$Var\left(R_{M_{3i}}\right)=E\left(R_{M3i}^2\right)-\left[E\left(R_{M_{3i}}\right)\right]^2\ .$$

Now, consider

$$\left(R_{M_{3i}}^2\right)=\left[D_{3i}Y_i\left(1+\alpha S_i\right)+D_{4i}Y_i\left(1-\beta S_i\right)+\left(1-D_{3i}-D_{4i}\right)\left(Y_i\right)\right]^2$$

After applying expectation,

$$E\left(R_{M_{3i}}^2\right)=E\left[D_{3i}\left(Y_i+\alpha Y_iS_i\right)+\left(D_{4i}\right)\left(Y_i-\beta Y_iS_i\right)+\left(1-D_{3i}-D_{4i}\right)\left(Y_i\right)\right]^2$$
$$E\left(R_{M_3}^2\right)=E\left(D_{31}^2\right)\left\{E\left(Y_i^2\right)+\alpha^2E\left(Y_i^2\right)E\left(S_i^2\right)\right\}+E\left(D_{4i}\right)^2\left[\left\{E\left(Y_i^2\right)+\beta^2E\left(Y_i^2\right)E\left(S_i^2\right)\right\}+E\left(1-D_{3i}-D_{4i}\right)^2E\left(Y_i^2\right)\right.$$
$$E\left(R_{M_{3i}}^2\right)=\left(p_1\right)\left\{E\left(Y_i^2\right)+\alpha^2E\left(Y_i^2\right)E\left(S_i^2\right)\right\}+\left(p_2\right)\left[\left\{E\left(Y_i^2\right)+\beta^2E\left(Y_i^2\right)E\left(S_i^2\right)\right\}+\left(p_3\right)E\left(Y_i^2\right)\right.$$
$$E\left(R_{M_{3i}}^2\right)=\left(\sigma_y^2+\mu_y^2\right)+\frac{\alpha\beta\left(\alpha+\beta\right)\left(\sigma_y{}^2+\mu_y{}^2\right)\sigma_s{}^2}{\alpha+\beta+\gamma}$$

Thus, the variance of $R_{M_{3i}}$ is given by

$$Var\left(R_{M_{3i}}\right) = E\left(R^2_{M_{3i}}\right) - \left(E\left(R_{M_{3i}}\right)\right)^2$$

$$= \left(\sigma_y^2 + \mu_y^2\right) + \frac{\alpha\beta\left(\alpha+\beta\right)\left(\sigma_y{}^2 + \mu_y{}^2\right)\sigma_s{}^2}{\alpha+\beta+\gamma} - \mu_y^2$$

$$= \sigma_y{}^2 + \frac{\alpha\beta\left(\alpha+\beta\right)\mu'_{2y}\sigma_s{}^2}{\alpha+\beta+\gamma}$$

Hence, the variance of $\hat{\mu}_{M_3}$ is given by:

$$Var\left(\hat{\mu}_{M_3}\right) = \frac{1}{n}\left[\ \sigma_y{}^2 + \frac{\alpha\beta\left(\alpha+\beta\right)\mu'_{2y}\sigma_s{}^2}{\alpha+\beta+\gamma}\ \right].$$

SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 2, June 2026
DOI: https://doi.org/10.64060/JASR2V2i4


# Bayesian Monitoring of Waiting-Time Processes using Exponential Distribution

Samina Kanwal [1], Muhammad Zafar Iqbal [1], Muhammad Sohaib Asad [1], Muhammad Kashif [1*]

[1]Department of Mathematics and Statistics, University of Agriculture Faisalabad, Pakistan

* Corresponding Email: mkashif@uaf.edu.pk

Received: 14 March 2026 / Revised: 04 May 2026 / Accepted: 10 May 2026 / Published online: 17 May 2026



ABSTRACT

The Bayesian control chart is an advanced statistical tool for monitoring process stability by incorporating prior knowledge with observed data. This study proposes a Bayesian control chart for monitoring waiting-time processes assuming an exponential distribution with a gamma conjugate prior. Posterior distributions of the process parameter are derived, and the posterior mean and standard deviation are used to construct dynamic control limits. A comprehensive simulation study is conducted by varying sample sizes and prior hyperparameters to evaluate the performance of the proposed chart. The results indicate that Bayesian control limits are generally narrower and more responsive to process shifts compared to classical control charts, leading to faster detection of out-of-control conditions. However, this increased sensitivity is associated with a higher false alarm rate, reflecting a trade-off between detection speed and in-control stability. The applicability of the proposed method is further demonstrated using real-life data, which supports the simulation findings. The study highlights the effectiveness of Bayesian approaches in process monitoring, particularly in situations with limited or uncertain data, while also emphasizing the importance of appropriate prior selection and control limit design. Future work may focus on developing probability-based or asymmetric control limits tailored for non-normal distributions.

**Keywords**:Control chart; Bayesian control chart; Gamma Distribution; Exponential Distribution

## 1. Introduction

The two primary categories for control charting systems are classical and Bayesian. In contrast to traditional control structures, control charting structures based on Bayesian approaches are considered when the process parameters are unclear. Usually, previous knowledge is used to quantify parametric uncertainty, which the Bayesian setup effectively manages. The practical challenge with the Bayesian technique is that it requires more precise and easily accessible information about the process structure.

A control chart (CC) is a statistical technique designed to verify and then track the statistical stability of a process, often permitting graphical application. A control chart's primary function is to keep an eye on the process throughout the manufacturing phase (Abujiya and Muttlak, 2004). The manufacturing process keeps producing goods if the procedure is managed, according to the control chart. In this case, no more action is needed. The required actions are taken by the quality engineers to regain control of the process. Whether any movement in the process is depicted on the control chart (Hajej et al., 2021). The control chart serves as a tool for determining when corrective action is necessary. Consequently, the CC serves as the tool for assessing the necessity of corrective actions (Roberts, 1966). The CL, UCL, and LCL are the three main components of a univariate Shewhart chart (Bakir, 2004). Statistical methods for quality control are essential in many industrial and service sectors. Control of statistical quality is a subfield of manufacturing data mainly concerned with SPC, capacity analysis, design of experiments, and acceptance sampling. To avoid the process of overreacting or underreacting, one of the primary functions of control plots is to differentiate between differences resulting from assignable sources and dissimilarities resulting from random causes (Ali et al., 2016). The collection of data points about a particular process or product attribute is the foundation of control charts. These data points might be metrics like weights, counts, or dimensions, and they are usually collected regularly (Koutras et al., 2007). Control limits specify the range in which the process should

function normally. In general, they are set to three standard deviations of the process mean. Using data collected about the process under observation, control charts are created. Product weights, dimensions, defect counts, and any other pertinent quality attribute measures might be included in this data (Sullivan and Jones, 2002). More information on the control chart can be seen in (Aslam et al., 2015), (Riaz and Ali, 2016) and (Abid et al., 2018).

A prior distribution represents the beliefs, knowledge, or uncertainty about the parameters of a statistical model before observing any data (Gelman, 2002). The parameters of interest were random variables with a prior distribution (PD). PD was determined by the information available at the time (Singh et al., 2013). The prior appropriately could need a lot of data. Therefore, a major influence on the Bayesian approach's accuracy might come from the quality of the priors (Touw, 2009). The Bayesian analysis focuses on the fact that how the prior distribution affects the posterior (Ghaderinezhad and Ley, 2019). The Bayesian method bases its decision on the appropriate control strategy on the posterior likelihood, upgraded after each sample that the process is out of control by the application of the Bayes theorem (Makis, 2008). The economic-statistical design of the Bayesian control chart is based on a predictive distribution with two types of informative and non-informative prior distributions (Seirani et al., 2023). A Bayesian chart is comparable to a Shewhart chart in which the bounds are quickly modified following each sample to reflect the most recent information available. Using Bayesian approaches, the decision-maker can include historical knowledge about the process's current state, which might come from expert opinion, past observations, or possibly other unrelated data (Celano et al., 2008). At each phase, the decision-maker obtains the likelihood that the process will eventually stabilize. This framework draws inspiration from the real options methodology (Moltchanova, 2017). The flexibility of Bayesian control charts allows them to adjust to varying sample sizes and amounts of previous knowledge. When there is a lack of expert knowledge or previous data, Bayesian control charts can nevertheless offer valuable insights by utilizing the information at hand to increase monitoring accuracy (Rasay et al., 2024).

Most existing research (Kalisetti et al., 2024) on control charts has primarily focused on classical estimation methods for parameter determination and chart construction. However,(Supharakonsakun, 2024) comparatively limited attention has been devoted to Bayesian estimation approaches that incorporate prior knowledge through suitable prior distributions. (Ali et al., 2025) Evaluating the effectiveness of classical and Bayesian control charts. (Supharakonsakun, 2024) introduced an improved classical control chart for monitoring nonconformities via the empirical Bayes approach. (Labha et al., 2025) discussed a Bayesian Exponentially Weighted Moving Average (EWMA) control chart designed for monitoring the shape parameter of lifetime Pareto distribution processes. Recent developments, such as the study by (Nafisah et al., 2025), proposed a variable sample size-based EWMA control chart with an exponential scaling mechanism for production process monitoring. (Joghee and Mathew, 2025) developed process shifts in the case of Six-Sigma based Bayesian Control Charts. Then, (Kanwal and Abbas, 2025) introduced a comparative analysis of quantile-based control charts for the French distribution, highlighting the increasing interest in exploring alternative statistical methodologies for process monitoring.

In this context, the objective of the present work is to develop a Bayesian control chart for the exponential distribution based on a gamma prior, and to evaluate its performance by comparing it with the traditional classical method. The exponential distribution is widely used in reliability analysis and lifetime data modeling, making it a natural candidate for quality control applications where failure times or inter-arrival times are of interest. The gamma prior, on the other hand, provides a mathematically convenient and flexible choice for Bayesian analysis, allowing incorporation of prior beliefs about the process parameter. This study not only aims to demonstrate the advantages of Bayesian approaches in terms of incorporating prior information but also to provide deeper insights into the efficiency and robustness of Bayesian control charts in real-world applications.

Although the Gamma–Exponential conjugate analysis is well known in Bayesian inference, its explicit use for constructing adaptive Shewhart-type control charts for exponential lifetime data has received very limited attention. Existing SPC literature primarily relies on classical estimators, and very few studies provide a complete Bayesian framework with dynamically updated limits. The present work fills this gap by integrating posterior updating into control limit calculation and systematically evaluating the effect of prior parameters and sample size on chart performance.

## 2. Methodology

### *2.1 Exponential Distribution*

The single-parameter exponential distribution serves as an educational tool for demonstrating parameter estimation methods (Elfessi and Reineke, 2001). The exponential density serves as a fundamental starting point for studying more general distributions because its analytical adjustments remain easy to perform. Using mathematical operations because they are straightforward enables us to explore and better understand the relationship between classical and Bayesian estimation methods (Rasheed and Abd, 2023).

The random variable follows a one-parameter Exponential Distribution where θ (θ>0) represents the parameter value in the associated Probability Density function for the continuous random variable X.

$$f(x,\theta) = \theta e^{-\theta x};\ 0< x < \infty \tag{1}$$

### *2.2 Gamma Distribution*

The GD is a continuous probability distribution. The probability of waiting for a firm amount of time before an event occurs. The gamma distribution predicts the wait time until future events. It has a connection to the normal distribution, the ED, and the Chi-squared distribution. "Γ" represents the Gamma function. A GD random variable θ with shape parameter "a" and rate parameter "b" (Son and Oh, 2006).

$$\theta \sim \Gamma(a,b) = Gamma(a,b) \tag{2}$$

In this study, the utilization of an informative conjugate prior is deemed suitable for the Bayesian approach. The corresponding PDF in the shape-rate parameterization is

$$f\left(\theta \mid a,b\right) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}; \quad \theta > 0 \ and \ a,b > 0 \tag{3}$$

The posterior distribution can be derived as follows:

$$P(\theta/x) = \frac{p(\theta).L(x/\theta)}{\int_{\theta} p(\theta).L(x/\theta)d\theta} \tag{4}$$

Where $L(x;\theta)$ represents the likelihood function of the Exponential probability function. Therefore, the Posterior distribution can be presented as follows:

$$f(\theta/x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \tag{5}$$

The Posterior distribution of the parameter $\theta$ is a gamma distribution, denoted as Gamma (a+n, $b+\sum_{i=1}^{n} x_i$ ). It can be expressed in the following form:

$$f(\theta/x) = \frac{(b+\sum_{i=1}^{n} x_i)^{a+n}}{\Gamma(a+n)}\theta^{(a+n)-1}e^{-(b+\sum_{i=1}^{n} x_i)\theta} \tag{6}$$

is the posterior distribution of the Gamma Prior Distribution for the case of Exponential distribution.

***2.3 Bayes Estimation***

The estimation of the parameter θ of the posterior distribution is determined such as Posterior Mean and Posterior Variance

2.3.1 Posterior Mean

The Posterior mean is defined as

$$mean = \mu = E(\theta/x) \tag{7}$$

As we know that the formula of expectation is

$$E(\theta/x) = \int_0^{\infty} \theta.f(\theta/x)d\theta$$

$$E(\theta/x) = \int_0^{\infty} \theta.\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\theta$$

$$E(\theta/x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}(\frac{v}{\beta})^{(\alpha+1)-1}e^{-v}\frac{dv}{\beta}$$

$$E(\theta/x) = \frac{\alpha}{\beta} \tag{8}$$

Hence, the mean of the posterior distribution is

$$where \ \alpha = a+n \quad and \quad \beta = b+\sum_{i=1}^{n} x_i$$

$$E(\theta/x) = \frac{a+n}{b+\sum_{i=1}^{n} x_i} \tag{9}$$

2.3.2 Posterior Variance

The Posterior Variance is defined as

$$\text{var}(\theta) = \sigma^2 = E(\theta^2/x) \tag{10}$$

As we know that the formula of variance is

$$\text{var}(\theta/x) = E(\theta^2/x) - [E(\theta/x)]^2$$

$$E(\theta^2/x) = \int_0^\infty \theta^2 . f(\theta/x) d\theta$$

Now,

$$E(\theta^2/x) = \int_0^\infty \theta^2 . \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta$$

$$E(\theta^2/x) = \frac{(a+n+1)a+n}{(b+\sum_{i=1}^{n} x_i)^2}$$

Put this in the variance formula,

$$var(\theta/x) = \frac{\alpha(\alpha+1)}{\beta^2} - \left[\frac{\alpha}{\beta}\right]^2$$

Hence, the variance of the posterior distribution is

$$var(\theta/x) = \frac{a+n}{(\text{b}+\sum_{i=1}^{n} x_i)^2} \tag{11}$$

3.3.3 Posterior Standard Deviation

The Standard Deviation of posterior distribution is

$$S.D(\theta/x) = \sqrt{\frac{a+n}{(\text{b}+\sum_{i=1}^{n} x_i)^2}} \tag{12}$$

### *2.4 Control Limits*

Establishing control limits is typically based on estimating the Posterior mean and posterior standard deviation. The control limit of the posterior distribution is defined as follows:

2.4.1 Lower Control Limit

The Lower Control Limit of Posterior distribution is defined as

$$LCL = mean - 3\ S.D \tag{13}$$

$$LCL = \mu - 3\sigma$$

$$LCL = \frac{a+n}{\text{b}+\sum_{i=1}^{n} x_i} - 3\sqrt{\frac{a+n}{(\text{b}+\sum_{i=1}^{n} x_i)^2}} \tag{14}$$

2.4.2 Upper Control Limit

The Upper Control Limit of Posterior distribution is defined as

$$UCL = mean + 3\ S.D \tag{15}$$

$$UCL = \mu + 3\sigma$$

$$UCL = \frac{a+n}{\text{b}+\sum_{i=1}^{n} x_i} + 3\sqrt{\frac{a+n}{(\text{b}+\sum_{i=1}^{n} x_i)^2}} \tag{16}$$

2.4.3 Central Limit

The Central Limit of Posterior distribution is defined as

$$CL = mean$$
$$CL = \frac{a+n}{\mathrm{b}+\sum_{i=1}^{n} x_i} \tag{17}$$

In the proposed Bayesian control chart, the monitoring statistic is the posterior distribution of the exponential rate parameter θ rather than the predictive distribution of future observations. This choice is intentional and consistent with the class of parameter-based Bayesian control charts discussed in earlier studies (Makis, 2008; Celano et al., 2008; Seirani et al., 2023). For lifetime and reliability-type data, the rate parameter θ directly reflects the underlying failure behavior of the process; therefore, tracking its evolution provides a more stable and interpretable signal of process changes compared to charts based solely on raw observations, which may exhibit high variability. Posterior-based monitoring naturally incorporates prior knowledge and new information at every sampling stage, producing smoother control limits and reducing the likelihood of false alarms, particularly when sample sizes are small or moderate. By updating θ after each subgroup, the proposed chart focuses on detecting genuine shifts in the process failure rate, which aligns with the practical objectives of reliability-oriented process monitoring. The implications of this design choice and its relevance to exponential-based processes are now explicitly discussed to clarify the advantages of parameter-driven monitoring within a Bayesian framework.

**Algorithm for Simulation Study**

**Pseudo-Code**

**Step 1:** Select parameter values
- Choose rate parameter θ = 1, 1.5, 2.
- Choose sample sizes n = 25, 50, 100, 150
- Choose prior parameters α = 1, 1.5, 2 and β = 0.5, 1, 1.5, 2.

**Step 2:** Generate data
- Generate random samples X1,X2,...,Xn ~ Exponential (θ)

**Step 3:** Specify prior distribution
- Assume θ ~ Gamma (α,β)

**Step 4:** Compute posterior distribution
- Update parameters:
  - $\alpha* = \alpha+n$
  - $\beta* = \beta+\sum Xi$

**Step 5:** Compute posterior estimates
- Posterior Mean:

$$\mu=\alpha*/\beta*$$

- Posterior Variance:

$$\sigma^2 = \alpha*/ (\beta*)^2$$

- Posterior Standard Deviation:

$$\sigma = \sqrt{\sigma^2}$$

**Step 6:** Construct control limits
- Central Line (CL): μ
- Upper Control Limit (UCL): μ + 3σ
- Lower Control Limit (LCL): μ − 3σ

**Step 7:** Repeat simulation
- Repeat Steps 2–6 for all combinations of parameters

**Step 8:** Compare results
- Compare Bayesian control limits with classical control limits
- Evaluate performance using ARL measures

The flowchart of the methodology applied in the simulation study is shown in Figure 1.

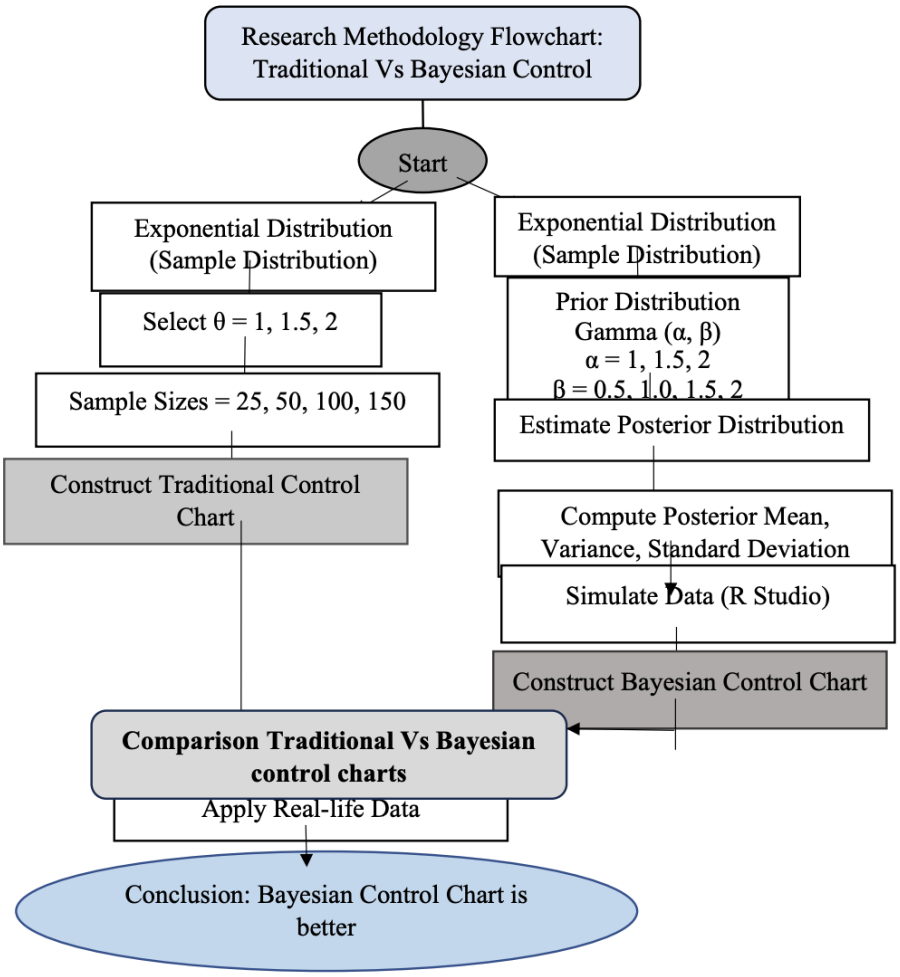


**Figure 1:** Flowchart of the methodologyy applied in the simulation study

## 3 Results

This study aims to develop and evaluate a Bayesian control chart specifically for the exponential distribution using a gamma prior. The research seeks to establish a more flexible and insightful approach to quality management by integrating Bayesian statistics with traditional control chart methodologies. This research aims to bridge the gap between traditional quality control methods and modern Bayesian statistical techniques by integrating prior knowledge and real-time data to enhance decision-making in process monitoring. Through comprehensive simulation studies utilizing R software, the research compares the performance of these Bayesian control charts with traditional control charts derived from frequentist methods. The numerical simulation study considers parameter values of n = 25, 50, 100, 150, and θ = 1, 1.5, 2.0. The hyper parameter values of the gamma prior for the proposed Bayesian method are set to (α, β) = (1, 1.5, and 2.0) and (0.5, 1.0, 1.5, and 2.0). These parameter values are applied in the proposed Bayesian method to determine the control limit such as UCL, LCL, and CL.

Although the control limits in this study are constructed using the conventional mean ± 3 standard deviation approach, it is important to note that the exponential distribution is positively skewed. Therefore, symmetric control limits may not be fully efficient for such distributions. The use of 3-sigma limits in this study is primarily intended to ensure comparability with classical control charts. Future research may consider the development of probability-based or quantile-based control limits that better accommodate the asymmetric nature of the exponential distribution.

**Table 1:** Control Limits for the Posterior Distribution of Exponential Distribution with θ =1 by using gamma prior with α=1 at different values of β

| α =1.5, θ =1 | | | | | | | |
|---|---|---|---|---|---|---|---|
| β | n | μ | $\sigma^2$ | σ | CL | UCL | LCL |
| 0.5 | 25 | 1.0981 | 0.0263 | 0.1622 | 1.0981 | 1.5849 | 0.6113 |
| | 50 | 1.1274 | 0.0378 | 0.1945 | 1.1274 | 1.7112 | 0.5437 |
| | 100 | 1.1672 | 0.0376 | 0.1939 | 1.1672 | 1.7489 | 0.5854 |
| | 150 | 1.1695 | 0.0467 | 0.2162 | 1.1695 | 1.8182 | 0.5208 |
| 1 | 25 | 1.0743 | 0.0252 | 0.1587 | 1.0743 | 1.5505 | 0.5980 |
| | 50 | 1.1030 | 0.0362 | 0.1903 | 1.1030 | 1.6741 | 0.5319 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 100 | 1.1418 | 0.0359 | 0.1897 | 1.1418 | 1.7110 | 0.5727 |
| | 150 | 1.1441 | 0.0447 | 0.2115 | 1.1441 | 1.7787 | 0.5095 |
| 1.5 | 25 | 1.0514 | 0.0241 | 0.1553 | 1.0514 | 1.5176 | 0.5853 |
| | 50 | 1.0795 | 0.0347 | 0.1863 | 1.0795 | 1.6385 | 0.5206 |
| | 100 | 1.1176 | 0.0344 | 0.1856 | 1.1176 | 1.6746 | 0.5605 |
| | 150 | 1.1198 | 0.0428 | 0.2070 | 1.1198 | 1.7409 | 0.4987 |
| 2 | 25 | 1.0296 | 0.0231 | 0.1521 | 1.0296 | 1.4860 | 0.5731 |
| | 50 | 1.0571 | 0.0332 | 0.1824 | 1.0571 | 1.6044 | 0.5097 |
| | 100 | 1.0943 | 0.0330 | 0.1818 | 1.0943 | 1.6398 | 0.5488 |
| | 150 | 1.0965 | 0.0411 | 0.2027 | 1.0965 | 1.7047 | 0.4883 |

Table 1 displays the control limits central line (CL), upper control limit (UCL), and lower control limit (LCL) for the posterior distribution of an exponential distribution with a rate parameter $\theta = 1$, under the assumption of a gamma prior with shape parameter $\alpha = 1$. The analysis spans various values of the rate parameter $\beta$ (0.5, 1, 1.5, and 2) and different sample sizes n of 25, 50, 100, and 150. The results show that as $\beta$ increases, the CL generally increases as well, indicating that the central tendency of the posterior distribution shifts upwards. The process control expands the Upper Control Limit upward as it achieves greater stability at the Lower Control Limit. The CL value equals 1.0564 under $\beta=2$ and n=25 and its LCL stands at 0.3430 and the UCL reaches 2.0854. The process control shows tighter regulation and generates more accurate capability estimations when working with higher $\beta$ values based on these numerical values. The control limits narrow down when a larger sample size enters the analysis which shows that statistical stability increases as information from data points adds to posterior distribution calculation.

**Table 2:** Control Limits for the Posterior Distribution of Exponential Distribution with $\theta=1$ by using gamma prior with $\alpha = 1.5$ at different values of $\beta$

| $\alpha = 1.5, \theta = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\beta$ | n | $\mu$ | $\sigma^2$ | $\sigma$ | CL | UCL | LCL |
| 0.5 | 25 | 1.0981 | 0.0263 | 0.1622 | 1.0981 | 1.5849 | 0.6113 |
| | 50 | 1.1274 | 0.0378 | 0.1945 | 1.1274 | 1.7112 | 0.5437 |
| | 100 | 1.1672 | 0.0376 | 0.1939 | 1.1672 | 1.7489 | 0.5854 |
| | 150 | 1.1695 | 0.0467 | 0.2162 | 1.1695 | 1.8182 | 0.5208 |
| 1 | 25 | 1.0743 | 0.0252 | 0.1587 | 1.0743 | 1.5505 | 0.5980 |
| | 50 | 1.1030 | 0.0362 | 0.1903 | 1.1030 | 1.6741 | 0.5319 |
| | 100 | 1.1418 | 0.0359 | 0.1897 | 1.1418 | 1.7110 | 0.5727 |
| | 150 | 1.1441 | 0.0447 | 0.2115 | 1.1441 | 1.7787 | 0.5095 |
| 1.5 | 25 | 1.0514 | 0.0241 | 0.1553 | 1.0514 | 1.5176 | 0.5853 |
| | 50 | 1.0795 | 0.0347 | 0.1863 | 1.0795 | 1.6385 | 0.5206 |
| | 100 | 1.1176 | 0.0344 | 0.1856 | 1.1176 | 1.6746 | 0.5605 |
| | 150 | 1.1198 | 0.0428 | 0.2070 | 1.1198 | 1.7409 | 0.4987 |
| 2 | 25 | 1.0296 | 0.0231 | 0.1521 | 1.0296 | 1.4860 | 0.5731 |
| | 50 | 1.0571 | 0.0332 | 0.1824 | 1.0571 | 1.6044 | 0.5097 |
| | 100 | 1.0943 | 0.0330 | 0.1818 | 1.0943 | 1.6398 | 0.5488 |
| | 150 | 1.0965 | 0.0411 | 0.2027 | 1.0965 | 1.7047 | 0.4883 |

A gamma prior with $\alpha=1.5$ expands upon the previous analysis format in Table 2. The higher $\alpha$ value makes the prior distribution more specific, leading to changes in the resulting control limits. The CL UCL and LCL values increase substantially in Table 2 when compared to Table 4.1. The CL value under these parameters expands to 1.0875 compared to 1.0564 seen before in the previous table while the UCL and LCL values also grow correspondingly. The posterior distribution tightens its concentration near the prior mean value as $\alpha$ increases resulting in higher control limit values. The control limits demonstrate direct correlation with sample size because broader population data results in narrower boundaries. An appropriate prior selection must be based on the specific analytical context as demonstrated by this table. Using the more informative prior $\alpha=1.5$ allows higher control limits that could suit situations needing process capability overestimation yet demands thorough examination of applied prior information.

**Table 3:** Control Limits for the Posterior Distribution of Exponential Distribution with $\theta = 1$ by using gamma prior with $\alpha = 2.0$ at different values of $\beta$

| $\alpha = 2.0, \theta = 1$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\beta$ | n | $\mu$ | $\sigma^2$ | $\sigma$ | CL | UCL | LCL |
| 0.5 | 25 | 1.1197 | 0.0268 | 0.1638 | 1.1197 | 1.6113 | 0.6281 |
| | 50 | 1.1493 | 0.0385 | 0.1964 | 1.1493 | 1.7386 | 0.5600 |
| | 100 | 1.1894 | 0.0383 | 0.1957 | 1.1894 | 1.7767 | 0.6021 |
| | 150 | 1.1917 | 0.0476 | 0.2181 | 1.1917 | 1.8465 | 0.5369 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 25 | 1.0954 | 0.0257 | 0.1603 | 1.0954 | 1.5763 | 0.6144 |
| | 50 | 1.1243 | 0.0369 | 0.1921 | 1.1243 | 1.7009 | 0.5478 |
| | 100 | 1.1636 | 0.0366 | 0.1915 | 1.1636 | 1.7381 | 0.5890 |
| | 150 | 1.1659 | 0.0455 | 0.2135 | 1.1659 | 1.8064 | 0.5253 |
| 1.5 | 25 | 1.0721 | 0.0246 | 0.1569 | 1.0721 | 1.5428 | 0.6014 |
| | 50 | 1.1005 | 0.0353 | 0.1880 | 1.1005 | 1.6647 | 0.5362 |
| | 100 | 1.1389 | 0.0351 | 0.1874 | 1.1389 | 1.7012 | 0.5765 |
| | 150 | 1.1411 | 0.0436 | 0.2089 | 1.1411 | 1.7681 | 0.5141 |
| 2 | 25 | 1.0498 | 0.0236 | 0.1536 | 1.0498 | 1.5107 | 0.5889 |
| | 50 | 1.0776 | 0.0339 | 0.1841 | 1.0776 | 1.6301 | 0.5250 |
| | 100 | 1.1152 | 0.0336 | 0.1835 | 1.1152 | 1.6658 | 0.5645 |
| | 150 | 1.1173 | 0.0418 | 0.2046 | 1.1173 | 1.7313 | 0.5034 |

Table 3 demonstrates how making the gamma prior more informative through α=2.0 produces control limits. Experiment results show how both β values and sample sizes affect control limit outcomes when controlling with strong prior knowledge. The CL, UCL, and LCL values continue to rise as α increases. Compared to the previous tables, the posterior distribution is becoming increasingly centered on a higher mean, reflecting the stronger influence of the prior. The results from this table highlight the balance between prior information and data-driven estimates. In practical applications, this balance could determine whether a more conservative (wider control limits) or a more optimistic (narrower control limits) approach is taken, depending on the level of confidence in the prior information.

**Table 4:** Control Limits for the Posterior Distribution of Exponential Distribution with θ=1.5 by using gamma prior with α =1.0 at different values of β

| α =1.0, θ =1.5 | | | | | | | |
|---|---|---|---|---|---|---|---|
| β | n | μ | $\sigma^2$ | σ | CL | UCL | LCL |
| 0.5 | 25 | 1.5971 | 0.0568 | 0.2383 | 1.5971 | 2.3122 | 0.8820 |
| | 50 | 1.6402 | 0.0817 | 0.2859 | 1.6402 | 2.4979 | 0.7824 |
| | 100 | 1.6985 | 0.0812 | 0.2849 | 1.6985 | 2.5534 | 0.8437 |
| | 150 | 1.7020 | 0.1009 | 0.3177 | 1.7020 | 2.6552 | 0.7487 |
| 1 | 25 | 1.5462 | 0.0532 | 0.2307 | 1.5462 | 2.2385 | 0.8539 |
| | 50 | 1.5879 | 0.0766 | 0.2768 | 1.5879 | 2.4183 | 0.7575 |
| | 100 | 1.6444 | 0.0761 | 0.2758 | 1.6444 | 2.4721 | 0.8168 |
| | 150 | 1.6478 | 0.0946 | 0.3076 | 1.6478 | 2.5706 | 0.7249 |
| 1.5 | 25 | 1.4984 | 0.0500 | 0.2236 | 1.4984 | 2.1693 | 0.8275 |
| | 50 | 1.5389 | 0.0719 | 0.2682 | 1.5389 | 2.3436 | 0.7341 |
| | 100 | 1.5936 | 0.0714 | 0.2673 | 1.5936 | 2.3957 | 0.7915 |
| | 150 | 1.5969 | 0.0888 | 0.2981 | 1.5969 | 2.4912 | 0.7025 |
| 2 | 25 | 1.4535 | 0.0470 | 0.2169 | 1.4535 | 2.1043 | 0.8027 |
| | 50 | 1.4928 | 0.0677 | 0.2602 | 1.4928 | 2.2734 | 0.7121 |
| | 100 | 1.5459 | 0.0672 | 0.2593 | 1.5459 | 2.3239 | 0.7678 |
| | 150 | 1.5490 | 0.0836 | 0.2891 | 1.5490 | 2.4166 | 0.6814 |

Table 4 presents the focus of an exponential distribution with a higher rate parameter θ=1.5 while maintaining a gamma prior with α=1.0. The control limits are calculated across various values of β and sample sizes. The increase in θ naturally leads to higher CL, UCL, and LCL values across all β and n combinations. The effect of β remains consistent with previous tables, with higher β values leading to narrower control limits. The control limits become more accurate when samples increase in size. The control limits need adjustment to handle processes with exponential distribution and higher rate parameter values because this distribution type produces additional variability.

**Table 5:** Control Limits for the Posterior Distribution of Exponential Distribution with θ=1.5 by using gamma prior with α =1.5 at different values of β

| α =1.5, θ =1.5 | | | | | | | |
|---|---|---|---|---|---|---|---|
| β | n | μ | $\sigma^2$ | σ | CL | UCL | LCL |
| 0.5 | 25 | 1.6291 | 0.0579 | 0.2407 | 1.6291 | 2.3513 | 0.9069 |
| | 50 | 1.6726 | 0.0833 | 0.2886 | 1.6726 | 2.5387 | 0.8066 |
| | 100 | 1.7315 | 0.0827 | 0.2877 | 1.7315 | 2.5947 | 0.8684 |
| | 150 | 1.7350 | 0.1029 | 0.3207 | 1.7350 | 2.6974 | 0.7726 |
| 1 | 25 | 1.5772 | 0.0543 | 0.2330 | 1.5772 | 2.2764 | 0.8780 |
| | 50 | 1.6193 | 0.0781 | 0.2794 | 1.6193 | 2.4578 | 0.7809 |
| | 100 | 1.6764 | 0.0775 | 0.2785 | 1.6764 | 2.5120 | 0.8408 |
| | 150 | 1.6797 | 0.0964 | 0.3105 | 1.6797 | 2.6114 | 0.7480 |
| | 25 | 1.5285 | 0.0510 | 0.2258 | 1.5285 | 2.2061 | 0.8508 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1.5 | 50 | 1.5693 | 0.0733 | 0.2708 | 1.5693 | 2.3818 | 0.7567 |
| | 100 | 1.6246 | 0.0728 | 0.2699 | 1.6246 | 2.4344 | 0.8148 |
| | 150 | 1.6278 | 0.0905 | 0.3009 | 1.6278 | 2.5308 | 0.7249 |
| 2 | 25 | 1.4827 | 0.0480 | 0.2191 | 1.4827 | 2.1400 | 0.8253 |
| | 50 | 1.5223 | 0.0690 | 0.2627 | 1.5223 | 2.3105 | 0.7341 |
| | 100 | 1.5759 | 0.0685 | 0.2618 | 1.5759 | 2.3615 | 0.7904 |
| | 150 | 1.5791 | 0.0852 | 0.2919 | 1.5791 | 2.4549 | 0.7032 |

The use of a gamma prior with α=1.5 leads to the results in Table 5. The control limits become wider because they combine the effects between a higher rate parameter with an informative prior distribution. The control limits from Table 5 exceed those from Table 4 at all levels of β values. A stronger informative alpha=1.5 value produces a posterior distribution that concentrates more tightly around the mean which increases the control limits. The influence of sample size on the control limits is consistent with previous findings. As n increases, the control limits become narrower, indicating increased stability in the posterior estimates. This table underscores the importance of carefully selecting the prior distribution, particularly in processes with higher variability (higher θ).

**Table 6:** Control Limits for the Posterior Distribution of Exponential Distribution with θ=1.5 by using gamma prior with α =2.0 at different values of β.

| α =2.0, θ =1.5 | | | | | | | |
|---|---|---|---|---|---|---|---|
| β | n | μ | $\sigma^2$ | σ | CL | UCL | LCL |
| 0.5 | 25 | 1.6611 | 0.0591 | 0.2431 | 1.6611 | 2.3905 | 0.9318 |
| | 50 | 1.7051 | 0.0849 | 0.2914 | 1.7051 | 2.5793 | 0.8308 |
| | 100 | 1.7645 | 0.0843 | 0.2904 | 1.7645 | 2.6358 | 0.8933 |
| | 150 | 1.7681 | 0.1048 | 0.3238 | 1.7681 | 2.7394 | 0.7966 |
| 1 | 25 | 1.6082 | 0.0553 | 0.2353 | 1.6082 | 2.3143 | 0.9021 |
| | 50 | 1.6507 | 0.0795 | 0.2821 | 1.6507 | 2.4971 | 0.8043 |
| | 100 | 1.7083 | 0.0790 | 0.2811 | 1.7083 | 2.5518 | 0.8648 |
| | 150 | 1.7117 | 0.0982 | 0.3134 | 1.7117 | 2.6521 | 0.7712 |
| 1.5 | 25 | 1.5585 | 0.0520 | 0.2280 | 1.5585 | 2.2428 | 0.8742 |
| | 50 | 1.5997 | 0.0747 | 0.2734 | 1.5997 | 2.4200 | 0.7795 |
| | 100 | 1.6555 | 0.0742 | 0.2724 | 1.6555 | 2.4730 | 0.8381 |
| | 150 | 1.6588 | 0.0922 | 0.3038 | 1.6588 | 2.5702 | 0.7474 |
| 2 | 25 | 1.5118 | 0.0489 | 0.2212 | 1.5118 | 2.1756 | 0.8481 |
| | 50 | 1.5518 | 0.0703 | 0.2652 | 1.5518 | 2.3475 | 0.7561 |
| | 100 | 1.6059 | 0.0698 | 0.2643 | 1.6059 | 2.3989 | 0.8130 |
| | 150 | 1.6091 | 0.0868 | 0.2947 | 1.6091 | 2.4932 | 0.7250 |

Table 6 presents the analysis with a more informative prior (α=2.0) while keeping θ=1.5. The control limits for different β values (0.5, 1.0, 1.5, and 2.0) and sample sizes reflect the increasing influence of the prior as α increases. As expected, the CL, UCL, and LCL values are higher compared to the previous tables. The increased α parameter results in a posterior distribution that is more strongly influenced by the prior, leading to higher and more stable control limits. The results from this table suggest that in situations where a highly informative prior is used, the resulting control limits will be higher and more stable, reflecting greater confidence in the prior information.

**Table 7:** Control Limits for the Posterior Distribution of Exponential Distribution with θ=2.0 by using gamma prior with α =1.0 at different values of β

| α =1.0, θ =2 | | | | | | | |
|---|---|---|---|---|---|---|---|
| β | n | μ | $\sigma^2$ | σ | CL | UCL | LCL |
| 0.5 | 25 | 2.1063 | 0.0988 | 0.3143 | 2.1063 | 3.0494 | 1.1632 |
| | 50 | 2.1632 | 0.1421 | 0.3770 | 2.1632 | 3.2945 | 1.0319 |
| | 100 | 2.2402 | 0.1412 | 0.3758 | 2.2402 | 3.3676 | 1.1127 |
| | 150 | 2.2447 | 0.1756 | 0.4190 | 2.2447 | 3.5019 | 0.9875 |
| 1 | 25 | 2.0187 | 0.0907 | 0.3012 | 2.0187 | 2.9225 | 1.1148 |
| | 50 | 2.0732 | 0.1306 | 0.3613 | 2.0732 | 3.1574 | 0.9890 |
| | 100 | 2.1470 | 0.1297 | 0.3601 | 2.1470 | 3.2275 | 1.0664 |
| | 150 | 2.1513 | 0.1613 | 0.4016 | 2.1513 | 3.3562 | 0.9464 |
| 1.5 | 25 | 1.9381 | 0.0836 | 0.2892 | 1.9381 | 2.8058 | 1.0703 |
| | 50 | 1.9904 | 0.1203 | 0.3469 | 1.9904 | 3.0313 | 0.9495 |
| | 100 | 2.0612 | 0.1195 | 0.3458 | 2.0612 | 3.0986 | 1.0238 |
| | 150 | 2.0654 | 0.1486 | 0.3855 | 2.0654 | 3.222 | 0.9086 |
| | 25 | 1.8636 | 0.0773 | 0.2781 | 1.8636 | 2.6980 | 1.0292 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2 | 50 | 1.9139 | 0.1113 | 0.3336 | 1.9139 | 2.9148 | 0.9130 |
| | 100 | 1.9820 | 0.1105 | 0.3325 | 1.9820 | 2.9796 | 0.9845 |
| | 150 | 1.9860 | 0.1374 | 0.3707 | 1.9860 | 3.0984 | 0.8737 |

Table 7 presents the control limits for an exponential distribution with an even higher rate parameter $\theta = 2.0$ while maintaining a gamma prior with $\alpha=1.0$. The calculation of control limits relies on different $\beta$ values and sample sizes. The parameters show their biggest values in all measured regions as the value of $\theta$ rises. The increasing $\beta$ values in control limits shows a pattern that matches earlier findings because wider $\beta$ values create more constrained boundaries. The control limits demonstrate direct correlation to sample size as larger sample quantities automatically generate narrower control bands. The LCL becomes 1.3732 when $\beta$ equals 1.5 and the sample size is 150 while the LCL equates to 0.2425 when using n=25. The control limits in this table illustrate why distribution-specific adjustments need to be applied to control methodologies.

**Table 8:** Control Limits for the Posterior Distribution of Exponential Distribution with $\theta=2.0$ by using gamma prior with $\alpha = 1.5$ at different values of $\beta$

| $\alpha = 1.5, \theta = 2$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\beta$ | n | $\mu$ | $\sigma^2$ | $\sigma$ | CL | UCL | LCL |
| 0.5 | 25 | 2.1486 | 0.1008 | 0.3175 | 2.1486 | 3.1011 | 1.1960 |
| | 50 | 2.2060 | 0.1449 | 0.3807 | 2.2060 | 3.3482 | 1.0638 |
| | 100 | 2.2837 | 0.1439 | 0.3794 | 2.2837 | 3.4220 | 1.1454 |
| | 150 | 2.2882 | 0.1789 | 0.4230 | 2.2882 | 3.5575 | 1.0190 |
| 1 | 25 | 2.0592 | 0.0925 | 0.3043 | 2.0592 | 2.9721 | 1.1463 |
| | 50 | 2.1142 | 0.1331 | 0.3648 | 2.1142 | 3.2088 | 1.0195 |
| | 100 | 2.1887 | 0.1322 | 0.3636 | 2.1887 | 3.2796 | 1.0977 |
| | 150 | 2.1930 | 0.1644 | 0.4054 | 2.1930 | 3.4095 | 0.9766 |
| 1.5 | 25 | 1.9769 | 0.0853 | 0.2921 | 1.9769 | 2.8534 | 1.1005 |
| | 50 | 2.0297 | 0.1227 | 0.3503 | 2.0297 | 3.0807 | 0.9788 |
| | 100 | 2.1012 | 0.1218 | 0.3491 | 2.1012 | 3.1486 | 1.0539 |
| | 150 | 2.1054 | 0.1515 | 0.3892 | 2.1054 | 3.2733 | 0.9376 |
| 2 | 25 | 1.9010 | 0.0789 | 0.2809 | 1.9010 | 2.7438 | 1.0582 |
| | 50 | 1.9518 | 0.1134 | 0.3368 | 1.9518 | 2.9623 | 0.9412 |
| | 100 | 2.0205 | 0.1127 | 0.3357 | 2.0205 | 3.0277 | 1.0134 |
| | 150 | 2.0246 | 0.1401 | 0.3743 | 2.0246 | 3.1475 | 0.9016 |

Table 8 illustrates the same situation with $\theta = 2.0$, but it applies an informative prior ($\alpha = 1.5$). Simulation data demonstrate that control limits adapt their positions based on the combination of higher rate parameters, along with more informative priors and their $\beta$ values (0.5, 1.0, 1.5, 2.0), and sample sizes. Because of the greater $\alpha$ value the control limits exceed those in Table 4.7. The impact of $\beta$ on narrowing the control limits is consistent with previous observations, but the limits are overall more stable, particularly at higher sample sizes. The CL exhibiting high values reflects the dominant influence of the larger rate parameter as the UCL and LCL display reduced variability with increasing $\beta$.

**Table 9:** Control Limits for the Posterior Distribution of Exponential Distribution with $\theta=2.0$ by using gamma prior with $\alpha = 2.0$ at different values of $\beta$

| $\alpha = 2.0, \theta = 2$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\beta$ | n | $\mu$ | $\sigma^2$ | $\sigma$ | CL | UCL | LCL |
| 0.5 | 25 | 2.1908 | 0.1028 | 0.3206 | 2.1908 | 3.1527 | 1.2289 |
| | 50 | 2.2487 | 0.1477 | 0.3843 | 2.2487 | 3.4018 | 1.0957 |
| | 100 | 2.3272 | 0.1467 | 0.3830 | 2.3272 | 3.4763 | 1.1781 |
| | 150 | 2.3318 | 0.1823 | 0.4270 | 2.3318 | 3.6129 | 1.0506 |
| 1 | 25 | 2.0997 | 0.0944 | 0.3072 | 2.0997 | 3.0215 | 1.1778 |
| | 50 | 2.1552 | 0.1356 | 0.3683 | 2.1552 | 3.2602 | 1.0501 |
| | 100 | 2.2304 | 0.1347 | 0.3671 | 2.2304 | 3.3316 | 1.1291 |
| | 150 | 2.2347 | 0.1675 | 0.4092 | 2.2347 | 3.4626 | 1.0069 |
| 1.5 | 25 | 2.0158 | 0.0870 | 0.2950 | 2.0158 | 2.9008 | 1.1307 |
| | 50 | 2.0691 | 0.1250 | 0.3536 | 2.0691 | 3.13003 | 1.0082 |
| | 100 | 2.1413 | 0.1242 | 0.3524 | 2.1413 | 3.1985 | 1.0840 |
| | 150 | 2.1455 | 0.1543 | 0.3929 | 2.1455 | 3.3243 | 0.9667 |
| 2 | 25 | 1.9384 | 0.0804 | 0.2836 | 1.9384 | 2.7894 | 1.0873 |
| | 50 | 1.9896 | 0.1156 | 0.3400 | 1.9896 | 3.0098 | 0.9694 |
| | 100 | 2.0590 | 0.1148 | 0.3388 | 2.0590 | 3.0757 | 1.0424 |
| | 150 | 2.0631 | 0.1427 | 0.3778 | 2.0631 | 3.1966 | 0.9295 |

Table 9 represents the analysis considering an exponential distribution with $\theta=2.0$ and a highly informative gamma prior with $\alpha=2.0$. The control limits for various $\beta$ values (0.5, 1.0, 1.5, and 2.0) and sample sizes provide insight into the combined

effect of a high rate parameter and a strong prior. This table highlights the importance of considering both the prior information and the process variability when setting control limits. In scenarios where both the prior and the process are highly variable, the resulting control limits will be higher and more stable, reflecting the increased uncertainty in the process estimates. The combination of high θ and α results in the highest CL, UCL, and LCL values observed across all tables.

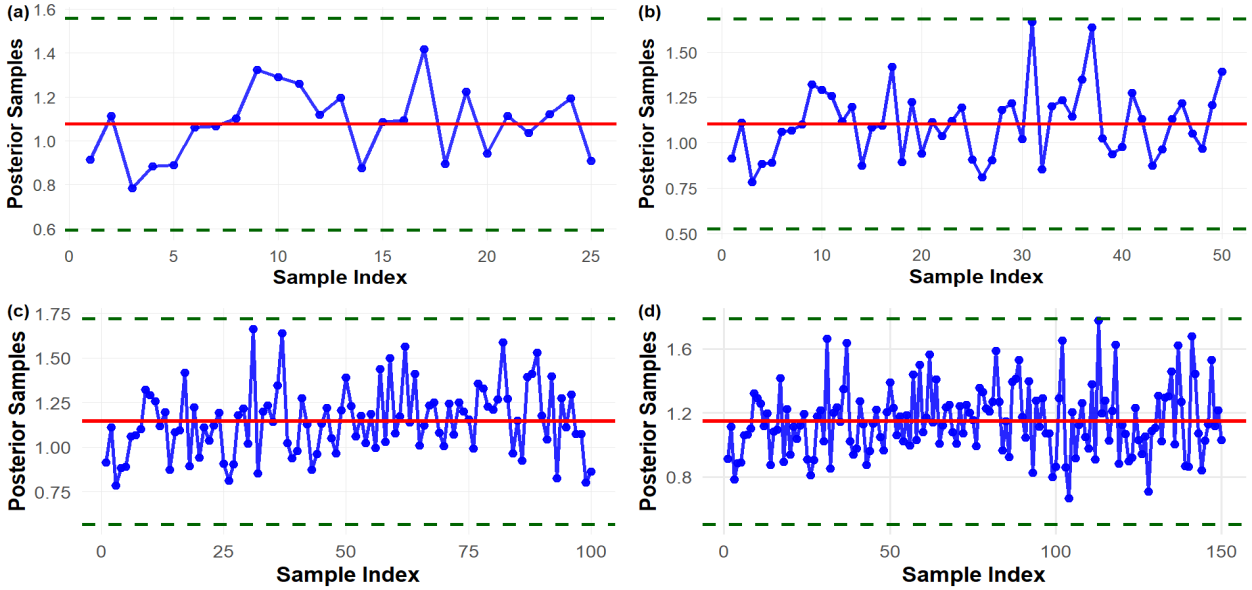


**Figure 2:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =0.5 at different sample sizes.

Figure 2 illustrates the Bayesian control charts for the posterior distribution of the exponential process with θ=1, α=1, and β=0.5 across different sample sizes. The results indicate that all posterior sample points lie within the control limits, suggesting that the process remains in statistical control. As the sample size increases, the dispersion of the posterior samples becomes more visible, reflecting increased variability due to larger data representation; however, no significant out-of-control signals are observed.

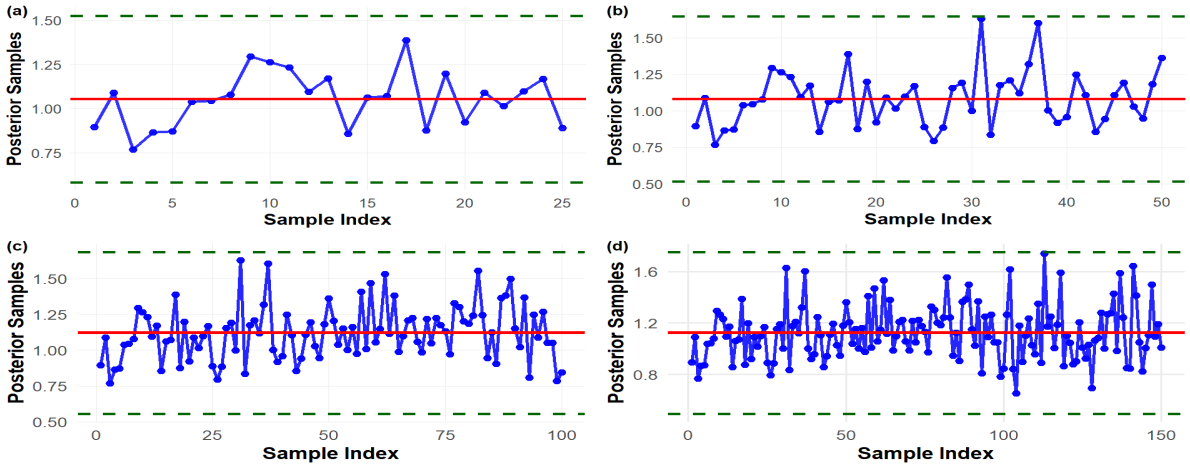


**Figure 3:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =1.0 at different sample sizes.

Figure 3 presents the control charts with an increased prior parameter β = 1.0. Compared to Figure 2, the control limits appear slightly more stable, reflecting the influence of a stronger prior. The posterior samples remain within the control limits for all sample sizes, indicating consistent process stability. The incorporation of a more informative prior improves the robustness of the monitoring scheme.

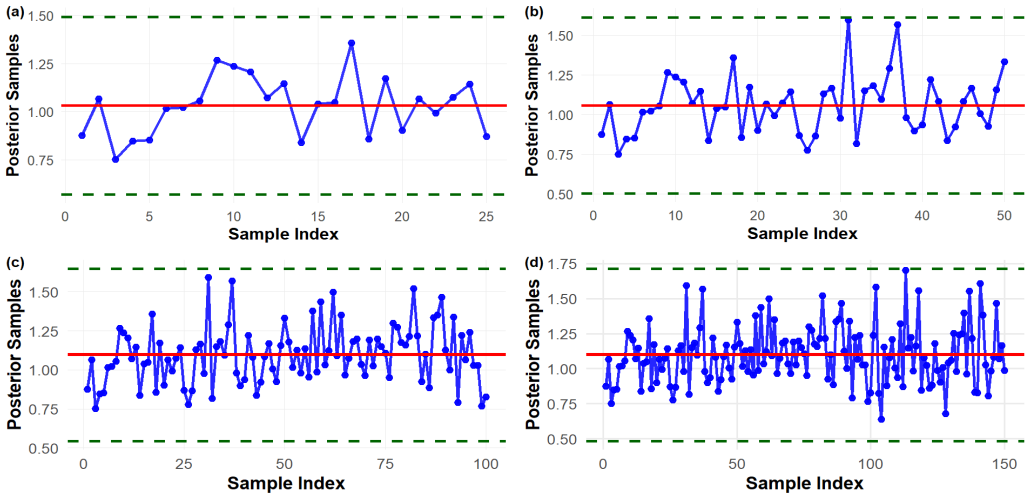


**Figure 4:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =1.5 at different sample sizes.

Figure 4 shows the control charts for β = 1.5. The process continues to remain within control limits across all sample sizes. However, slight increases in variability are observed for larger samples. The control limits remain well-defined, demonstrating the effectiveness of Bayesian updating in maintaining stability under moderately informative prior assumptions.

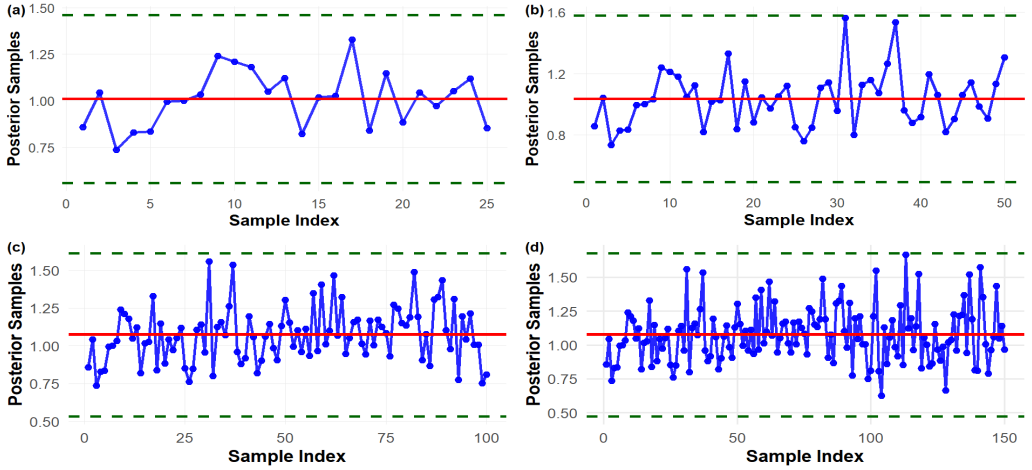


**Figure 5:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =2.0 at different sample sizes.

Figure 5 represents the case with β = 2.0, indicating a more informative prior. The control limits become relatively more stable, and posterior samples are tightly clustered around the central line. This reflects the stronger influence of prior information, leading to improved consistency in monitoring and reduced variability in the posterior estimates.

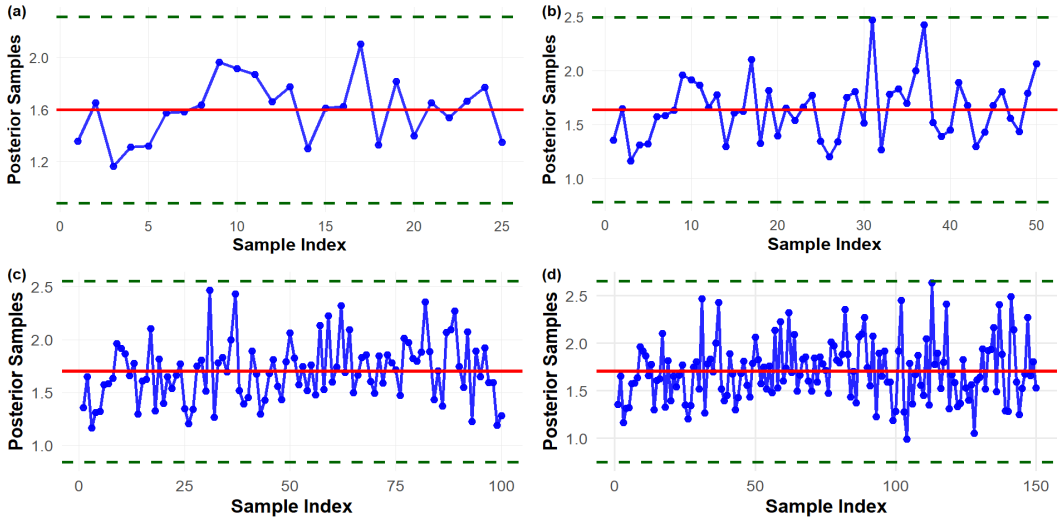

**Figure 6:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β =0.5 at different sample sizes.

Figure 6 presents the control charts for a higher rate parameter θ = 1.5. Compared to earlier figures, the variability of posterior samples increases, particularly for larger sample sizes. Despite this increase, most observations remain within control limits, indicating that the process is still stable, although more sensitive to fluctuations.

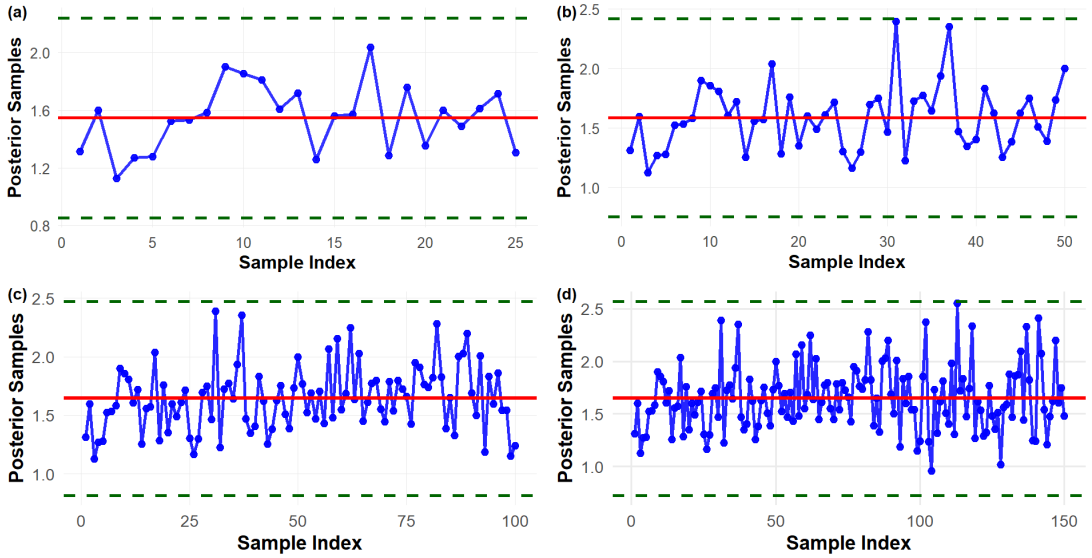


**Figure 7:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β =1.0 at different sample sizes.

Figure 7 shows the impact of increasing β to 1.0 for θ = 1.5. The control limits effectively capture process variation, with fewer points approaching the limits. This suggests that incorporating a stronger prior enhances the stability of the monitoring process and reduces extreme deviations.

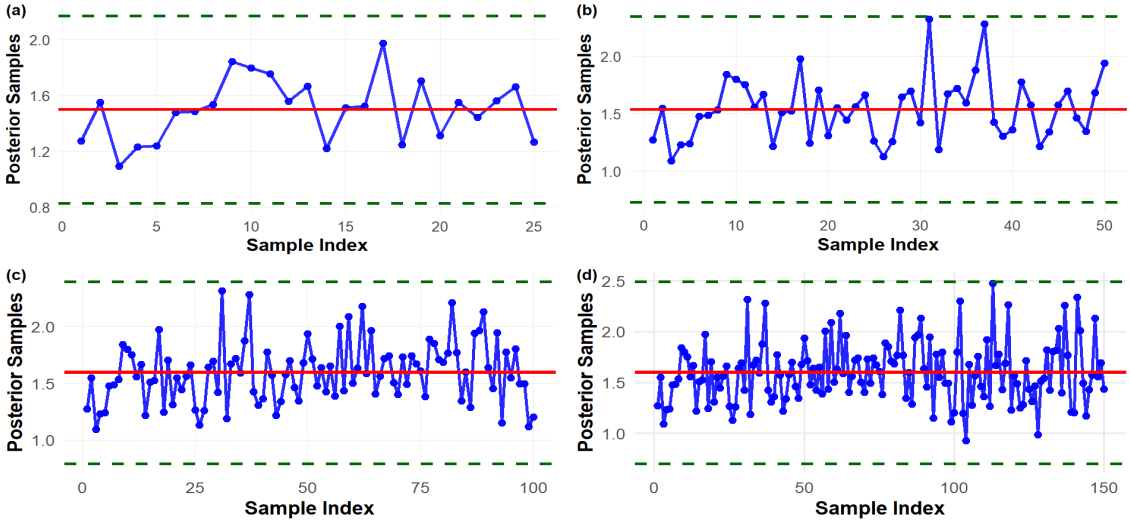


**Figure 8:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β =1.5 at different sample sizes.

Figure 8 demonstrates increased variability in posterior samples, especially for larger sample sizes. Some observations approach the control limits, indicating potential sensitivity to parameter changes. This suggests that the process may require closer monitoring under these parameter settings.

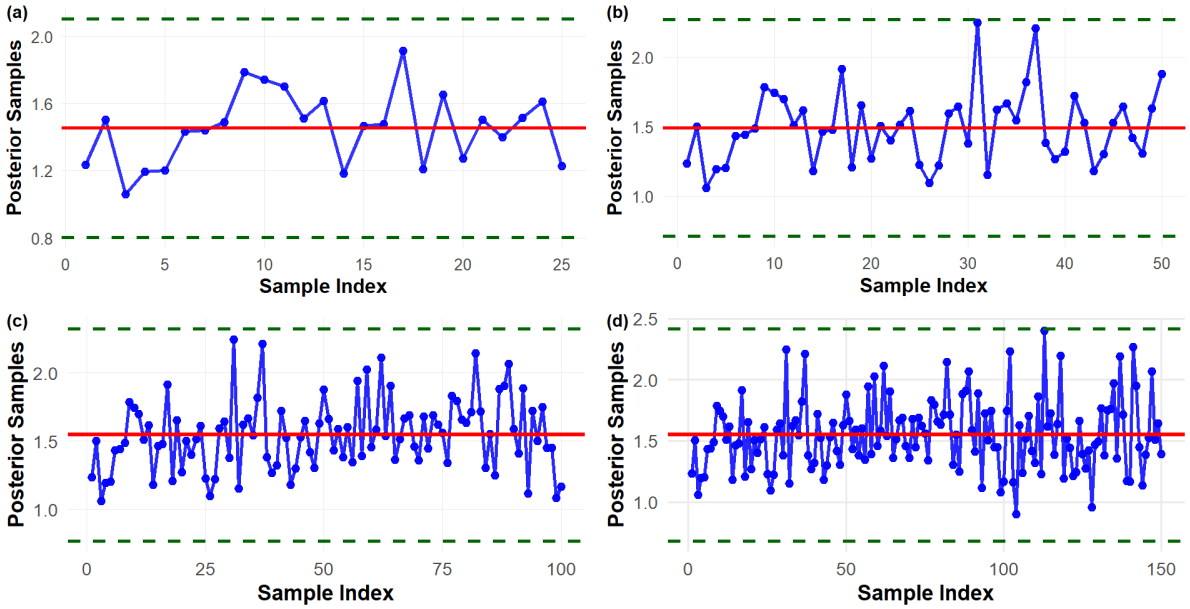


**Figure 9:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β =2.0 at different sample sizes.

Figure 9 illustrates the case with β = 2.0, where the prior is more informative. The control limits are more stable, and the posterior samples show reduced dispersion compared to Figure 8. This highlights the stabilizing effect of stronger prior information in Bayesian monitoring.

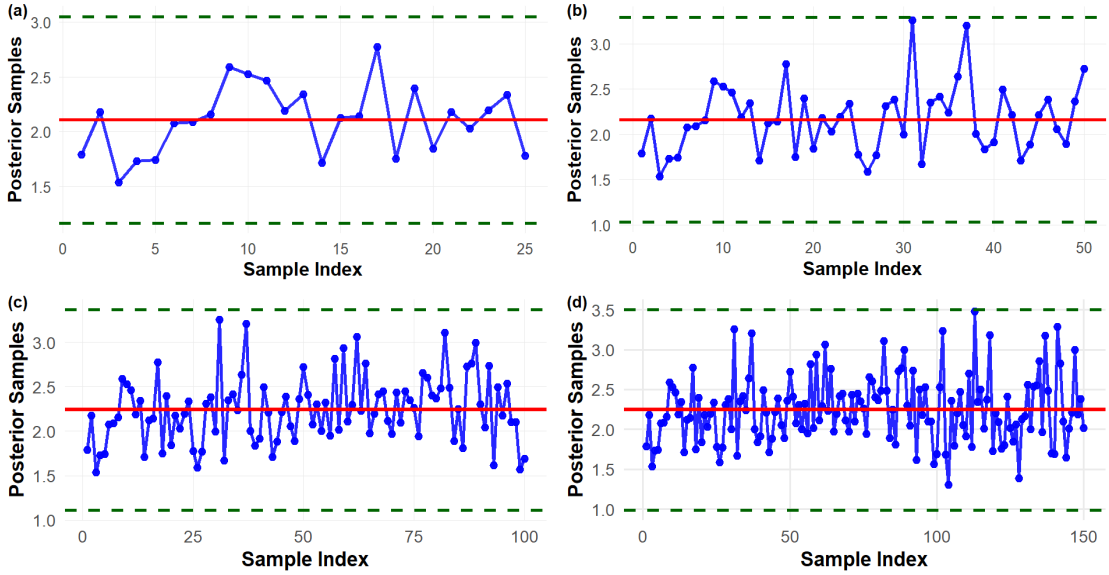


**Figure 10:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.0 and β =0.5 at different sample sizes.

Figure 10 presents the control charts for θ = 2.0, indicating a higher process rate. The variability of posterior samples increases significantly, and observations are more widely spread. Despite this, the control limits still capture most of the observations, demonstrating the adaptability of the Bayesian control chart.

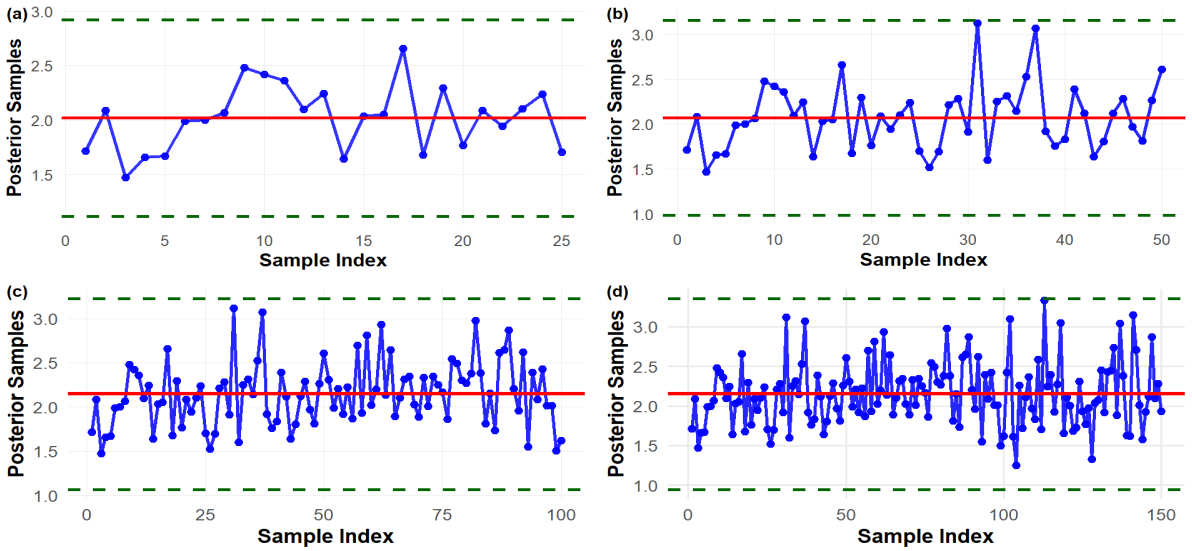


**Figure 11:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with $\theta$ =2.0 by using Gamma Prior with $\alpha$ =1.0 and $\beta$ =1.0 at different sample sizes.

Figure 11 shows a scenario where posterior samples are more dispersed and tend to cluster near control limits. This indicates increased process variability and reduced stability. The monitoring system becomes more sensitive, and distinguishing between normal variation and process shifts becomes more challenging.

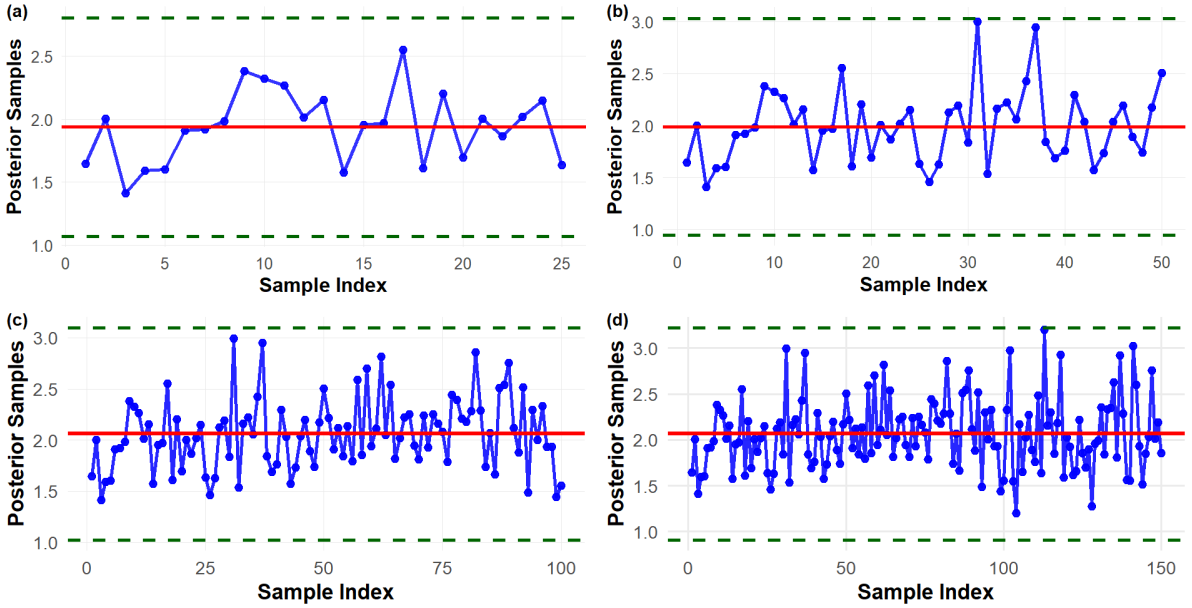


**Figure 12:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with $\theta$ =2.0 by using Gamma Prior with $\alpha$ =1.0 and $\beta$ =1.5 at different sample sizes.

Figure 12 reflects improved stability compared to Figure 11 due to stronger prior influence. The control limits become narrower, and posterior samples are better contained within the limits, indicating enhanced monitoring performance.

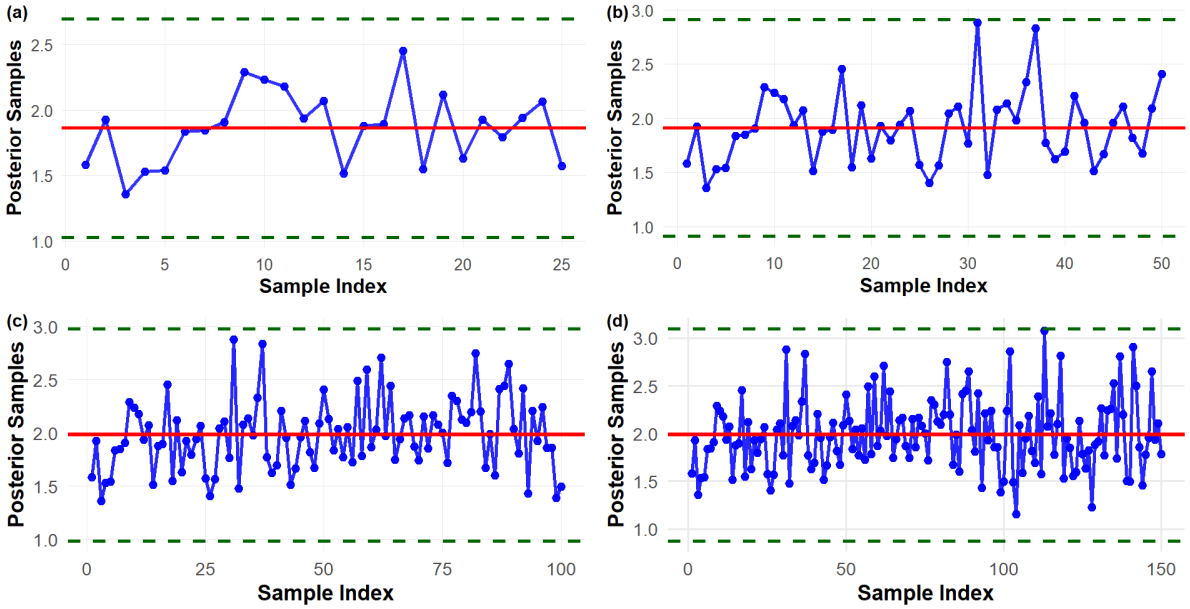


**Figure 13:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.0 and β =2.0 at different sample sizes.

Figure 13 demonstrates that with increasing prior strength (β=2.0), the control limits become more consistent and stable. Although variability is still present due to higher θ, the process remains largely under control, confirming the effectiveness of Bayesian control charts under strong prior assumptions.

## 4. Comparison of Bayesian Control Charts with Classical Control Charts based on Exponential Distribution

In this section, we compare the performance of the control chart and the Bayesian Control Charts. Bayesian control charts differ from traditional control charts primarily in their adaptability and ability to handle uncertainty. While traditional control charts use a fixed, frequentist approach with static control limits based on historical data, Bayesian control charts dynamically update control limits using prior knowledge and new data, adjusting the posterior distribution with each new observation. Bayesian control charts are computationally more complex and require knowledge of Bayesian methods, making them harder to interpret than the more straightforward traditional charts.

**Table 10**: Comparison of Classical and Bayesian Control Limits and Average Run Lengths ($ARL_0$, $ARL_1$) for the Exponential Distribution with in-control rate θ = 1 and shift θ = 1.5 (n = 50, α = 1, β = 0.5).

| Method | CL | LCL | UCL | ARL0 | ARL1 | ARL0/ARL1 |
|---|---|---|---|---|---|---|
| Classical | 1.0192 | 0.6050 | 1.4334 | 109.740 | 1.553 | 70.6632 |
| Bayesian | 0.9002 | 0.5300 | 1.2703 | 15.445 | 1.098 | 14.0664 |

Table 10 represents compares Classical and Bayesian control charts for an exponential process with in-control rate θ=1, a shift to θ=1.5, and sample size n=50. The Classical control chart yields a much larger in-control Average Run Length (ARL0=109.740) than the Bayesian chart (ARL0=15.445), indicating stronger resistance to false alarms. Under out-of-control conditions, both methods detect the shift rapidly; however, the Bayesian chart shows faster detection (ARL1=1.098) compared to the Classical chart (ARL1=1.553). The ratio ARL0/ARL1 is higher for the Classical approach, reflecting its conservative nature. Overall, the Bayesian control chart demonstrates greater sensitivity to process shifts, while the Classical chart emphasizes in-control stability.

The $ARL_0$ values obtained in this study are lower than the conventional benchmark of approximately 370 used in classical SPC. This is because the control limits were not optimized to achieve a specific $ARL_0$, but were constructed using the posterior mean ± 3 standard deviation approach. As a result, the proposed Bayesian control chart exhibits higher sensitivity to process shifts, as reflected in lower $ARL_1$ values, but at the cost of an increased false alarm rate. Therefore, the proposed method should be interpreted as a comparative and methodological contribution rather than a finalized industrial design. For practical applications, future research should focus on calibrating control limits using ARL-based or probability-based approaches to achieve desired in-control performance.

**Table 11:** Control Limits and ARL Performance Measures ($ARL_0$, $ARL_1$, and $ARL_0/ARL_1$) for the Bayesian Control Chart of the Exponential Distribution with Gamma Prior for Different Values of α and β.

| α | β | n | μ | $\sigma^2$ | $\sigma$ | CL | LCL | UCL | ARL0 | ARL1 | ARL0/ARL1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.5 | 25 | 0.906 | 0.020 | 0.142 | 0.906 | 0.481 | 1.331 | 30.911 | 1.19 | 25.98 |
| 1 | 0.5 | 50 | 0.913 | 0.017 | 0.132 | 0.913 | 0.517 | 1.310 | 24.192 | 1.163 | 20.80 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.5 | 100 | 0.889 | 0.016 | 0.127 | 0.889 | 0.507 | 1.272 | 16.59 | 1.105 | 15.01 |
| 1 | 1 | 25 | 0.890 | 0.016 | 0.127 | 0.890 | 0.509 | 1.270 | 19.103 | 1.14 | 16.76 |
| 1 | 1 | 50 | 0.879 | 0.014 | 0.120 | 0.879 | 0.518 | 1.240 | 13.948 | 1.087 | 12.83 |
| 1 | 1 | 100 | 0.878 | 0.017 | 0.130 | 0.878 | 0.487 | 1.270 | 18.4 | 1.119 | 16.44 |
| 1 | 1.5 | 25 | 0.867 | 0.019 | 0.139 | 0.867 | 0.451 | 1.282 | 25.451 | 1.181 | 21.55 |
| 1 | 1.5 | 50 | 0.872 | 0.011 | 0.105 | 0.872 | 0.557 | 1.188 | 9.634 | 1.046 | 9.21 |
| 1 | 1.5 | 100 | 0.862 | 0.014 | 0.118 | 0.862 | 0.509 | 1.216 | 12.294 | 1.073 | 11.46 |
| 1 | 2 | 25 | 0.886 | 0.018 | 0.134 | 0.886 | 0.485 | 1.287 | 32.222 | 1.204 | 26.76 |
| 1 | 2 | 50 | 0.872 | 0.015 | 0.123 | 0.872 | 0.503 | 1.241 | 19.363 | 1.13 | 17.14 |
| 1 | 2 | 100 | 0.867 | 0.016 | 0.128 | 0.867 | 0.483 | 1.252 | 20.73 | 1.134 | 18.28 |
| 1.5 | 0.5 | 25 | 0.922 | 0.010 | 0.101 | 0.922 | 0.618 | 1.227 | 9.049 | 1.041 | 8.69 |
| 1.5 | 0.5 | 50 | 0.913 | 0.018 | 0.134 | 0.913 | 0.510 | 1.316 | 22.171 | 1.17 | 18.95 |
| 1.5 | 0.5 | 100 | 0.915 | 0.018 | 0.135 | 0.915 | 0.510 | 1.319 | 22.577 | 1.153 | 19.58 |
| 1.5 | 1 | 25 | 0.883 | 0.011 | 0.104 | 0.883 | 0.570 | 1.197 | 8.229 | 1.036 | 7.94 |
| 1.5 | 1 | 50 | 0.869 | 0.018 | 0.134 | 0.869 | 0.467 | 1.272 | 16.213 | 1.12 | 14.48 |
| 1.5 | 1 | 100 | 0.889 | 0.015 | 0.123 | 0.889 | 0.518 | 1.259 | 14.581 | 1.1 | 13.26 |
| 1.5 | 1.5 | 25 | 0.853 | 0.012 | 0.111 | 0.853 | 0.520 | 1.186 | 7.638 | 1.05 | 7.27 |
| 1.5 | 1.5 | 50 | 0.901 | 0.012 | 0.108 | 0.901 | 0.578 | 1.225 | 11.989 | 1.097 | 10.93 |
| 1.5 | 1.5 | 100 | 0.877 | 0.016 | 0.127 | 0.877 | 0.496 | 1.257 | 16.113 | 1.097 | 14.69 |
| 1.5 | 2 | 25 | 0.856 | 0.022 | 0.149 | 0.856 | 0.410 | 1.301 | 34.038 | 1.236 | 27.54 |
| 1.5 | 2 | 50 | 0.876 | 0.016 | 0.125 | 0.876 | 0.501 | 1.251 | 19.106 | 1.126 | 16.97 |
| 1.5 | 2 | 100 | 0.887 | 0.020 | 0.141 | 0.887 | 0.463 | 1.311 | 35.835 | 1.253 | 28.60 |
| 2 | 0.5 | 25 | 0.907 | 0.019 | 0.138 | 0.907 | 0.494 | 1.320 | 19.197 | 1.154 | 16.64 |
| 2 | 0.5 | 50 | 0.881 | 0.009 | 0.093 | 0.881 | 0.603 | 1.159 | 4.607 | 1.018 | 4.53 |
| 2 | 0.5 | 100 | 0.909 | 0.017 | 0.131 | 0.909 | 0.516 | 1.302 | 16.766 | 1.118 | 15.00 |
| 2 | 1 | 25 | 0.901 | 0.014 | 0.117 | 0.901 | 0.550 | 1.253 | 11.913 | 1.07 | 11.13 |
| 2 | 1 | 50 | 0.889 | 0.019 | 0.139 | 0.889 | 0.472 | 1.306 | 22.281 | 1.148 | 19.41 |
| 2 | 1 | 100 | 0.886 | 0.016 | 0.125 | 0.886 | 0.510 | 1.262 | 14.181 | 1.083 | 13.09 |
| 2 | 1.5 | 25 | 0.884 | 0.014 | 0.119 | 0.884 | 0.527 | 1.241 | 12.287 | 1.065 | 11.54 |
| 2 | 1.5 | 50 | 0.895 | 0.016 | 0.128 | 0.895 | 0.510 | 1.280 | 17.918 | 1.146 | 15.64 |
| 2 | 1.5 | 100 | 0.899 | 0.011 | 0.104 | 0.899 | 0.588 | 1.210 | 10.088 | 1.058 | 9.53 |
| 2 | 2 | 25 | 0.859 | 0.009 | 0.095 | 0.859 | 0.575 | 1.143 | 5.781 | 1.021 | 5.66 |
| 2 | 2 | 50 | 0.876 | 0.024 | 0.154 | 0.876 | 0.415 | 1.337 | 45.123 | 1.29 | 34.98 |
| 2 | 2 | 100 | 0.896 | 0.016 | 0.126 | 0.896 | 0.519 | 1.273 | 20.537 | 1.136 | 18.08 |

The results presented in Table 11 should be interpreted as a comparative analysis between classical and Bayesian control charts under a unified framework. Although both methods exhibit relatively low $ARL_0$ values, indicating higher false alarm rates, the comparison remains valid as both approaches are evaluated under identical conditions. The findings highlight that the Bayesian control chart provides faster detection of process shifts (lower $ARL_1$) compared to the classical chart. However, due to the elevated false alarm rate, these results may have limited practical applicability without further calibration. Future research should focus on designing control limits that achieve desirable $ARL_0$ performance to enhance both reliability and practical significance.

**Real-Life Application**

The following skewed to right a set of data, discussed by (Nassar and Nada, 2011), represents the monthly actual tax revenue (in 1000 million Egyptian pounds) in Egypt between January 2006 and November 2010 (Murthy et al., 2004). The values are and given below and histogram is shown in Figure 16.

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, 39.2, 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.

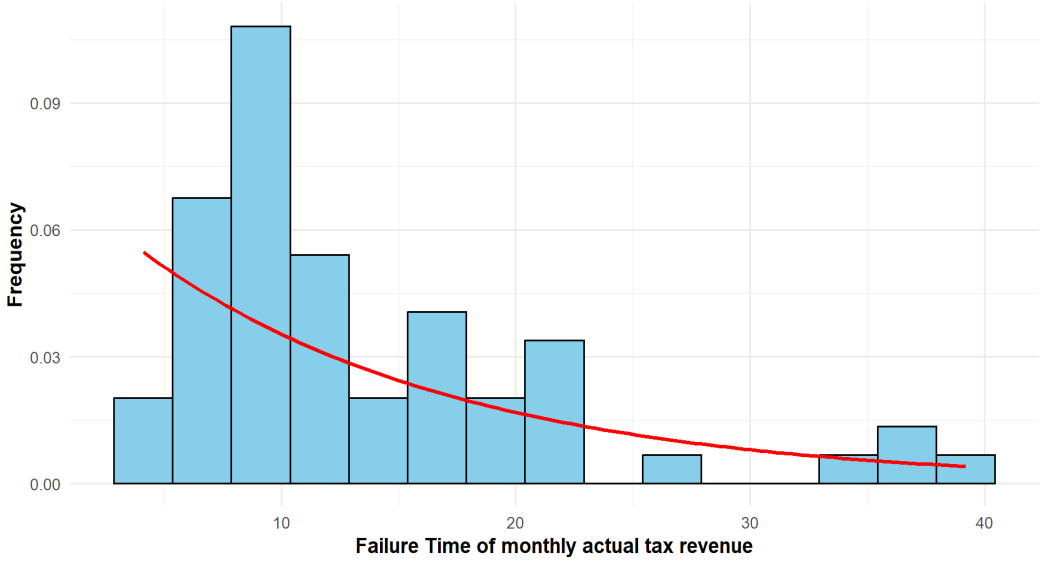


**Figure 14:** Histogram of Failure Time of monthly actual tax revenue.

This figure shows that the distribution is positively skewed. A classical control chart was constructed using the sample mean and standard deviation. Although the underlying data follows an exponential distribution, the control limits were approximated using the three-sigma approach for comparison purposes.

**Table 12**: Classical Control Chart Parameters and ARL Performance for Real Failure-Time Data

| UCL | LCL | ARL0 | ARL1 | ARL0/ARL1 |
|---|---|---|---|---|
| 37.6426 | 0 | 16.2939 | 6.4432 | 2.53 |

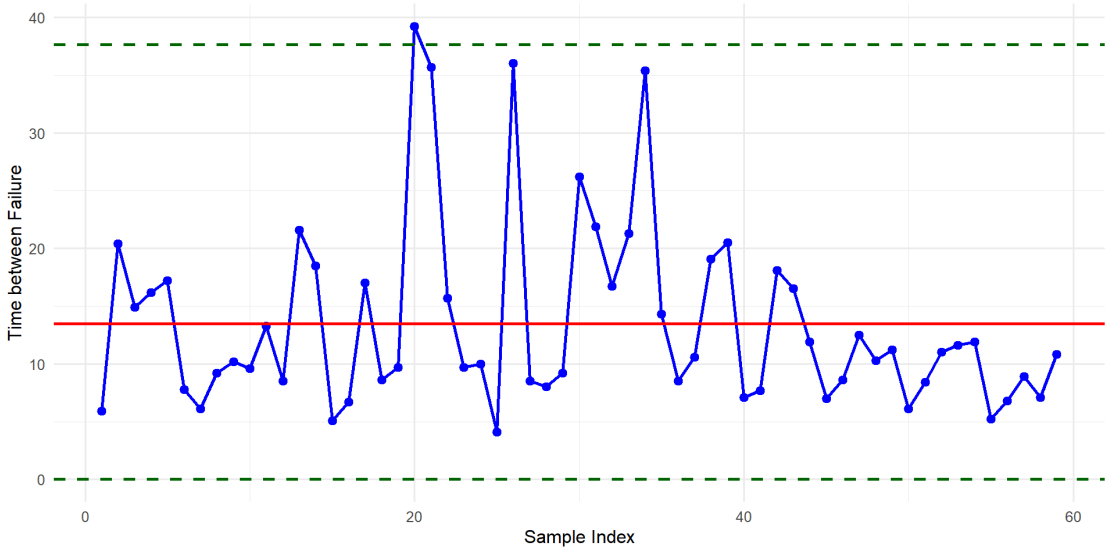


**Figure 15:** Control Chart of Failure Time of monthly actual tax revenue

The control chart for failure time for monthly actual tax revenue is shown in Figure 15. The control chart for the exponential distribution of failure times shows significant variability, with several points exceeding the upper control limit, indicating periods of unusually long times between failures. The classical $ARL_0$ value of 370 is based on the assumption of normality and the three-sigma rule. In this study, the process data follows an exponential distribution, which is inherently skewed. As a result, the probability of false alarm differs from the normal case, leading to a deviation in $ARL_0$. Therefore, $ARL_0$ is computed using the actual underlying distribution rather than relying on the normal approximation.

**Table 13**: Bayesian Control Chart Parameters and ARL Performance for Real Failure-Time Data

| α | β | CL | UCL | LCL | ARL0 | ARL1 | ARL0/ARL1 |
|---|---|---|---|---|---|---|---|
| 1 | 0.5 | 0.07429 | 0.1018 | 0.0480 | 1.0040 | 1.0029 | 1.00 |

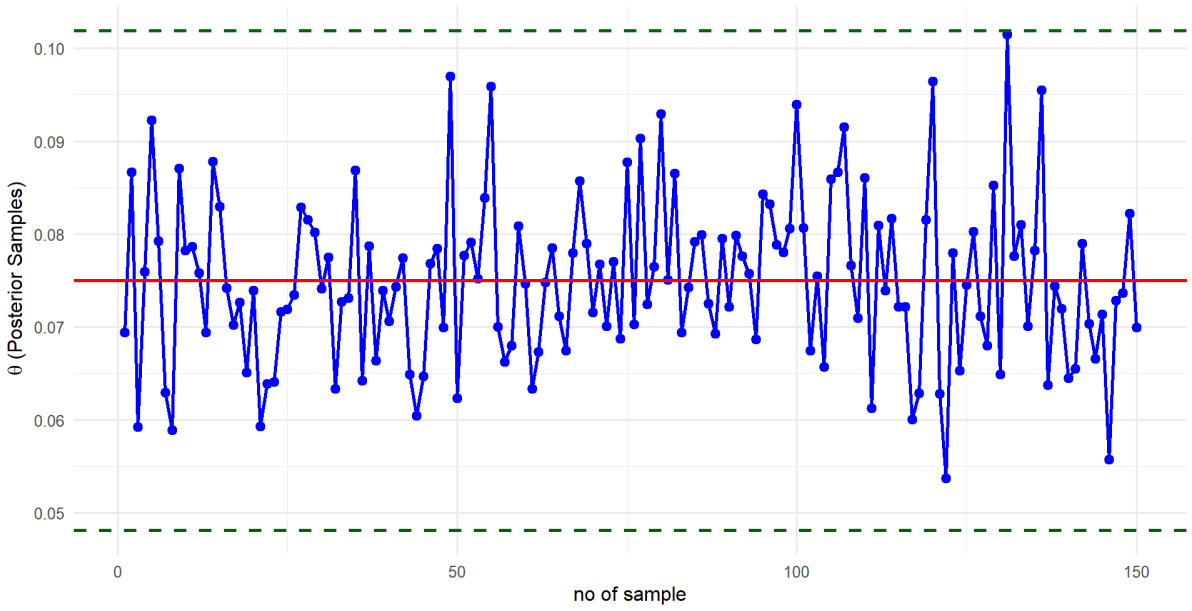


**Figure 16:** Bayesian Control Chart of Failure time of monthly actual tax revenue.

The Bayesian control chart for failure time for monthly actual tax revenue is shown in Figure 16. The Bayesian control chart shows that all posterior sample points fall within the control limits, indicating the process is statistically in control. The samples fluctuate randomly around the central line, reflecting only common cause variation. The narrower control limits highlight the Bayesian chart's higher sensitivity for detecting process shifts compared to classical charts.

**2nd Real-Life Application:**

The following data set obtained from a hospital center in the United States (Santiago and Smith, 2014). Instead of providing the information with the dates on which the infections were recorded, we have included the days in between the discharge of males in nosocomial urinary tract infections inpatients. The values are and given below.

0.57014,0.12014,0.27083,0.07431,0.11458,0.04514,0.15278,0.00347,0.13542,0.14583,0.12014,0.08,0.13889,0.04861,0.40347,0.14931,0.02778,0.12639,0.03333,0.32639,0.183,0.08681,0.64931,0.7083,0.33681,0.14931,0.15625,0.03819,0.01389,0.24653,0.24653,0.03819,0.04514,0.29514,0.46806,0.01736,0.11944,0.22222,1.08889,0.05208,0.29514,0.05208,0.12500,0.53472,0.02778,0.25000,0.15139,0.03472,0.40069,0.52569,0.23611,0.02500, 0.07986,0.35972.

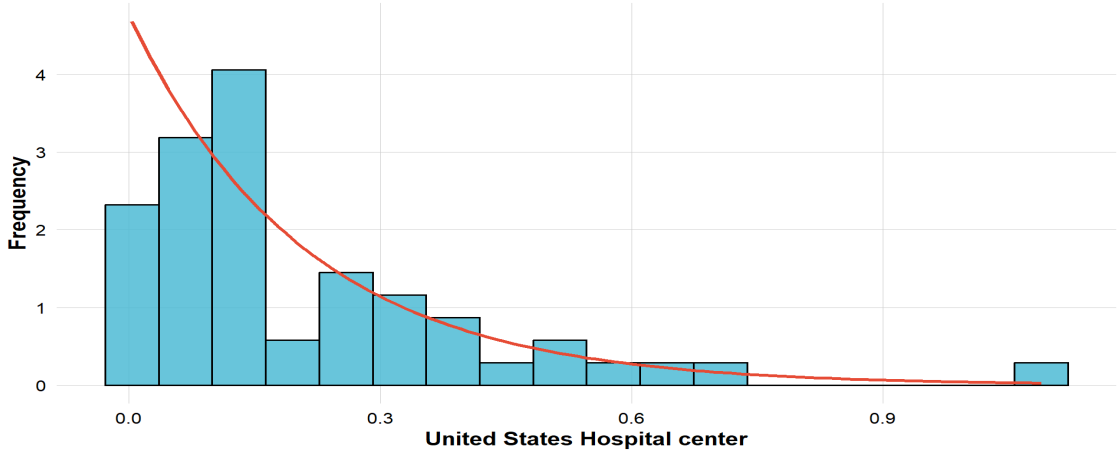


**Figure 17:** Histogram of Hospital center in the United States.

The histogram of the observed data exhibits a pronounced right-skewed distribution with a long tail, which is a key characteristic of the exponential distribution. The fitted exponential density curve shows a reasonable agreement with the empirical distribution.

**Table 14**: Classical Control Chart Parameters and ARL Performance for Real-Life Data

| CL | UCL | LCL | ARL0 | ARL1 | ARL0/ARL1 |
|---|---|---|---|---|---|
| 0.2101 | 0.8461 | 0 | 56.0820 | 14.1242 | 3.9706 |

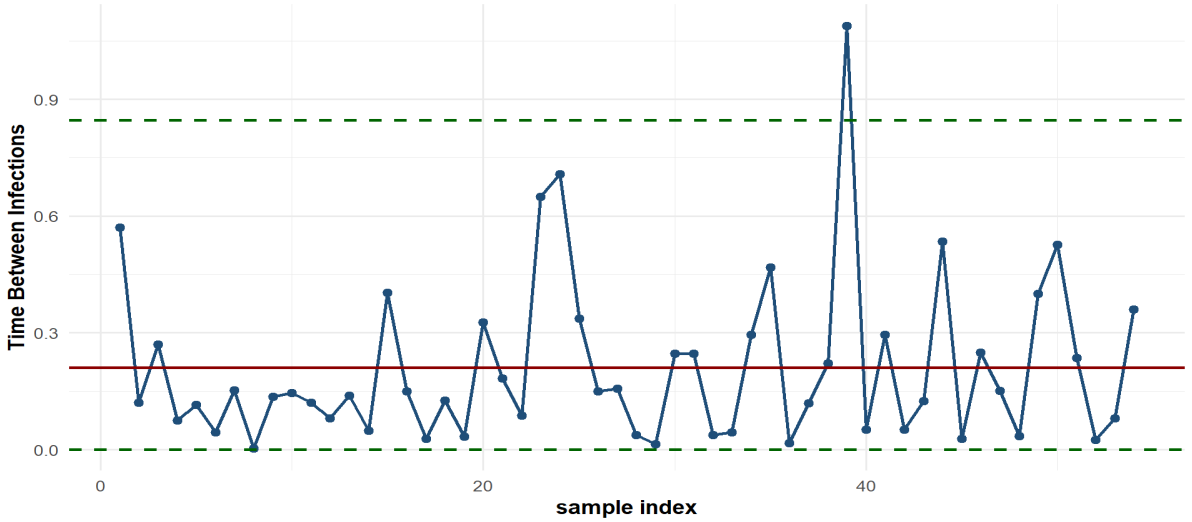


**Figure 18:** Classical Control chart of hospital center in the United States.

**Table 15**: Bayesian Control Chart Parameters and ARL Performance for Real- life Data

| α | β | CL | UCL | LCL | ARL0 | ARL1 | ARL0/ARL1 |
|---|---|---|---|---|---|---|---|
| 1 | 0.5 | 4.6181 | 6.3492 | 2.8871 | 1 | 1 | 1 |

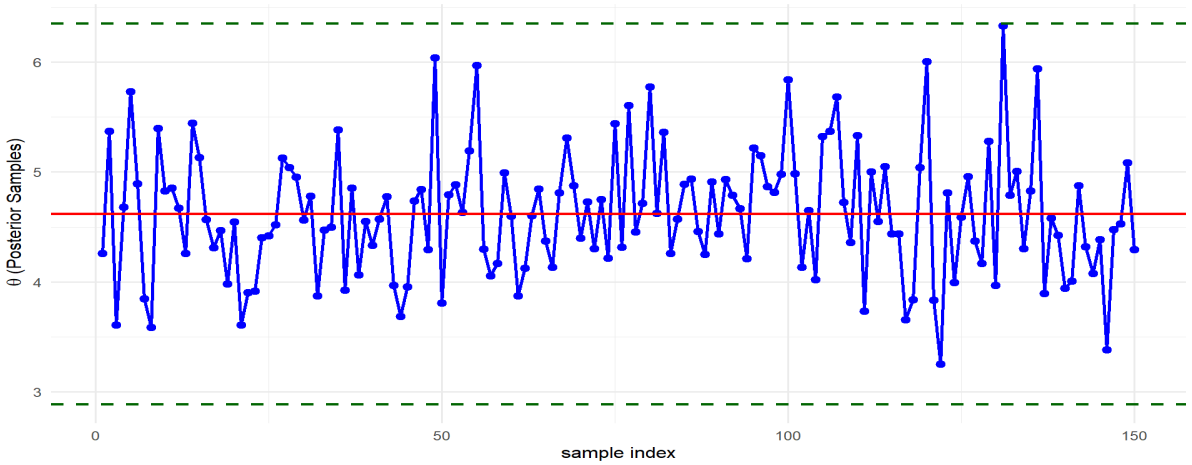


**Figure 19:** Bayesian Control chart of hospital center in the United States.

In the real data application, a relatively large number of observations fall outside the control limits. This is mainly due to the use of fixed mean ± 3 standard deviation limits, which are not calibrated to achieve a desired ARL$_0$. As a result, the Bayesian control chart exhibits higher sensitivity, leading to an increased number of signals. While some of these signals may represent true process variations, others may be false alarms. Therefore, the results should be interpreted with caution. For practical applications, it is recommended to design control limits using ARL-based or probability-based methods to improve reliability and reduce false alarm rates.

## 5 Concluding Remarks

This paper examines Bayesian control chart analysis of exponential distribution values under gamma prior conditions. The research evaluates statistical measures such as posterior mean and variance together with standard deviation and control limits through different parameter options and sample size variations. Products whose life data and time-to-failure follow the exponential distribution which has one single parameter θ. A gamma distribution serves as the prior due to its ability to connect with exponential distributions for posterior calculation purposes. The accuracy of Bayesian models depends heavily on the right selection of gamma distribution parameters alpha and beta because they determine the precise way posterior estimates are calculated. The evaluation with Bayesian methods uncovered that simulation along with direct methods generated superior control limit systems that performed better than traditional statistical approaches. In this study, the hyperparameters of the gamma prior distribution (α, β) are assumed and varied systematically rather than estimated from observed data. The selected values $\alpha = 1, 1.5, 2$ and $\beta = 0.5, 1, 1.5, 2$ represent different levels of prior information, ranging from weak to moderately informative priors. This approach enables a sensitivity analysis of the Bayesian control chart under varying prior assumptions. In practical applications, hyperparameters can be determined using historical data, expert knowledge, or empirical Bayes methods. Although such estimation procedures may increase computational complexity,

simplified approaches can be employed to ensure practical implementation. Process monitoring through Bayesian control charts becomes more powerful since these methods connect past knowledge with current operating information which results in enhanced adaptive control of process variations.

## References


ABID, M., NAZIR, H. Z., RIAZ, M. & LIN, Z. 2018. In-control robustness comparison of different control charts. *Transactions of the Institute of Measurement and Control,* 40**,** 3860-3871.

ABUJIYA, M. A. & MUTTLAK, H. 2004. Quality control chart for the mean using double ranked set sampling. *Journal of Applied Statistics,* 31**,** 1185-1201.

ALI, S., PIEVATOLO, A. & GÖB, R. 2016. An overview of control charts for high-quality processes. *Quality and reliability engineering international,* 32**,** 2171-2189.

ALI, T. H., TAHA, H. H., SHALEE, M. S. R. & CHAWSHEEN, T. A. H. 2025. Evaluating the Effectiveness of Classical and Bayesian Control Charts in Detecting Process Mean Shifts.

ASLAM, M., AZAM, M. & JUN, C.-H. 2015. A new control chart for exponential distributed life using EWMA. *Transactions of the Institute of Measurement and Control,* 37**,** 205-210.

BAKIR, S. T. 2004. A distribution-free Shewhart quality control chart based on signed-ranks. *Quality Engineering,* 16**,** 613-623.

CELANO, G., COSTA, A., FICHERA, S. & TROVATO, E. 2008. One-sided Bayesian S2 control charts for the control of process dispersion in finite production runs. *International Journal of Reliability, Quality and Safety Engineering,* 15**,** 305-327.

ELFESSI, A. & REINEKE, D. M. 2001. A Bayesian look at classical estimation: the exponential distribution. *Journal of Statistics Education,* 9.

GELMAN, A. 2002. Prior distribution. *Encyclopedia of environmetrics,* 3**,** 1634-1637.

GHADERINEZHAD, F. & LEY, C. 2019. On the impact of the choice of the prior in Bayesian statistics. *Bayesian inference on complicated data.* IntechOpen.

HAJEJ, Z., NYOUNGUE, A. C., ABUBAKAR, A. S. & MOHAMED ALI, K. 2021. An integrated model of production, maintenance, and quality control with statistical process control chart of a supply chain. *Applied Sciences,* 11**,** 4192.

JOGHEE, R. & MATHEW, I. M. 2025. Influence of process shifts in case of Six Sigma-based Bayesian control charts. *Communications in Statistics-Theory and Methods,* 54**,** 4191-4206.

KALISETTI, Y., KATNENI, N. D. & KRALETI, S. R. 2024. Application of Additive Exponential Distribution in Design of $\bar{X}$ Control Chart-Statistical and Economic Perspective. *Journal of The Institution of Engineers (India): Series C,* 105**,** 437-455.

KANWAL, T. & ABBAS, K. 2025. Comparative analysis of quantile-based control charts for Frechet distribution. *Quality Engineering***,** 1-16.

KOUTRAS, M., BERSIMIS, S. & MARAVELAKIS, P. 2007. Statistical process control using Shewhart control charts with supplementary runs rules. *Methodology and Computing in Applied Probability,* 9**,** 207-224.

LABHA, F., KHAN, N. & TAHIR, M. 2025. A Bayesian Approach to EWMA Control Charts for Enhanced Monitoring of Pareto-Based Processes. *Quality and Reliability Engineering International*.

MAKIS, V. 2008. Multivariate Bayesian control chart. *Operations Research,* 56**,** 487-496.

MOLTCHANOVA, E. 2017. Bayesian real options control chart. *Quality and Reliability Engineering International,* 33**,** 2205-2213.

MURTHY, D. P., XIE, M. & JIANG, R. 2004. *Weibull models,* John Wiley & Sons.

NAFISAH, I. A., ALMAZAH, M. M., AL-REZAMI, A. & HUSSAIN, S. 2025. Variable sample size based EWMA control chart with an exponential scaling mechanism for production process monitoring. *Scientific Reports,* 15**,** 30964.

NASSAR, M. & NADA, N. 2011. The beta generalized Pareto distribution. *Journal of Statistics: Advances in Theory and Applications,* 6**,** 1-17.

RASAY, H., HADIAN, S. M., NADERKHANI, F. & AZIZI, F. 2024. Optimal condition based maintenance using attribute Bayesian control chart. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* 238**,** 754-763.

RASHEED, H. A. & ABD, M. N. 2023. Bayesian Estimation for two parameters of exponential distribution under different loss functions. *Ibn AL-Haitham Journal For Pure and Applied Sciences,* 36**,** 289-300.

RIAZ, M. & ALI, S. 2016. On process monitoring using location control charts under different loss functions. *Transactions of the Institute of Measurement and Control,* 38**,** 1107-1119.

ROBERTS, S. 1966. A comparison of some control chart procedures. *Technometrics,* 8**,** 411-430.

SANTIAGO, E. & SMITH, J. 2014. Control charts based on the exponential distribution: Adapting runs rules for the t chart. *Quality control and applied statistics,* 59**,** 23-24.

SEIRANI, R., TORABIAN, M., BEHZADI, M. H. & SEIF, A. 2023. Economic–statistical design of Bayesian $\bar{X}$ control chart based on the predictive distribution. *International Journal of Quality & Reliability Management,* 40**,** 1925-1939.

SINGH, S. K., SINGH, U. & SHARMA, V. K. 2013. Expected total test time and Bayesian estimation for generalized Lindley distribution under progressively Type-II censored sample where removals follow the beta-binomial probability law. *Applied Mathematics and Computation,* 222**,** 402-419.

SON, Y. S. & OH, M. 2006. Bayesian estimation of the two-parameter Gamma distribution. *Communications in Statistics-Simulation and Computation,* 35**,** 285-293.

SULLIVAN, J. H. & JONES, L. A. 2002. A self-starting control chart for multivariate individual observations. *Technometrics,* 44**,** 24-33.

SUPHARAKONSAKUN, Y. 2024. Empirical Bayes Prediction for an Attribute Control Chart in Quality Monitoring. *IEEE Access*.

TOUW, A. E. 2009. Bayesian estimation of mixed Weibull distributions. *Reliability Engineering & System Safety,* 94**,** 463-473.

## Declaration

**Conflict of Study:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
**Funding Statement:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
**Author Contribution Statement**: All the authors worked together as a cell group and designed the research; Analyzed and interpreted the data, and wrote the paper.
**Availability of Data and Material**: Data will be made available by the corresponding author on reasonable request.
**Ethical Approval**: Not Applicable
**Consent to Publish**: Not Applicable
**Consent to Participate**: Not Applicable
**Acknowledgement**: Not Applicable

## Author(s) Bio / Authors' Note

***Samina Kanwal*** *Samina Kanwal received her Bachelor of Science degree from Post Graduate College and completed her M.Phil. degree in Statistics from University of Agriculture Faisalabad. Her research interests include Bayesian statistics, Bayesian control charts, process capability analysis, machine learning, and quality control. Her research work focuses on comparing classical and Bayesian approaches using different probability distributions and prior distributions. She has experience in statistical computing software including R, SPSS, and Minitab. Currently, she is working as a visiting faculty teacher and is actively engaged in research in applied and Bayesian statistics. Email:* saminakanwal189@gmail.com

***Muhammad Zafar Iqbal*** *from the Department of Mathematics and Statistics, University of Agriculture Faisalabad, Pakistan. Email:* mzts2004@uaf.edu.pk

***Muhammad Sohaib Asad*** *Muhammad Sohaib Asad received his M.Phil. degree in Statistics from University of Agriculture Faisalabad and is currently pursuing his Ph.D. in Statistics at the same institution. His research interests include control charts, Bayesian control charts, process capability analysis, and bootstrap confidence intervals. His research work focuses on statistical quality control and the application of Bayesian and resampling techniques for process monitoring and improvement. He has experience in statistical computing and data analysis using R, SPSS, and Minitab. Currently, he is serving as a Visiting Lecturer at Government College University Faisalabad while actively engaged in academic research. Email:* msohaibasad@gmail.com

***Muhammad Kashif*** *Dr. Muhammad Kashif received his M.Sc. and M.Phil. degree in Statistics from University of Agriculture, Faisalabad (UAF), Pakistan and Ph.D. degree from King Abdul Aziz, University Jeddah, Saudi Arabia. He has worked as researcher at various reputable research institutes of Pakistan. Currently, he has been working as Assistant Professor in the Department of Mathematics and Statistics, UAF. He has authored and co-authored over 50 technical papers and 1 book chapter in applied Statistics. His research interest include non-normal distribution, process capability indices, time series analysis , design of experiments and Multivariate Techniques. Dr. Kahif have also command on modern statistical packages like R, Python, SPSS, and Minitab. Email:* *mkashif@uaf.edu.pk*

## Appendix

COMPLETE FIGURES OF TABLES 1 TO 9

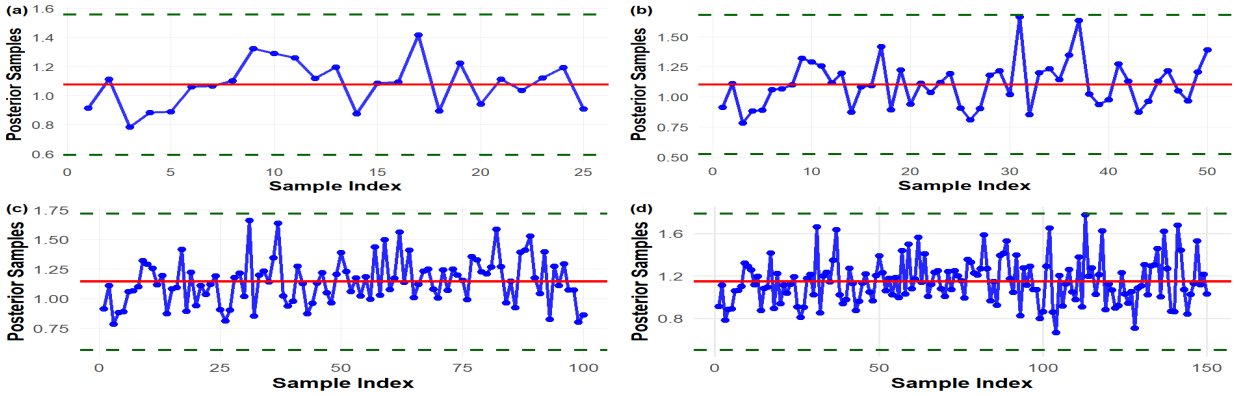


**Figure 2:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =0.5 at different sample sizes.

Note: Figures 2–5 represent the graphical results corresponding to Table 1.

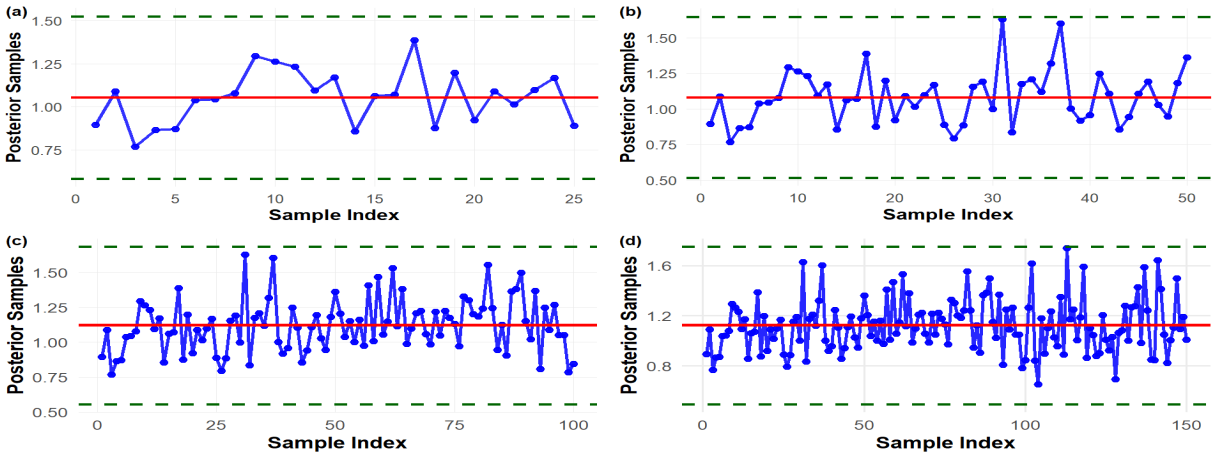


**Figure 3:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =1.0 at different sample sizes.

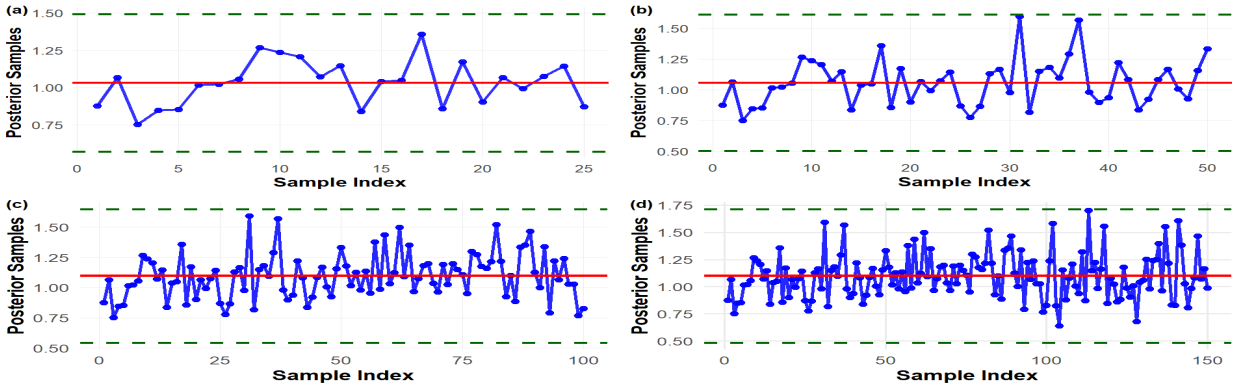


**Figure 4:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =1.5 at different sample sizes.

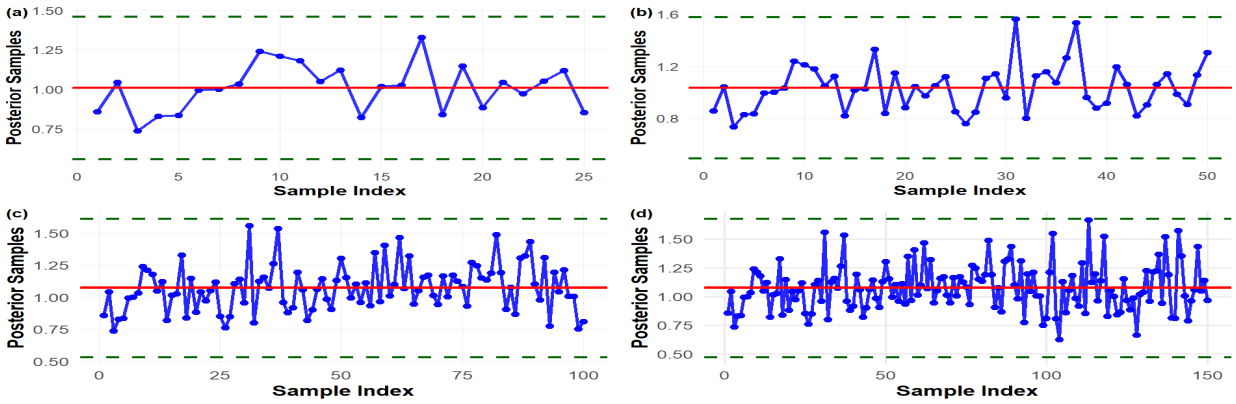


**Figure 5:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1 and β =2.0 at different sample sizes.

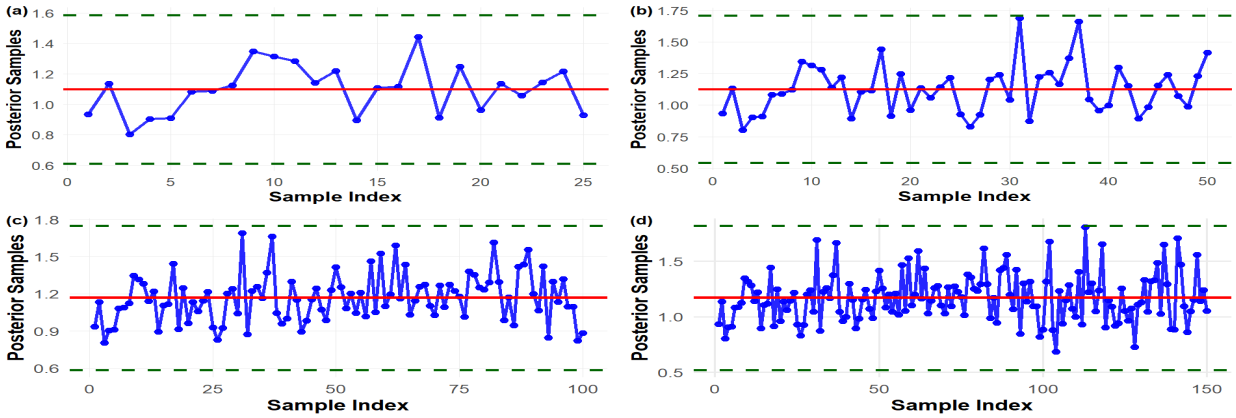


**Figure 6:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1.5 and β = 0.5 at different sample sizes.

Note: Figures 6–9 represent the graphical results corresponding to Table 2.

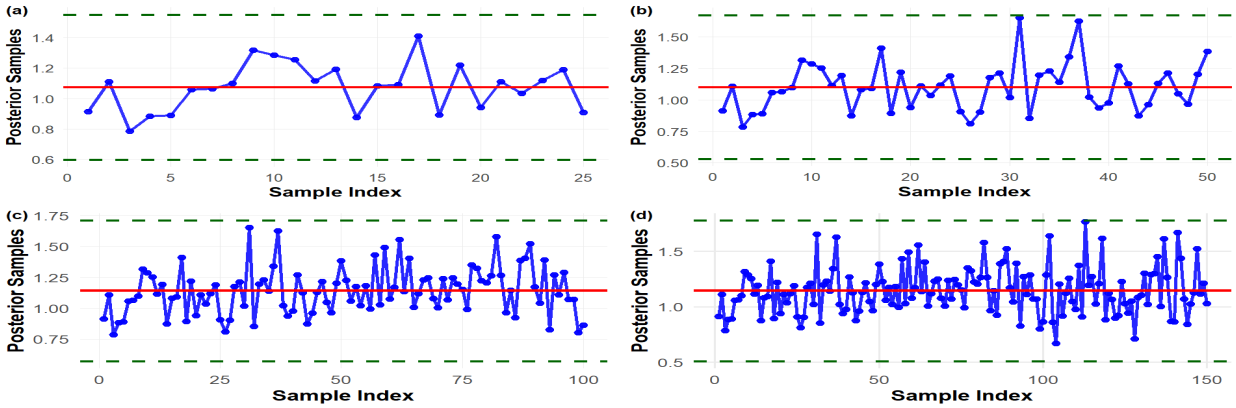


**Figure 7:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1.5 and β = 1.0 at different sample sizes.

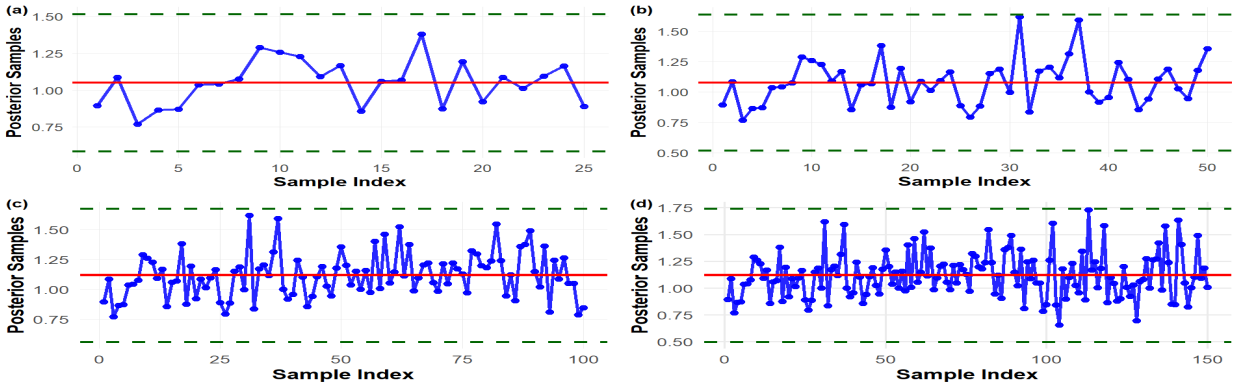


**Figure 8:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1.5 and β = 1.5 at different sample sizes.

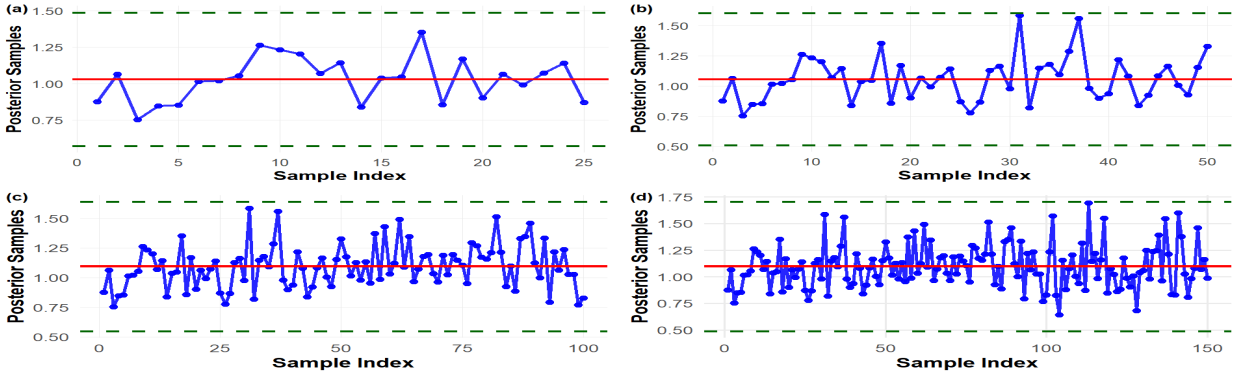


**Figure 9:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =1.5 and β = 2 at different sample sizes.

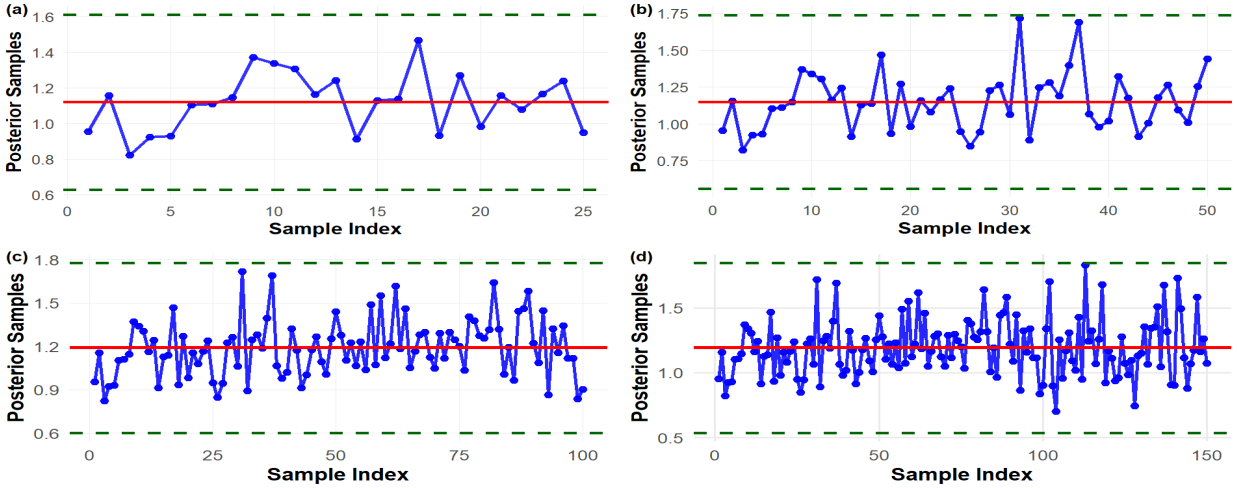


**Figure 10:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =2 and β = 0.5 at different sample sizes.

Note: Figures 10–13 represent the graphical results corresponding to Table 3.

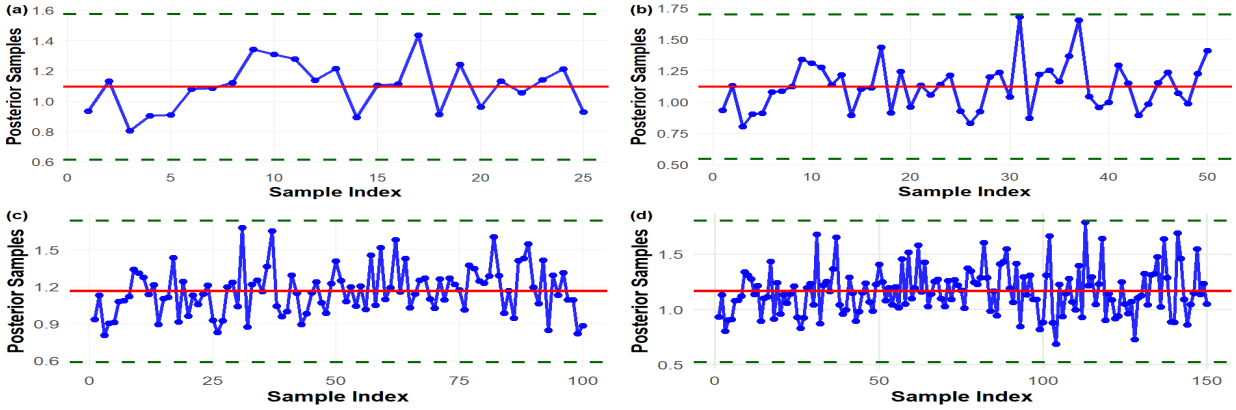


**Figure 11:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =2 and β = 1.0 at different sample sizes.

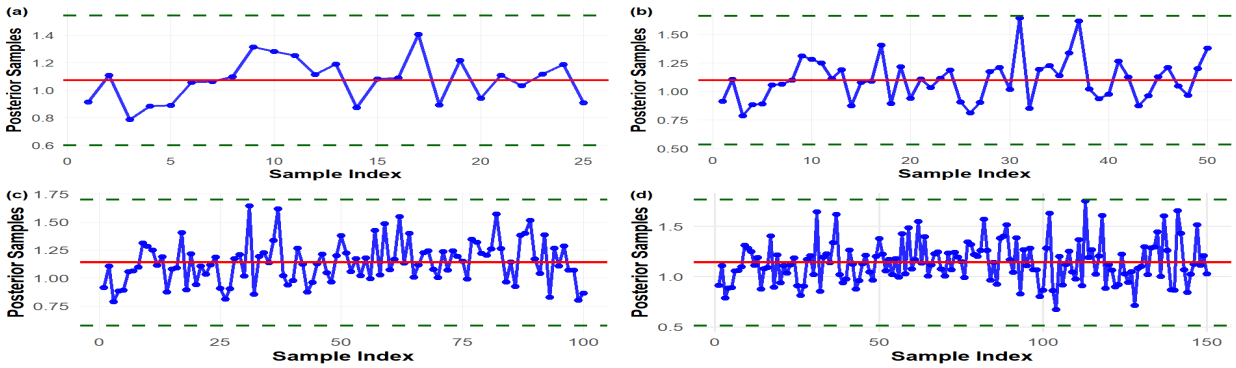


**Figure 12:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =2 and β = 1.5 at different sample sizes.

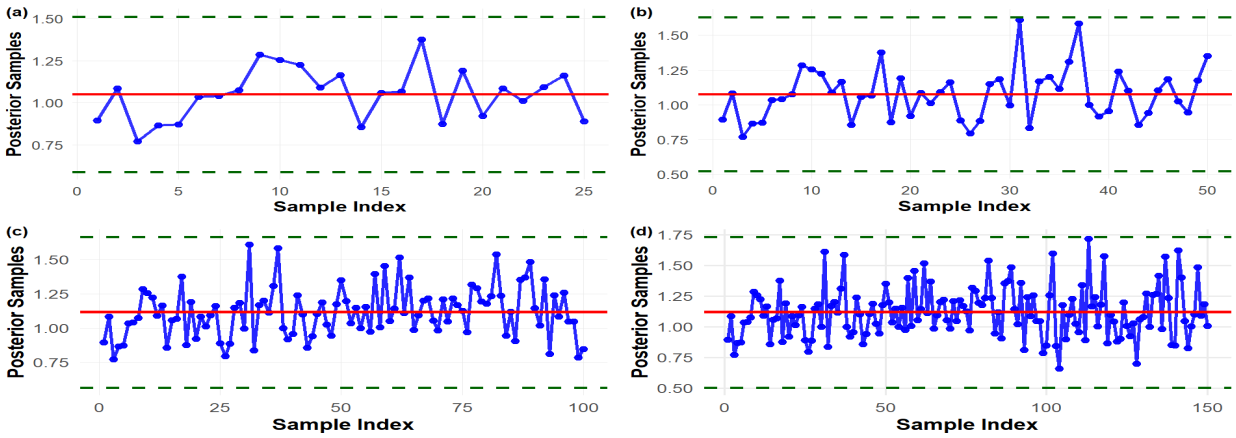


**Figure 13:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1 by using Gamma Prior with α =2 and β = 2.0 at different sample sizes.

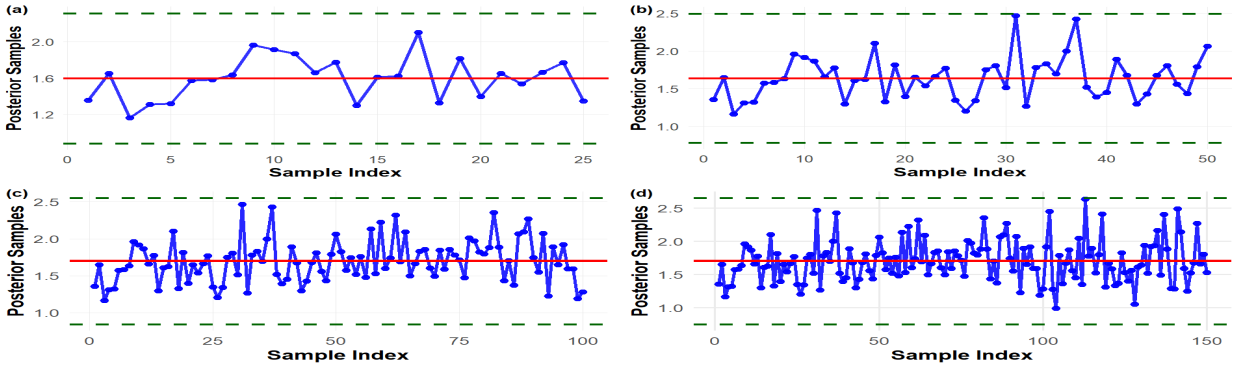


**Figure 14:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β = 0.5 at different sample sizes.

Note: Figures 14–17 represent the graphical results corresponding to Table 4.

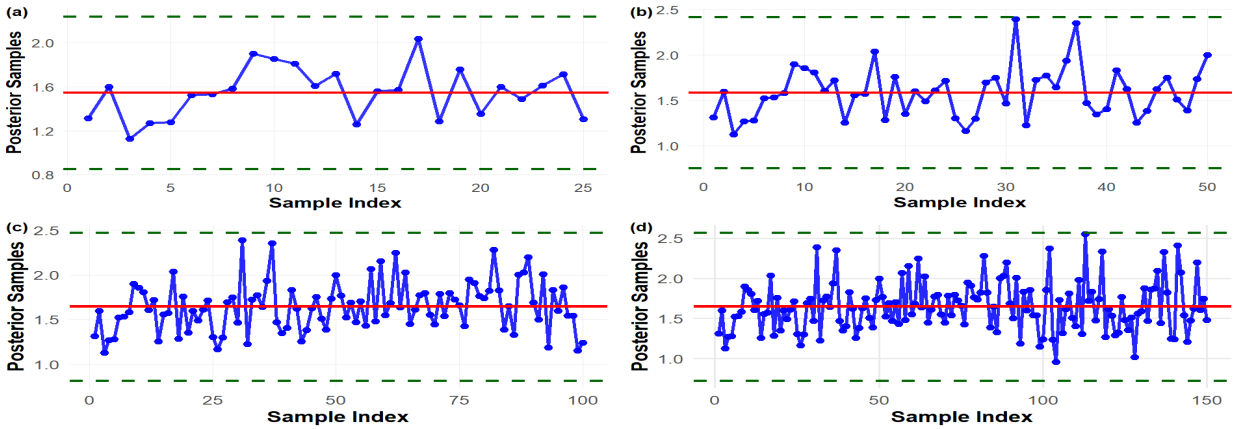


**Figure 15:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β = 1.0 at different sample sizes.

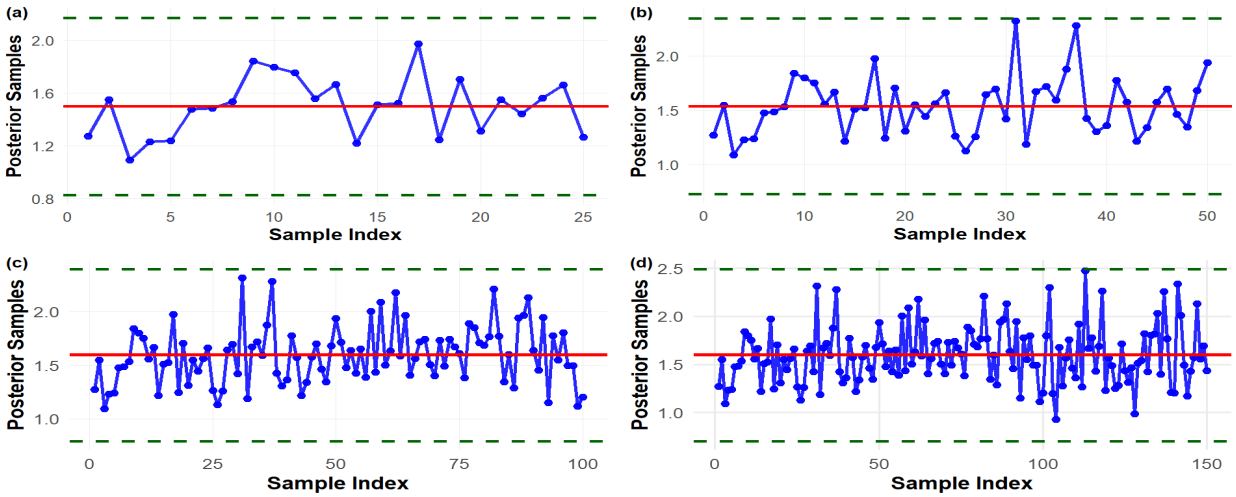


**Figure 16:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β = 1.5 at different sample sizes.

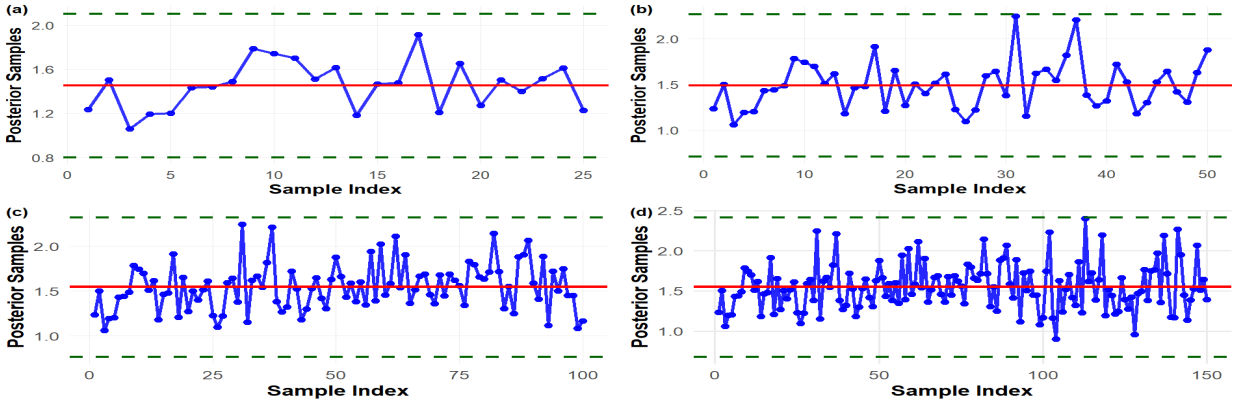


**Figure 17:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.0 and β = 2.0 at different sample sizes.

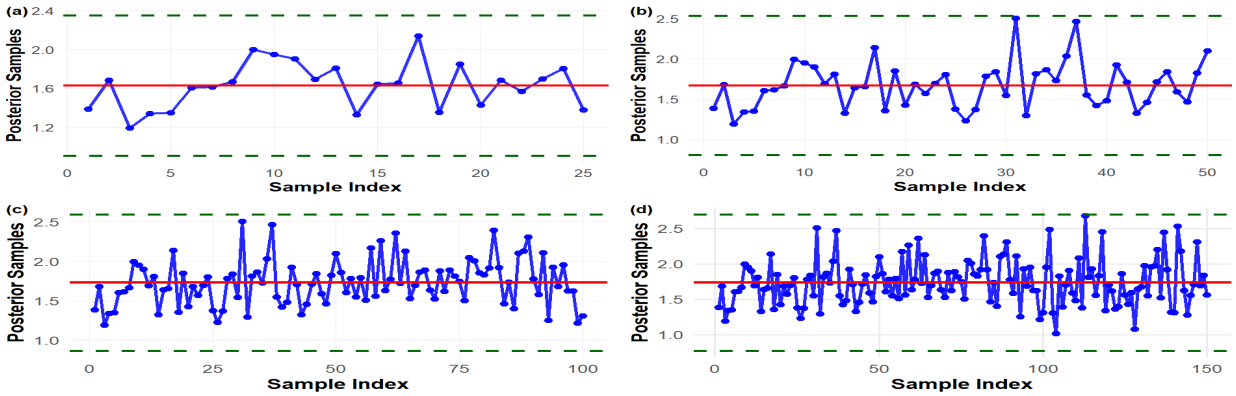


**Figure 18:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.5 and β = 0.5 at different sample sizes.

Note: Figures 18–21 represent the graphical results corresponding to Table 5.

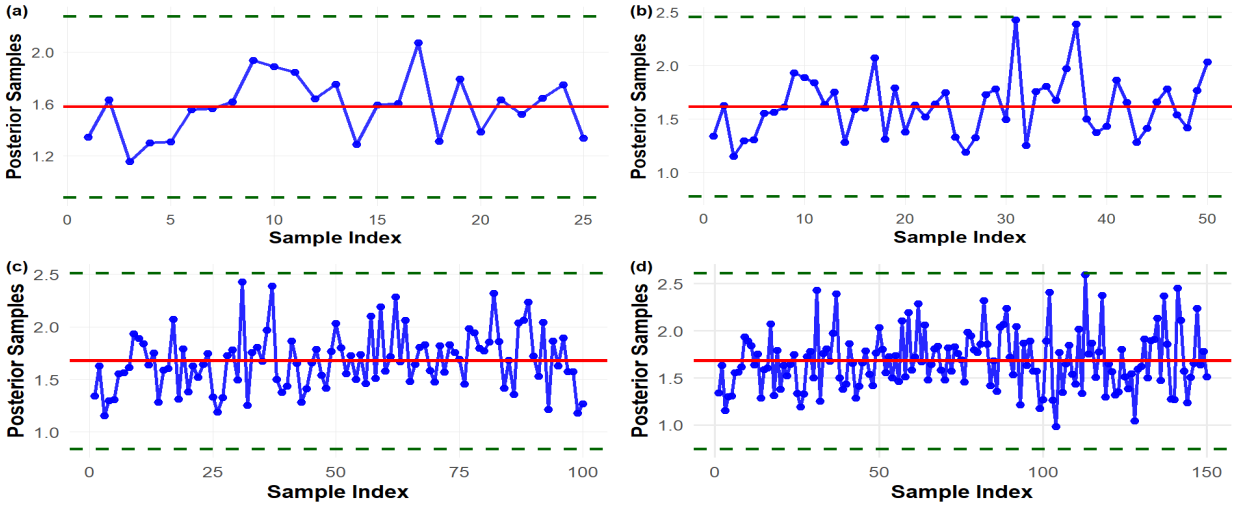


**Figure 19:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.5 and β = 1 at different sample sizes.

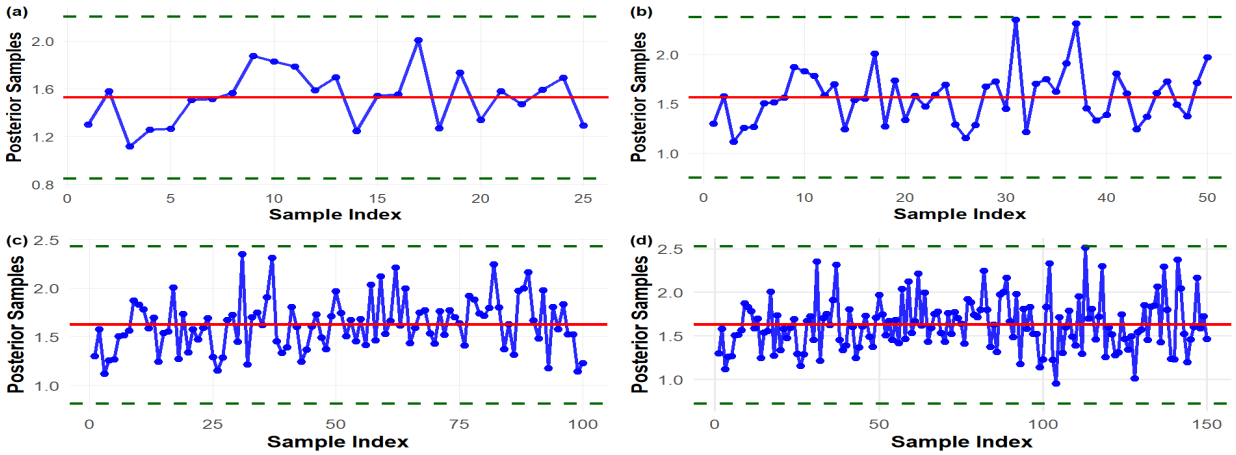


**Figure 20:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.5 and β = 1.5 at different sample sizes.

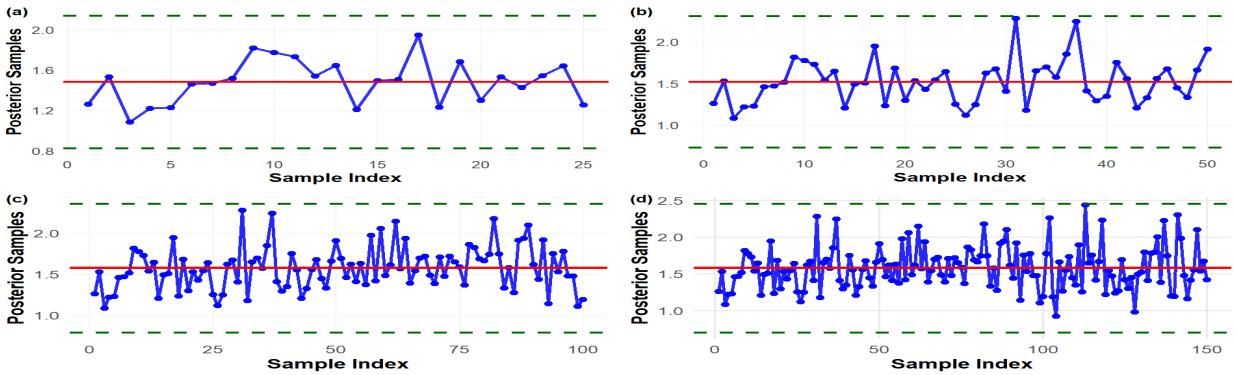


**Figure 21:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =1.5 and β = 2.0 at different sample sizes.

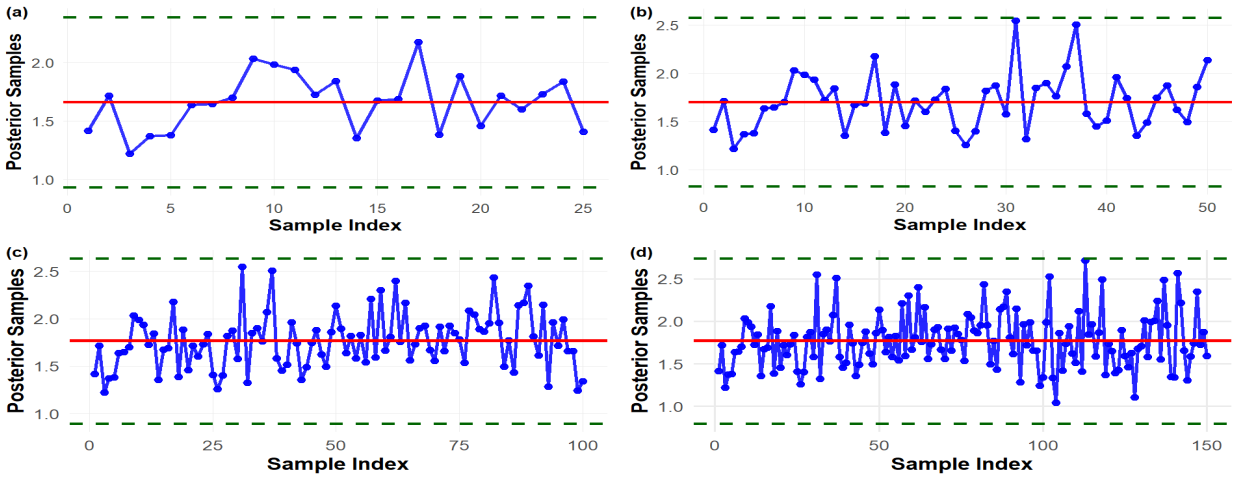


**Figure 22:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =2.0 and β = 0.5 at different sample sizes.

Note: Figures 22–25 represent the graphical results corresponding to Table 6.

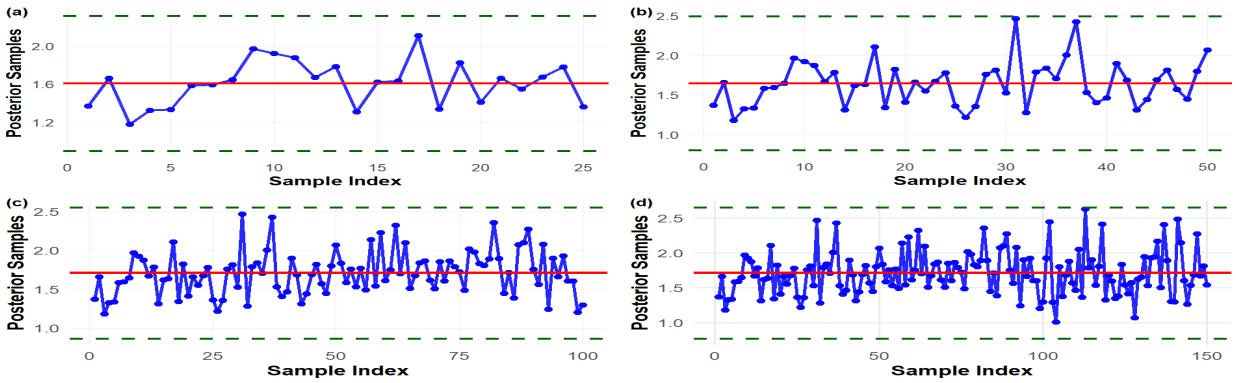


**Figure 23:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =2.0 and β = 1.0 at different sample sizes.

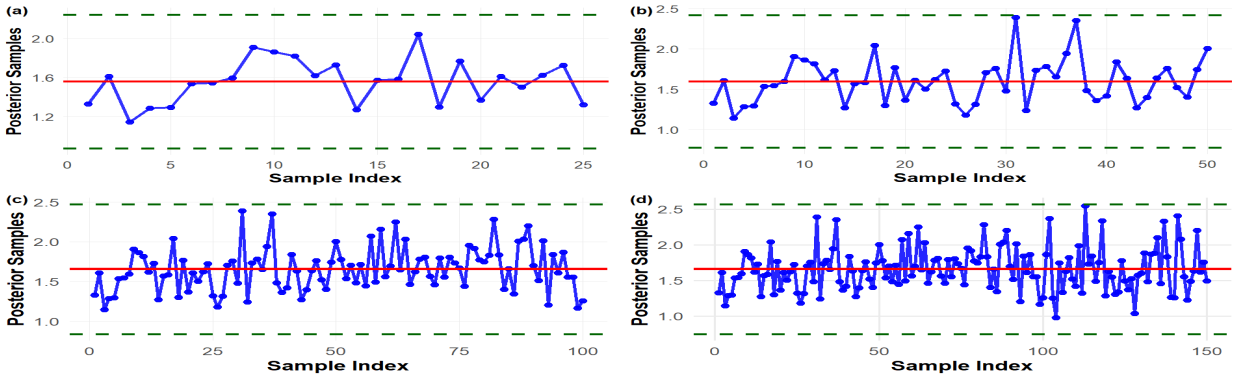


**Figure 24:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =2.0 and β = 1.5 at different sample sizes.

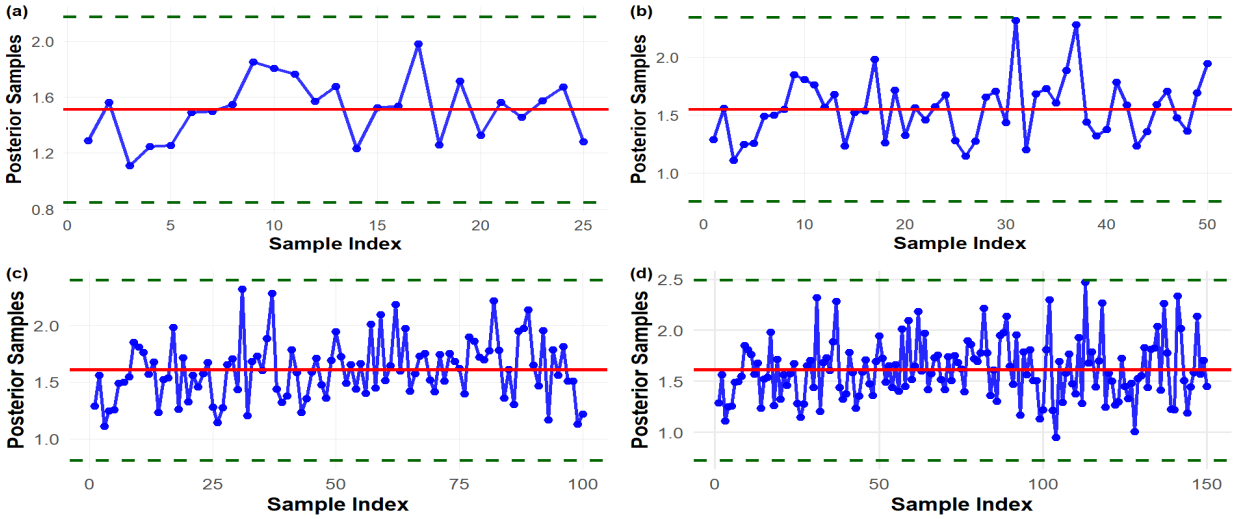


**Figure 25:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =1.5 by using Gamma Prior with α =2.0 and β = 2.0 at different sample sizes.

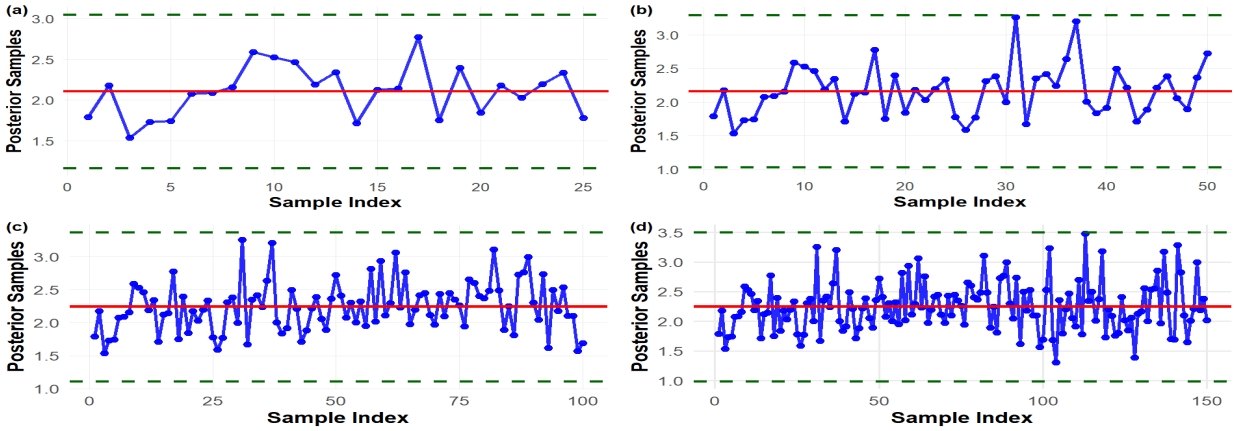


**Figure 26:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.0 and β = 0.5 at different sample sizes.

Note : Figures 26–29 represent the graphical results corresponding to Table 7.

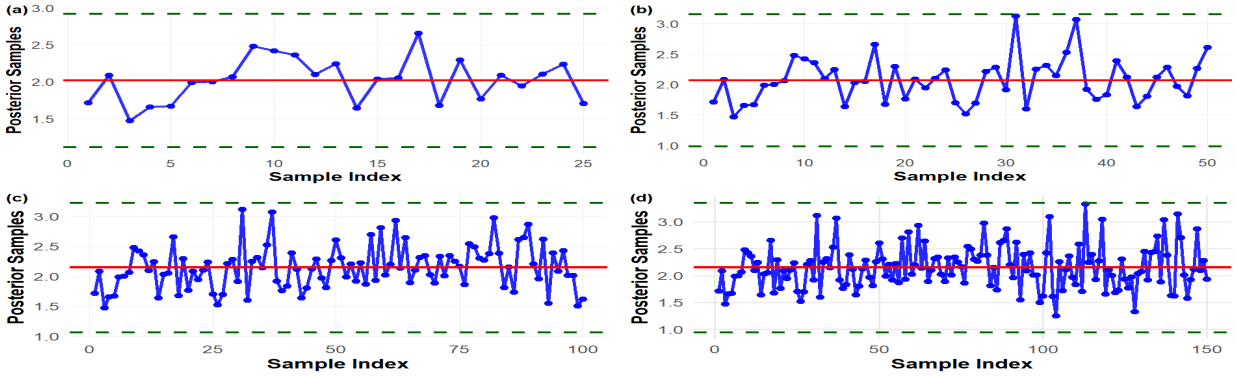


**Figure 27:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.0 and β = 1.0 at different sample sizes.

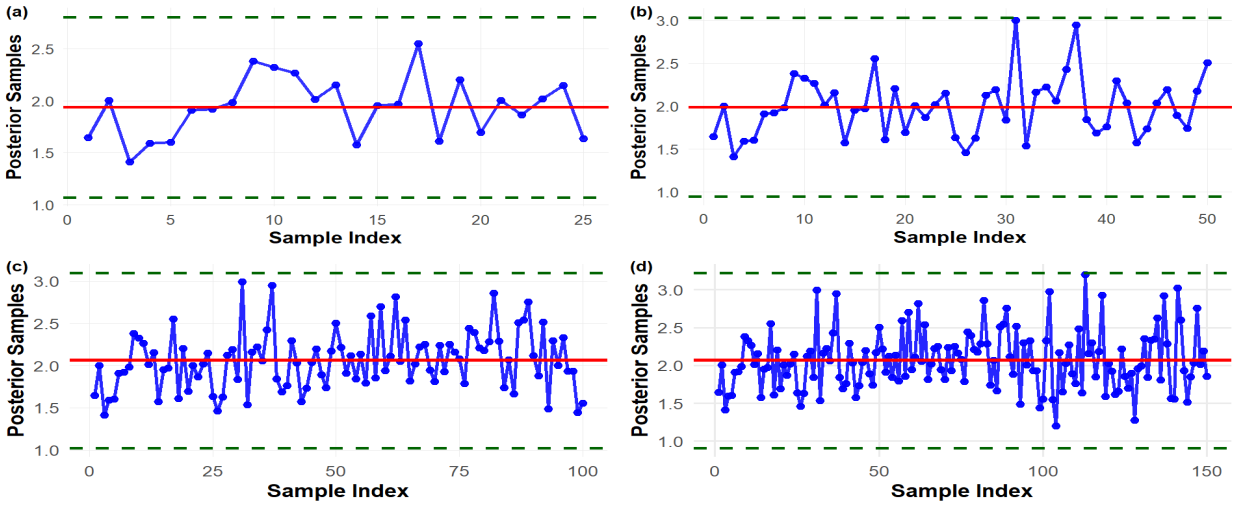


**Figure 28:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.0 and β = 1.5 at different sample sizes.

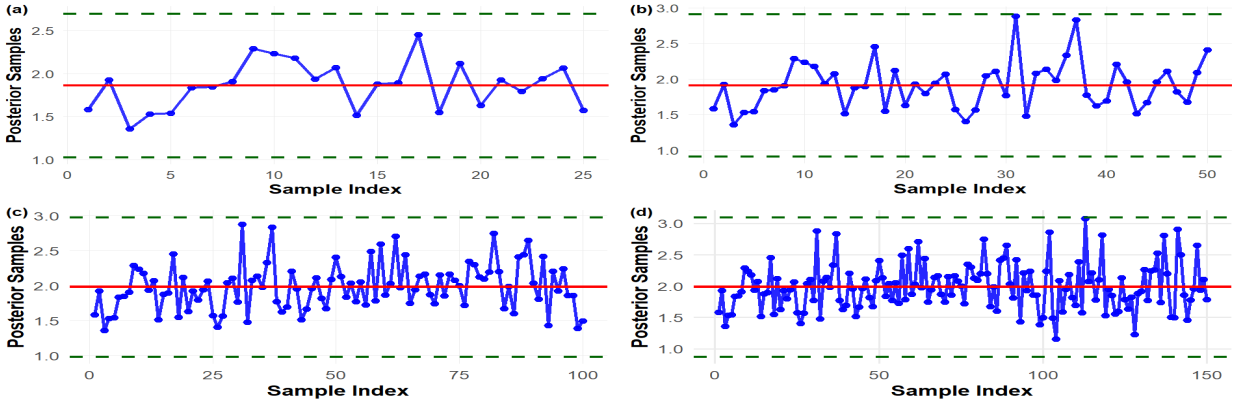


**Figure 29:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.0 and β = 2.0 at different sample sizes.

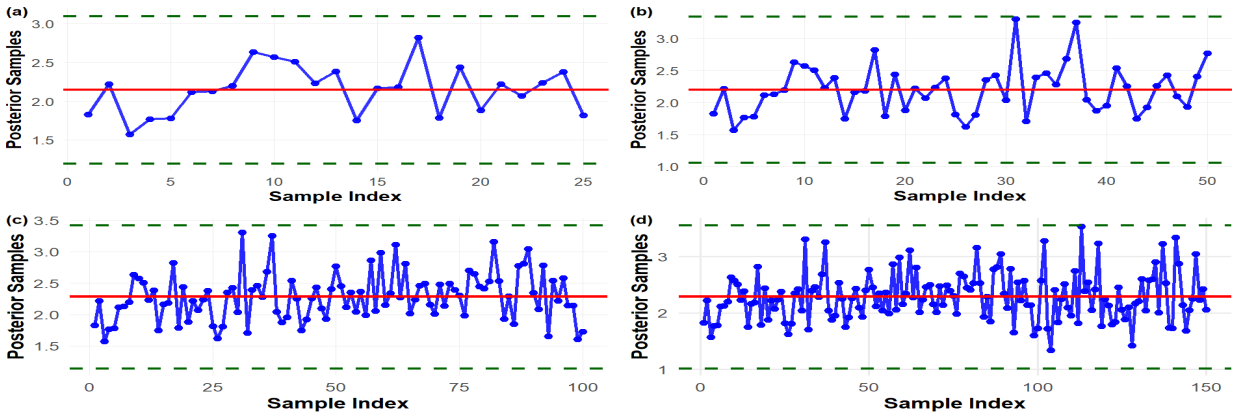


**Figure 30:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.5 and β = 0.5 at different sample sizes.

Note: Figures 30–33 represent the graphical results corresponding to Table 8.

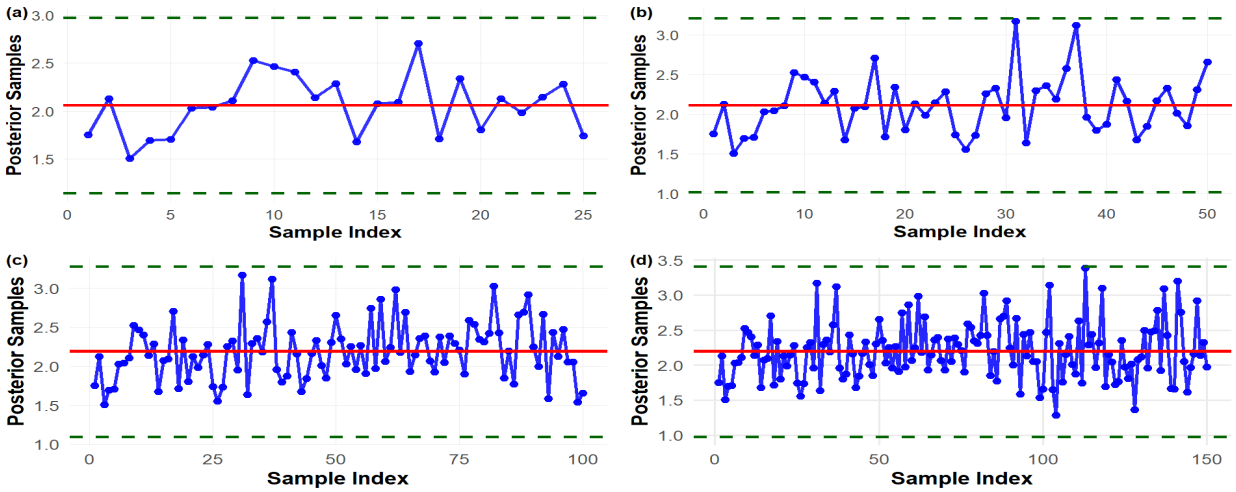


**Figure 31:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.5 and β = 1.0 at different sample sizes.

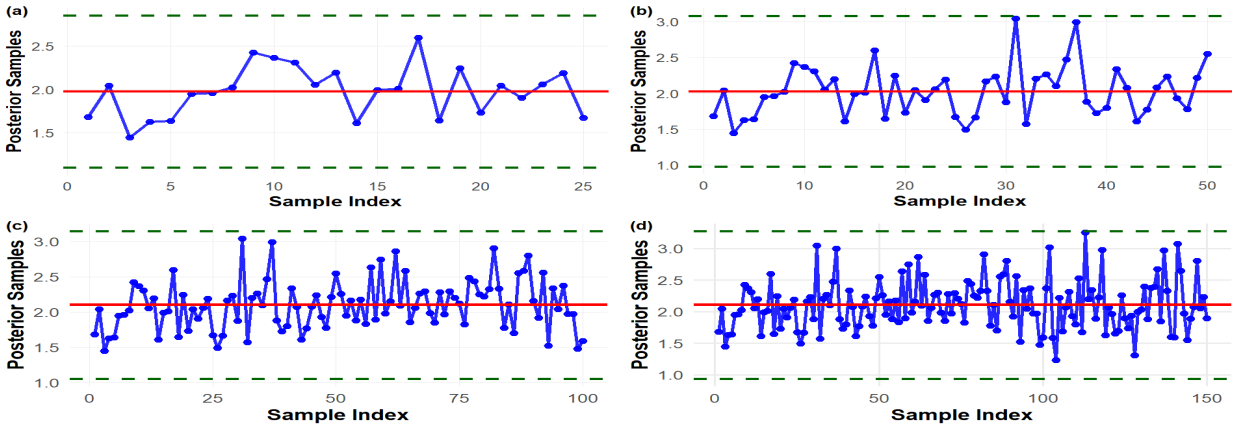


**Figure 32:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.5 and β = 1.5 at different sample sizes.

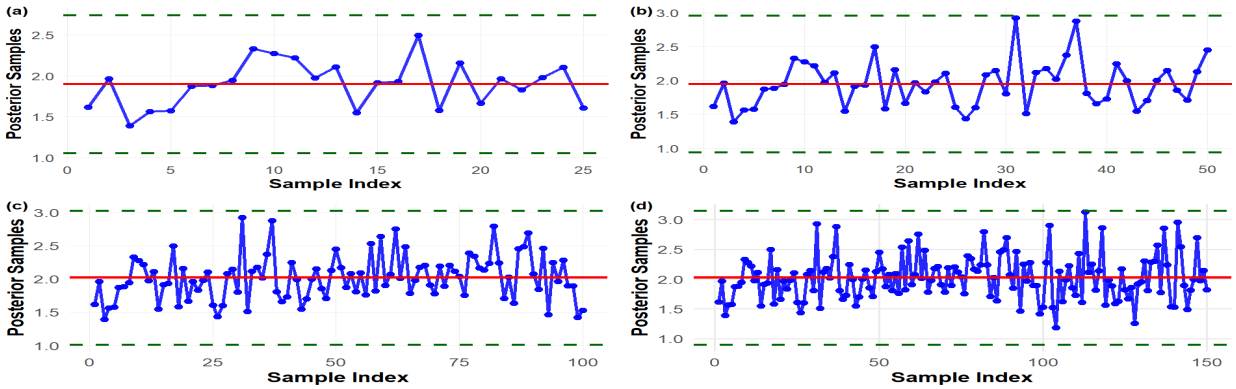


**Figure 33:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =1.5 and β = 2.0 at different sample sizes.

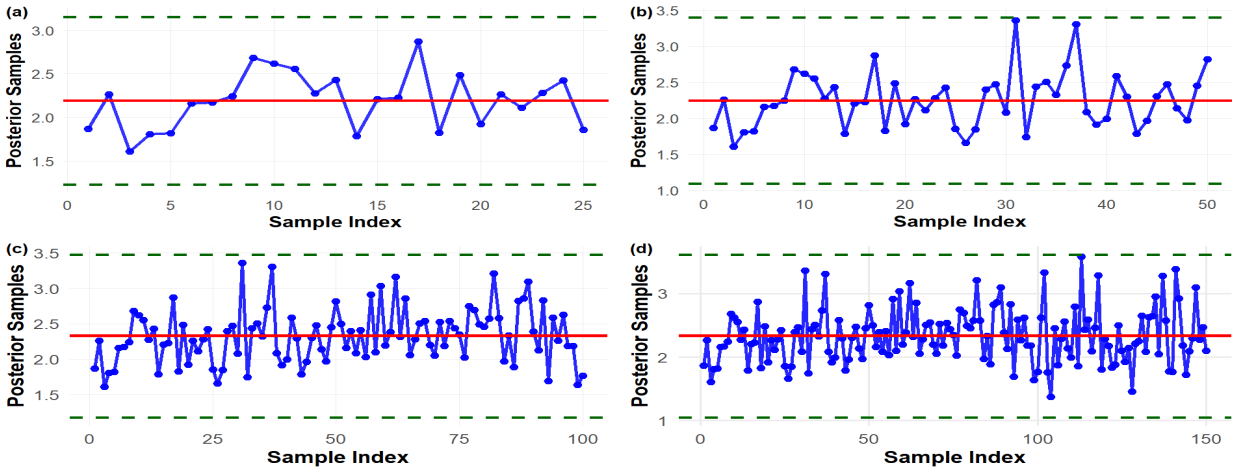


**Figure 34:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =2.0 and β = 0.5 at different sample sizes.

Note : Figures 34–37 represent the graphical results corresponding to Table 9.

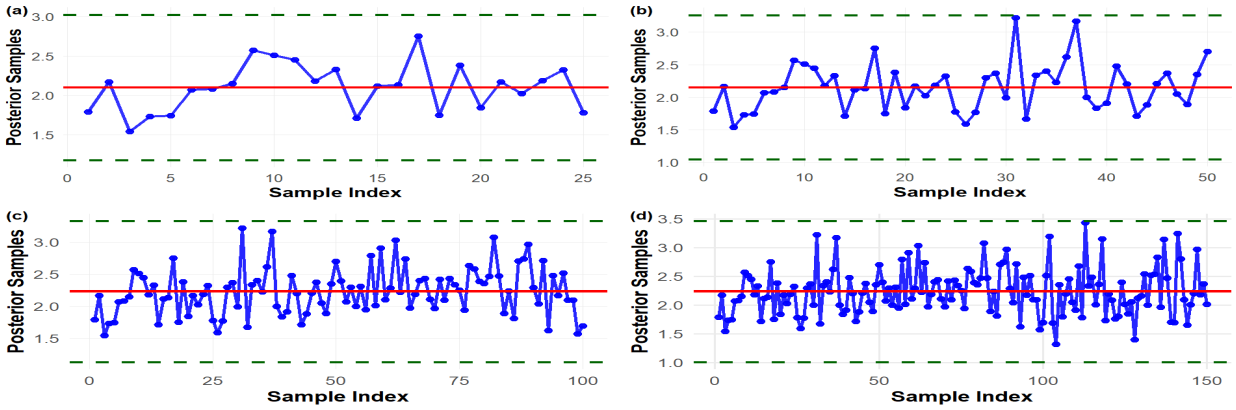


**Figure 35:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =2.0 and β = 1.0 at different sample sizes.

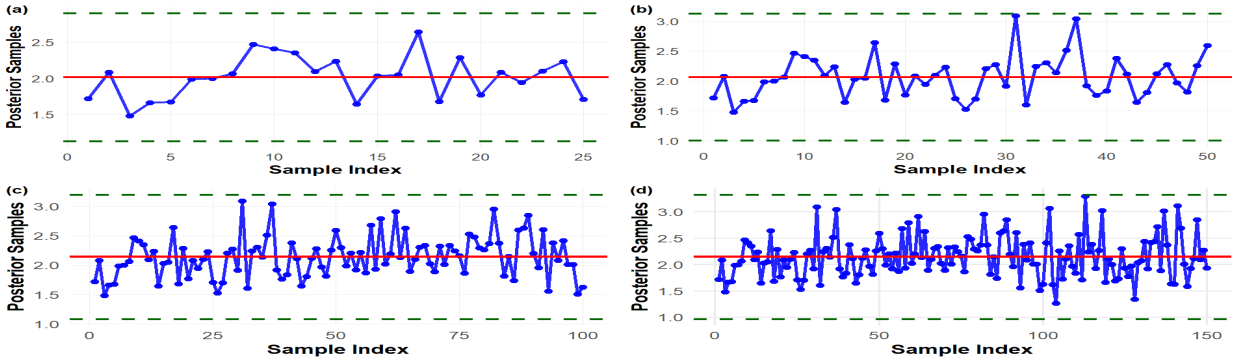


**Figure 36:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =2.0 and β = 1.5 at different sample sizes.

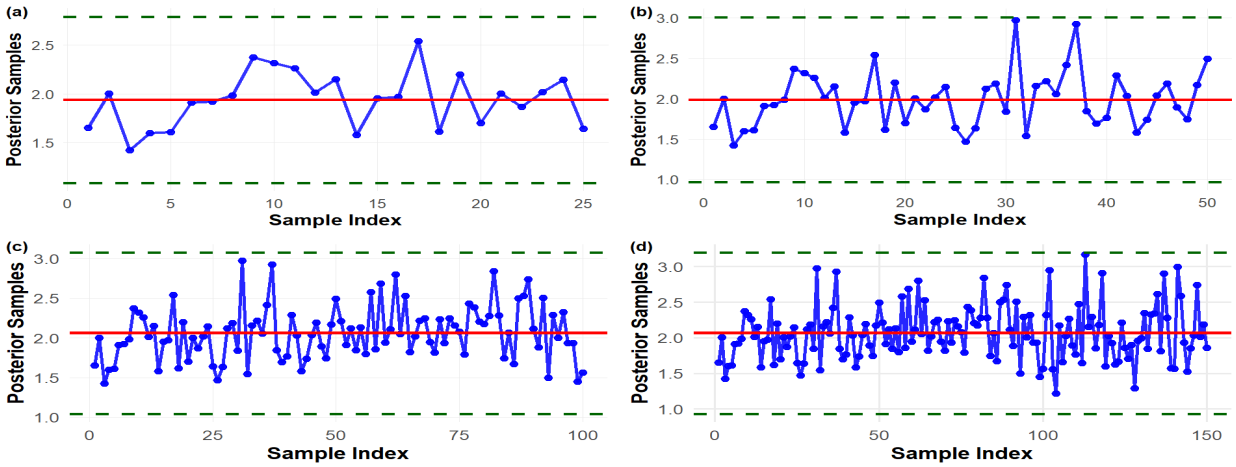


**Figure 37:** Bayesian Control Chart for the Posterior distribution of Exponential distribution with θ =2.0 by using Gamma Prior with α =2.0 and β = 2.0 at different sample sizes.

SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 2, June 2026
DOI: https://doi.org/10.64060/JASR2V2i5


# Development and Statistical Analysis of Modified Kies Perks Distribution

Sarwat Rashid [1*]

[1]Forman Christian College (A Chartered University) Lahore, Pakistan

* Corresponding Email: sarwatrashid1984@gmail.com

Received: 18 March 2026 / Revised: 01 May 2026 / Accepted: 10 May 2026 / Published online: 17 May 2026



ABSTRACT

This paper introduces a novel development of the Perks Distribution called the Modified Kies Perks distribution (MKPD), a highly adaptable probabilistic model characterized by its versatile probability density function, distribution function (DF) and hazard rate function. The MKPD can take various shapes, such as a single peak (unimodal), leaning to the right (right skewed), leaning to the left (left-skewed), a curve resembling a bathtub, and more, demonstrating its flexibility in statistical modelling and it is one of the most important and well-established ideas in reliability analysis. Several fundamental characteristics of the proposed model have been derived, including its PDF, CDF, quantile function, Hazard rate, survival function, cumulative hazard function, Odds Ratio and Average Failure Rate. Furthermore, a Monte Carlo simulation study is performed to assess the efficiency of different estimation methods across varying sample sizes confirming the MKPD's flexibility and strong fit for lifetime, epidemiological and actuarial data. To validate its practicality, the proposed distribution is applied to real datasets. The results demonstrate that the MKPD is a valuable contribution to both statistical theory and applied statistical work.

**Keywords**:Perk distribution; Modified Kies distribution; estimation; Reliability measures

## 1. Introduction

Statistical modelling of lifetime and reliability data plays an essential role in many scientific disciplines, including engineering, biomedical sciences, actuarial studies, and risk analysis. Reliable statistical models enable researchers to analyse failure mechanisms, estimate survival probabilities, and make informed decisions regarding system reliability. Classical probability distributions, such as exponential, gamma, and Weibull distributions, have been widely used in reliability and survival analysis due to their mathematical simplicity and interpretability. However, these classical models often lack sufficient flexibility to represent complex data patterns, particularly when the hazard rate function exhibits non-monotonic behaviours such as bathtub-shaped or unimodal patterns (Kotz & Balakrishnan, 2012).

The development of flexible lifetime distributions has been a central focus in reliability and actuarial analysis. Early contributions by William Perks laid the foundation through the introduction of the Perks distribution for modeling mortality data (Perks, 1932), while Harold A. Kies extended this framework by proposing the Kies family of distributions for life testing applications (Kies, 1958, 1959). Subsequent studies have introduced several modifications and generalizations of the Kies distribution to enhance its flexibility and applicability to lifetime data, including modified and power versions that accommodate diverse data behaviors (El-Bassiouny et al., 2016; Yousof et al., 2017; Shanker & Mishra, 2013). In parallel, a wide range of classical distributions such as the Weibull, exponential, beta, and Pareto distributions have been extensively studied and applied in modeling failure rates and survival data (Weibull, 1951; Balakrishnan & Basu, 1995; Nadarajah & Kotz, 2003; Arnold, 2015).

More recently, researchers have focused on developing generalized and transmuted families of distributions using innovative generator techniques to improve modeling flexibility and capture complex characteristics such as skewness, heavy tails, and varying hazard rate shapes (Cordeiro & de Castro, 2011; Nofal et al., 2017; Afify et al., 2016). Methods for generating new distribution families have also been widely explored, enabling the construction of models with enhanced adaptability for real-world data (Alzaatreh et al., 2013; Baharith & Aljuhani, 2021). Contemporary studies continue to

advance this area by proposing new lifetime distributions through transformation and stochastic generator approaches, demonstrating improved performance in biomedical, environmental, and engineering applications (Bakr et al., 2022; Elbatal et al., 2022; Yusuf et al., 2024; Dutta & Yadav, 2025). Despite these advancements, there remains a need for more flexible models that combine rich shape characteristics with tractable properties and robust estimation performance. This gap motivates further extensions of the Kies–Perks framework, leading to the development of more versatile models such as the Modified Kies Perks Distribution.

The Weibull distribution, introduced by Waloddi Weibull, has been extensively applied for modelling failure time data due to its flexibility in representing increasing and decreasing hazard rates. Similarly, the Lindley distribution, proposed by Dennis V. Lindley, has found widespread applications in reliability analysis and Bayesian statistics. Despite their usefulness, these classical distributions are sometimes inadequate for modelling real datasets with complex shapes, skewness, and varying failure patterns. One of the earliest contributions to actuarial and mortality modelling was made by William Perks, who introduced the Perks distribution to model mortality rates in life tables. The Perks model improved representation of mortality patterns compared with simpler models used at the time. Later, Harold A. Kies introduced the Kies distribution for life testing experiments, providing a flexible framework for constructing probability models capable of capturing diverse failure behaviors. Subsequent research explored extensions of the Kies distribution to improve flexibility and applicability. For instance, Abd El-Basit El-Bassiouny and colleagues proposed a modified Kies distribution with additional parameters to capture complex lifetime patterns, while Hesham M. Yousof and co-authors developed the power Kies distribution through power transformations. In parallel, significant research has focused on developing generalized families of distributions by applying transformation or generator mechanisms to baseline models. Generator-based approaches introduce additional shape parameters that enhance modelling flexibility. The method proposed by Fathi Alzaatreh et al. (2013) has been widely adopted to construct new families of probability distributions, and generalized distribution families developed by Gauss M. Cordeiro and collaborators have attracted attention for modeling skewed and heavy-tailed data. Recent studies, including those by Elbatal et al. (2022), Bakr et al. (2022), Baharith and Aljuhani (2021), Yusuf et al. (2024), Dutta and Yadav (2025), Mecha et al. (2025), and Pandey et al. (2024), have demonstrated the effectiveness of generator-based methods in constructing flexible distributions for engineering, biomedical, and actuarial applications.

Despite these advances, there remains a need for distributions that combine the structural characteristics of classical models with the flexibility of modern generator techniques. Motivated by these developments, this study proposes a new probability distribution called the Modified Kies Perks Distribution (MKPD). The MKPD integrates the structural properties of the Perks distribution with the flexibility of the Kies transformation, introducing additional parameters that allow the probability density function and hazard rate function to assume various shapes. This makes the MKPD suitable for modelling diverse types of lifetime data. The main objectives of this study are to develop the mathematical formulation of the MKPD, investigate its statistical properties, and perform parameter estimation using the maximum likelihood method. A Monte Carlo simulation study is conducted to evaluate estimator performance, and the applicability of the proposed model is demonstrated using real datasets. The proposed approach provides a flexible and robust framework for lifetime and reliability analysis, addressing limitations of both classical and existing generalized distributions.

The growing complexity of real-world data in fields such as reliability analysis, survival studies, epidemiology, and actuarial science has created a strong demand for flexible probability distributions capable of capturing diverse behaviors including skewness, heavy tails, and varying hazard rate shapes such as increasing, decreasing, and bathtub forms. Although classical models, including the Perks distribution, have been widely used due to their theoretical importance, they often lack the flexibility required to accurately model such complex data patterns. Existing extensions of these models also remain limited, as many fail to simultaneously provide adaptable shape characteristics, tractable mathematical properties, and strong performance across different datasets. Furthermore, there is a noticeable lack of comprehensive studies that evaluate multiple estimation techniques under varying sample sizes using simulation methods, along with limited real data validation to confirm practical applicability. These limitations highlight a clear research gap in developing a more versatile and efficient model. To address this gap, the Modified Kies Perks Distribution (MKPD) is proposed as a highly flexible extension that accommodates a wide range of distributional shapes and hazard rate behaviors, while also offering well-defined statistical properties and improved estimation performance, making it suitable for both theoretical development and practical applications

## 2. Methodology

Unser this section the development of the MKPD is explained, the PDF and CDF of the proposed models along with graphical representation is shown.

### *2.1 The Modified Kies (MK) Family*

The Modified Kies family is a modern class of statistical distributions designed to extend classical lifetime models, offering greater flexibility in reliability analysis, survival studies, and actuarial applications. It generalizes baseline distributions to

capture diverse hazard rate shapes, including monotonic, non-monotonic, and bathtub curves. Al-Babtain et al (2020) proposed a modified Kies (MKi) family with the CDF and PDF as follows:

$$F(x;\omega) = 1 - \exp\left(-\left[\frac{G(x;\omega)}{1-G(x;\omega)}\right]^{\gamma}\right); x > 0, \gamma > 0 \quad (1)$$

$$f(x;\omega) = \gamma \cdot g(x;\omega) \cdot [G(x;\omega)]^{\gamma-1}. \quad [1 - G(x;\omega)]^{-(\gamma+1)} \cdot \exp\left(-\left[\frac{G(x;\omega)}{1-G(x;\omega)}\right]^{\gamma}\right); \quad x > 0,\ \gamma > 0 \quad (2)$$

### *2.2 Background of the Perks Distribución*

The Perks distribution is a flexible statistical model widely applied in reliability engineering and actuarial science, particularly for analyzing lifetime data. Its distinctive feature lies in its ability to capture both monotonic and non-monotonic failure rate behaviors, including the well-known bathtub-shaped hazard function. In this study, we adopt the Perks distribution, parameterized by α and λ, as the baseline model for subsequent analysis.

So, the CDF and PDF of Perks Distribution are

$$G(x) = \frac{\alpha(e^{\lambda x}-1)}{1+\alpha e^{\lambda x}} \quad ; \quad \alpha, \lambda > 0, x > 0 \quad (3)$$

$$g(x) = \frac{\alpha\lambda e^{\lambda x}(1+\alpha)}{(1+\alpha e^{\lambda x})^2}; \quad \alpha, \lambda > 0, x > 0 \quad (4)$$

Where $\alpha, \lambda$ *are shape and scale parameter respectively*

### *2.3 Formulation of the Proposed Distribution*

By embedding the Perks distribution within the Modified Kies family framework, this study establishes a robust baseline for comparative analysis of lifetime models. This integration enables a systematic exploration of their structural properties, parameter estimation methods, and practical applications in reliability and survival contexts. Specifically, we propose the Modified Kies Perks Distribution (MKPD), constructed by applying the Kies transformation to the Perks distribution. The resulting model is characterized by three positive parameters, $\alpha$, $\lambda$, and $\gamma$, which collectively enhance its flexibility in capturing diverse hazard rate behaviors

Using the CDF and PDF in equations (3) and (4) in equations (1) and (2), the cumulative distribution function (CDF) and probability density function (PDF)of the MKPD is defined as

$$F(x;\alpha,\lambda,\gamma) = 1 - exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right); x > 0, \alpha, \lambda, \gamma > 0 \quad (4)$$

$$f(x;\alpha,\lambda,\gamma) = \frac{\gamma\alpha^{\gamma}\lambda e^{\lambda x}}{(1+\alpha)^{\gamma}}.\left(e^{\lambda x}-1\right)^{\gamma-1} exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right); x > 0, \alpha, \lambda, \gamma > 0 \quad (5)$$

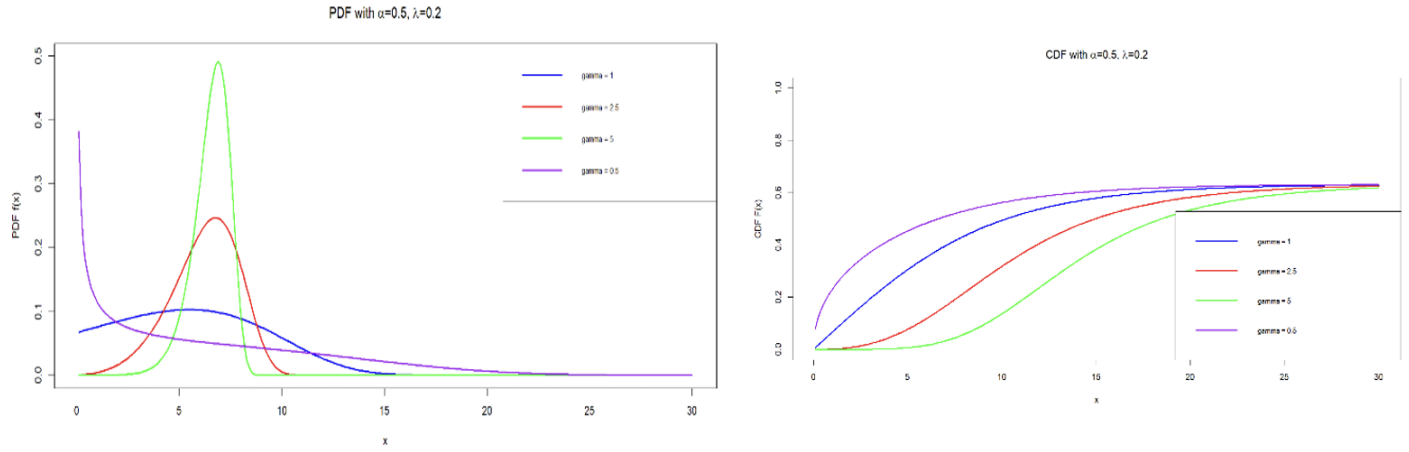


**Figure 1**: PDF plot of MKPD (left) and CDF plot of MKPD (right)

**Theorem 1.** Show that the CDF in Eq. (4) is valid CDF.

**Proof.** Consider X follows the MKPD with CDF in Eq. (4), then

$$F(-\infty) = F(0) = 1 - exp\left(-\left(\frac{\alpha\left(e^{\lambda(0)}-1\right)}{1+\alpha}\right)^{\gamma}\right) = 1 - Exp(0) = 0$$

$$F(\infty) = 1 - exp\left(-\left(\frac{\alpha\left(e^{\lambda(\infty)}-1\right)}{1+\alpha}\right)^{\gamma}\right) = 1 - Exp(-\infty) = 1$$

So, CDF in Eq. (4) is a valid CDF.

## 3. Properties of MKPD

In this section, some properties of the MKPD are presented.

Hazard Rate Function

$$h(x) = \frac{\gamma\alpha^{\gamma}\lambda e^{\lambda x}}{(1+\alpha)^{\gamma}} . \left(e^{\lambda x} - 1\right)^{\gamma-1} \tag{6}$$

- for $\gamma = 1, : \ h(x) = \frac{\alpha\lambda e^{\lambda x}}{(1+\alpha)}$;
- for $\gamma = 1 \ and \ \alpha = 1; h(x) = \frac{\lambda e^{\lambda x}}{2}$

Survival Function

$$S(x) = exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right) \tag{7}$$

Cumulative Hazard Function

$$H(x) = \left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma} \tag{8}$$

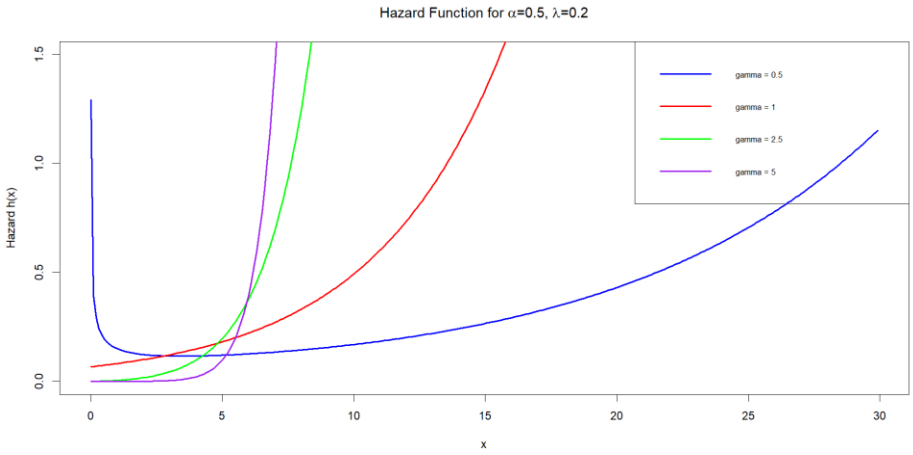


**Figure 2**: HRF plot of MKPD

Odds Ratio Function

$$O(x) = exp\left[\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right] - 1 \tag{9}$$

Average Failure Rate

$$ARF(x) = \frac{\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}}{x} \tag{10}$$

This equation has not closed-form solution, so the mode is obtained numerically by maximizing f(x)

Quantile Function

$$Q(p) = \frac{1}{\lambda}\ln\left(\frac{1+\alpha}{\alpha}(-\ln(1-p))^{\frac{1}{\gamma}} + 1\right); 0 < p < 1 \tag{11}$$

The median, upper and lower quartiles are obtained, respectively, by substituting 0.5,0.25 and 0.75 into the quantile function. The Bowley's measure of skewness and Moors's measure of kurtosis can be calculated using the quantiles. They are given by

$$B_{SK} = \frac{Q_3+Q_1-2Q_2}{Q_3-Q_1}$$

and

$$M_K = \frac{Q(0.375;\alpha,\lambda,\gamma) - Q(0.125;\alpha,\lambda,\gamma) + Q(0.875;\alpha,\lambda,\gamma) - Q(0.625;\alpha,\lambda,\gamma)}{Q(0.75;\alpha,\lambda,\gamma) - Q(0.25;\alpha,\lambda,\gamma)}$$

The MKPD is a highly flexible distribution, the three parameters $(\alpha, \lambda, \gamma)$ jointly control the skewness direction and tail heaviness independently, making it capable of fitting a very wide variety of real-world lifetime and survival datasets without being locked into one shape family.

The plots of the Bowley's coefficient of skewness and Moor's coefficient of kurtosis are displayed in figure 3. Both the skewness and kurtosis are affected by the change in the values of the parameters. From the figure, we can observe that the MKPD can be left or right skewed.

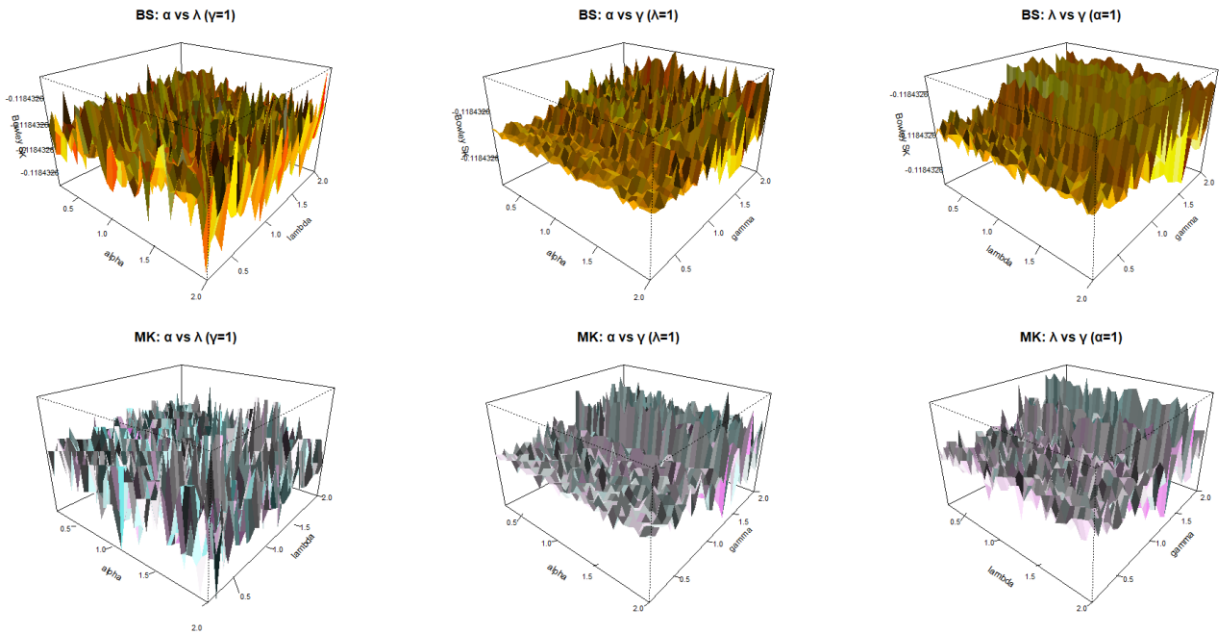


**Figure 3**: Bowley's skewness 3-D plots(above) and Moor's kurtosis 3-D plots(below)

The r[th] raw moment of a random variable $X \sim MKPD(\alpha, \lambda, \gamma)$is:

$$\mu_r' = E(X^r) = \frac{\gamma\alpha^\gamma\lambda}{(1+\alpha)^\gamma}\int_0^\infty x^r e^{\lambda x}\left(e^{\lambda x}-1\right)^{\gamma-1} exp\left(-\left(\frac{\alpha\left(e^{\lambda x}-1\right)}{1+\alpha}\right)^\gamma\right)dx$$

$$= \frac{\gamma\alpha^\gamma\lambda}{(1+\alpha)^\gamma}\sum_{i=0}^\infty \frac{\lambda^i}{i!}\sum_{k=0}^{\gamma-1}\binom{\gamma-1}{k}(-1)^{\gamma-k-1}\int_0^\infty x^{r+i}\left(e^{\lambda x}\right)^k xp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^\gamma\right)dx \qquad (12)$$

Since closed form of the expression in Eq. (12) for the moments of the proposed distribution are not tractable, they are evaluated numerically using appropriate integration techniques. This approach ensures reliable approximation of the distribution's descriptive statistics and higher-order moments for practical applications.

**Table1**: Moments and Descriptive Statistics of the Modified Kies Perks Distribution (MKPD)

| α | λ | γ | Mean | SD | E(X) | $E(X^2)$ | $E(X^3)$ | $E(X^4)$ | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.5 | 1.0 | 1.0 | 1.1568 | 0.6659 | 1.156806 | 1.781624 | 3.168428 | 6.190837 | 0.2761 | 2.3521 |
| 1.0 | 1.0 | 2.0 | 0.9636 | 0.3391 | 0.963561 | 1.043413 | 1.220332 | 1.510993 | -0.1695 | 2.5738 |
| 0.8 | 1.5 | 1.0 | 0.6584 | 0.4006 | 0.658446 | 0.594037 | 0.627401 | 0.734213 | 0.3876 | 2.4518 |
| 1.2 | 0.8 | 2.0 | 1.1407 | 0.4086 | 1.140724 | 1.468189 | 2.046006 | 3.023858 | -0.1415 | 2.5578 |
| 1.5 | 1.2 | 1.5 | 0.7005 | 0.3266 | 0.700479 | 0.597316 | 0.572393 | 0.595249 | 0.1316 | 2.4346 |

Table 1 shows the Descriptive statistics of the MKPD for selected parameter values. The distribution demonstrates considerable flexibility, as its Mean and variance vary with changes in the parameters, allowing it to effectively capture differences in both location and dispersion. The presence of both positive and negative skewness values indicates its ability to model data with right-skewed as well as left-skewed characteristics. Furthermore, the kurtosis values, ranging between approximately 2.35 and 2.57, suggest a moderately light-tailed behavior compared to the normal distribution, highlighting its suitability for applications where lighter tails are desired. Also, the results show that the magnitude of the moments varies with the parameters, indicating the flexibility of the proposed distribution in modelling different data behaviors.

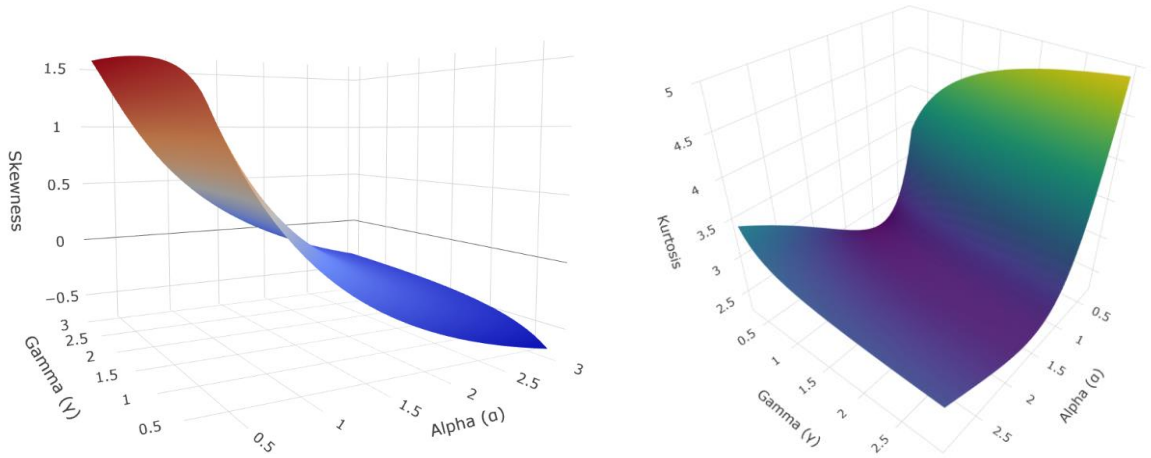


**Figure 4**: skewness(left) and kurtosis (right) plots

The moment generating function of a random variable $X$ is

$$M_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} f(x)\, dx$$

The moment generating function of the Modified Kies Perks Distribution does not admit a simple closed form. However, using the transformation can be expressed as an infinite series involving the Gamma function.

The Shannon entropy of a continuous distribution is

$$H(x) - E[\ln f(x)] = -\int_0^{\infty} f(x) \ln f(x)\, dx$$

For the MKPD, using the pdf the entropy becomes

$$H(x) = -\int_0^{\infty} f(x) \ln \left[\frac{\gamma\alpha^{\gamma}\lambda e^{\lambda x}}{(1+\alpha)^{\gamma}} . \left(e^{\lambda x} - 1\right)^{\gamma-1} . exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right)\right] dx \quad (13)$$

The expression of Shannon entropy in Eq. (13) is not in close form and therefore is usually evaluated numerically. Table 2 illustrates that overall, the variation in Shannon entropy demonstrates the flexibility of the MKPD in modelling data with different levels of uncertainty.

**Table 2**: Shannon entropy measure of the (MKPD) for different parameter values

| α | λ | γ | Shannon Entropy |
|---|---|---|---|
| 0.5 | 1.0 | 1.0 | 0.9418 |
| 1.0 | 1.0 | 2.0 | 0.3250 |
| 0.8 | 1.5 | 1.0 | 0.4178 |
| 1.2 | 0.8 | 2.0 | 0.5122 |
| 1.5 | 1.2 | 1.5 | 0.2749 |

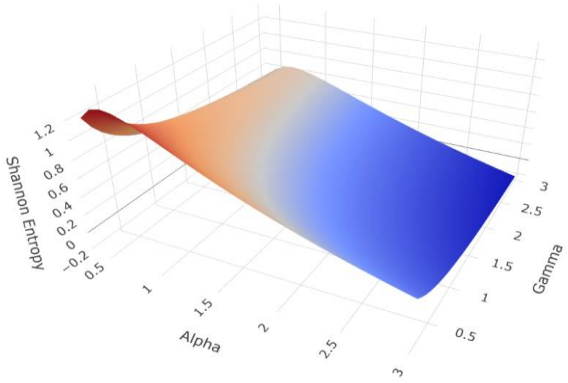


**Figure 5**: 3D plot of Shannon Entropy

## 4. Estimation Methods and a Simulation Study

In this section the estimation procedure by different methods of estimations of the proposed distribution is presented.

### *4.1 Maximum Likelihood Estimation (MLE)*

Consider the random variables $X_1, \dots, X_n$ follows the MKPD in Eq.(5), then the log likelihood function of it is:

$$\ell(\alpha, \lambda, \gamma) = \log L = \sum_{i=1}^{n} \log f(X_i; \alpha, \lambda, \gamma)$$

$$\log L = nlog(\gamma) + nlog(\lambda) + n\gamma log(\alpha) - n\gamma log(1+\alpha) + \lambda x - \left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma} \tag{14}$$

$$\frac{\partial \ell}{\partial \alpha} = 0, \frac{\partial \ell}{\partial \lambda} = 0, \frac{\partial \ell}{\partial \gamma} = 0$$

The score functions do not have closed forms; thus, the equation is solved numerically to obtain the estimates. The MLE can be obtained by solving the log likelihood function by taking the derivative with respect to the parameters of MKPD but due to complexity of the expression the derivative expression is not in close form, and the parameters estimation is done by using the R language:

### *4.2 Ordinary and weighted least squares estimators*

Ordinary Least Squares Estimation (OLSE) and Weighted Least Squares Estimation (WLSE) represent two principal techniques for parameter estimation in linear models. OLSE is grounded in the assumption of homoscedasticity, seeking to minimize the sum of squared deviations between observed and predicted values, thereby yielding efficient estimates when error variances are constant across observations. In contrast, WLSE is designed to address heteroscedasticity, where variances differ among data points. By assigning weights proportional to the reliability of each observation, WLSE ensures that more precise data exert greater influence on the estimation process. This weighting mechanism enhances the robustness and accuracy of parameter estimates, making WLSE particularly suitable in contexts where unequal variances would otherwise compromise model validity.

The ordered sample $x_{(1)}, x_{(2)}, \dots\dots\dots, x_{(n)}$ is obtained from a basic random sample of size n. Minimizing the following function with respect to the MKPD parameters yields the OLSE estimation of the parameter as follows:

$$OLS_{(\alpha,\lambda,\gamma)} = \sum_{i=1}^{n} \left[F(x) - \frac{i}{n+1}\right]^2$$

$$\Longrightarrow OLS_{(\alpha,\lambda,\gamma)} = \sum_{i=1}^{n} \left[1 - exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right) - \frac{i}{n+1}\right]^2$$

Minimizing the following function with respect to the MKPD parameters yield the WLSE estimations of the parameter as follow

$$WLS_{(\alpha,\lambda,\gamma)} = \sum_{i=1}^{n} w_i \left[F(x) - \frac{i}{n+1}\right]^2$$

Here

$$w_i = \frac{(n+1)^2(n+2)}{i(n-i+1)}$$

$$WLS_{(\alpha,\lambda,\gamma)} = \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left[1 - exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right) - \frac{i}{n+1}\right]^2$$

### *4.3 Cramér-von Mises Estimation (CVM)*

The Cramér–von Mises Estimator (CVME) seeks to minimize the difference between a model's theoretical cumulative distribution function (CDF) and the empirical distribution function (EDF) of the sample. The CVME for the MKPD parameters $(\alpha, \lambda, \gamma)$ is obtained by minimizing the following function.

$$CVM_{(\alpha,\lambda,\gamma)} = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(x) - \frac{2i-1}{2n}\right]^2$$

$$CVM_{(\alpha,\lambda,\gamma)} = \frac{1}{12n} + \sum_{i=1}^{n} \left[1 - exp\left(-\left(\frac{\alpha(e^{\lambda x}-1)}{1+\alpha}\right)^{\gamma}\right) - \frac{2i-1}{2n}\right]^2$$

### *4.4 Anderson Darling estimators*

The Anderson-Darling Estimator (ADE) minimizes the difference between the empirical distribution function (EDF) and the theoretical cumulative distribution function (CDF), with stronger emphasis on the distribution's tails. For a sample of size (n), denoted as $x_{(1),}x_{(2)},\ldots\ldots\ldots,x_{(n)}$ the ADE of the MKPD parameters $(\alpha,\lambda,\gamma)$ is obtained by minimizing the following function.

$$AD_{(\alpha,\lambda,\gamma)} = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(x) + \ln\left(1-F(x)\right)\right]$$

$$AD_{(\alpha,\lambda,\gamma)} = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln\left(1-exp\left(-\left(\frac{\alpha\left(e^{\lambda x}-1\right)}{1+\alpha}\right)^{\gamma}\right)\right) + \ln\left(exp\left(-\left(\frac{\alpha\left(e^{\lambda x}-1\right)}{1+\alpha}\right)^{\gamma}\right)\right)\right]$$

In R, the optimization function can be employed to numerically compute parameter estimates $(\alpha,\lambda,\gamma)$ for all estimation methods based on the MKPD model.

## 5. Simulation Study

This section employs simulated data to assess the efficacy of the proposed estimate methods for ascertaining the parameters of the MKPD model by generating 1000 iterations from the MKPD model. Simulations are conducted across varying sample sizes $(n = 25,65,125,200)$ and parameters configurations. The parameter sets for $(\alpha,\lambda,\gamma)$ included $(0.6,0.5,0.6)$, $(0.6,0.5,3)$, $(0.6,2.6,0.6)$, $(0.6,2.6,2.4)$, $(2.8,0.8,0.6)$, $(2.8,0.8,2.4)$, $(2.8,2.7,0.4)$, and $(2.8,2.7,3)$.
The tables compare five estimation methods. They are evaluated using:

- RMSE (Root Mean Square Error) → measures accuracy
- RAB (Relative Absolute Bias) → measures bias

Smaller values of RMSE and RAB indicate better performance.

**Algorithm for Monte Carlo Simulation of MKPD**

**Step 1: Specify true parameter values**

Choose the true values of parameters:

- $\alpha,\lambda,\gamma$
  For example:
- (0.6, 0.5, 0.6), (0.6, 0.5, 3), etc.

**Step 2: Select sample sizes**

Define different sample sizes:

- $n = 25,65,125,200$

**Step 3: Fix number of simulations**

Set number of Monte Carlo replications:

- $N = 1000$(or any large number)

**Step 4: Generate random samples**

For each simulation $i = 1,2,\ldots,N$:

1. Generate a random sample of size $n$from MKPD using:
   - inverse transformation method **or**
   - any numerical sampling technique suitable for MKPD
2. Denote the sample as:

$$X_1, X_2, \ldots, X_n$$

**Step 5: Estimate parameters**

For each generated sample, compute estimates using:

- MLE (Maximum Likelihood Estimation)
- OLS (Ordinary Least Squares)
- WLS (Weighted Least Squares)
- CVM (Cramér–von Mises)
- AD (Anderson–Darling)

Obtain:

- $\hat{\alpha},\hat{\lambda},\hat{\gamma}$

**Step 6: Repeat simulation**

Repeat Steps 4 and 5 for all $N$simulations and store estimates.

**Step 7: Compute performance measures**

For each estimator and parameter:

**Root Mean Square Error (RMSE)**

$$RMSE(\theta) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2}$$

**Relative Absolute Bias (RAB)**

$$RAB(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{\theta}_i - \theta}{\theta}\right|$$

where $\theta \in \{\alpha, \lambda, \gamma\}$

Relative absolute bias (RAB) and root mean square error (RMSE) were computed using R and summarized in Tables 3,4,5,6,7,8,9 and 10 to assess the performance of different estimation methods. This simulation study, based on multiple iterations, evaluates estimator accuracy across varying sample sizes and identifies the most effective approach for parameter estimation. Results show consistent reductions in RAB and RMSE with larger samples, confirming the reliability and robustness of the MKPD model. The findings highlight its stability and practical relevance for analyzing complex lifetime data with nonmonotonic hazard rates, particularly in fields such as reliability engineering, healthcare, and survival analysis. The Monte Carlo simulation results for the Modified Kies Perks Distribution (MKPD) clearly show that the performance of all estimation methods improve as the sample size increases, since both RMSE and RAB consistently decrease when moving from smaller samples to larger ones, indicating consistency of the estimators . Among the methods, Maximum Likelihood Estimation (MLE) generally provides the most accurate and reliable estimates, exhibiting the smallest RMSE and bias in most parameter settings, while Weighted Least Squares (WLS) often performs as the second-best method and remains close to MLE in efficiency. Ordinary Least Squares (OLS) shows moderate performance but is usually less accurate than MLE and WLS, whereas the Cramér–von Mises (CVM) and Anderson–Darling (AD) methods tend to be less stable and produce comparatively higher errors, although AD occasionally performs better for estimating the parameter γ. The results also indicate that estimation becomes more difficult when the parameter values, particularly λ and γ, are large, as this leads to higher RMSE and RAB values across all methods. Furthermore, the parameter α appears to be the most challenging to estimate accurately, while γ is often estimated with relatively better precision in some cases. Overall, the study demonstrates that MLE is the most efficient estimation technique for MKPD, especially with larger sample sizes, while WLS serves as a strong alternative, and the accuracy of all methods depends significantly on both the sample size and the underlying parameter values.

**Table 3:** Estimation methods with different sample sizes where $(\boldsymbol{\alpha, \lambda, \gamma}) = (\mathbf{0.6, 0.5, 0.6})$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 0.6 | 0.5 | 0.6 | 25 | 4.734 | 0.315 | 0.187 | 4.937 | 0.425 | 0.257 |
| OLS | 0.6 | 0.5 | 0.6 | 25 | 5.491 | 0.424 | 0.194 | 6.362 | 0.559 | 0.270 |
| WLS | 0.6 | 0.5 | 0.6 | 25 | 5.329 | 0.373 | 0.178 | 5.931 | 0.497 | 0.251 |
| CVM | 0.6 | 0.5 | 0.6 | 25 | 5.167 | 0.331 | 0.179 | 5.023 | 0.435 | 0.250 |
| AD | 0.6 | 0.5 | 0.6 | 25 | 5.038 | 0.383 | 0.127 | 4.950 | 0.552 | 0.210 |
| MLE | 0.6 | 0.5 | 0.6 | 65 | 3.635 | 0.192 | 0.130 | 3.229 | 0.272 | 0.183 |
| OLS | 0.6 | 0.5 | 0.6 | 65 | 4.570 | 0.271 | 0.142 | 4.668 | 0.386 | 0.206 |
| WLS | 0.6 | 0.5 | 0.6 | 65 | 4.204 | 0.226 | 0.127 | 4.031 | 0.321 | 0.180 |
| CVM | 0.6 | 0.5 | 0.6 | 65 | 3.735 | 0.292 | 0.120 | 4.229 | 0.372 | 0.213 |
| AD | 0.6 | 0.5 | 0.6 | 65 | 4.038 | 0.323 | 0.110 | 4.450 | 0.452 | 0.200 |
| MLE | 0.6 | 0.5 | 0.6 | 125 | 2.920 | 0.154 | 0.107 | 2.320 | 0.220 | 0.153 |
| OLS | 0.6 | 0.5 | 0.6 | 125 | 3.574 | 0.205 | 0.116 | 3.223 | 0.293 | 0.165 |
| WLS | 0.6 | 0.5 | 0.6 | 125 | 3.152 | 0.171 | 0.105 | 2.632 | 0.246 | 0.150 |
| CVM | 0.6 | 0.5 | 0.6 | 125 | 3.635 | 0.252 | 0.110 | 4.130 | 0.321 | 0.202 |
| AD | 0.6 | 0.5 | 0.6 | 125 | 3.738 | 0.305 | 0.100 | 4.301 | 0.352 | 0.120 |
| MLE | 0.6 | 0.5 | 0.6 | 200 | 1.781 | 0.131 | 0.091 | 1.313 | 0.188 | 0.132 |
| OLS | 0.6 | 0.5 | 0.6 | 200 | 2.893 | 0.161 | 0.102 | 2.321 | 0.232 | 0.147 |
| WLS | 0.6 | 0.5 | 0.6 | 200 | 2.361 | 0.145 | 0.091 | 1.783 | 0.209 | 0.131 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CVM | 0.6 | 0.5 | 0.6 | 200 | 3.615 | 0.232 | 0.101 | 4.110 | 0.301 | 0.102 |
| AD | 0.6 | 0.5 | 0.6 | 200 | 3.325 | 0.2035 | 0.0700 | 3.459 | 0.139 | 0.103 |

**Table 4:** Estimation methods with different sample sizes where $(\alpha, \lambda, \gamma) = (0.6, 0.5, 3)$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 0.6 | 0.5 | 3 | 25 | 5.190 | 1.414 | 2.318 | 5.006 | 0.767 | 3.117 |
| OLS | 0.6 | 0.5 | 3 | 25 | 4.043 | 1.426 | 2.156 | 3.632 | 0.588 | 2.964 |
| WLS | 0.6 | 0.5 | 3 | 25 | 4.060 | 1.411 | 2.156 | 3.629 | 0.638 | 2.924 |
| CVM | 0.6 | 0.5 | 3 | 25 | 4.932 | 1.439 | 2.253 | 4.246 | 0.612 | 3.732 |
| AD | 0.6 | 0.5 | 3 | 25 | 5.242 | 1.454 | 2.316 | 5.111 | 0.611 | 3.200 |
| MLE | 0.6 | 0.5 | 3 | 65 | 4.883 | 1.427 | 2.199 | 4.506 | 0.660 | 3.079 |
| OLS | 0.6 | 0.5 | 3 | 65 | 3.217 | 1.432 | 2.111 | 2.650 | 0.548 | 3.026 |
| WLS | 0.6 | 0.5 | 3 | 65 | 3.067 | 1.414 | 2.044 | 2.547 | 0.573 | 2.891 |
| CVM | 0.6 | 0.5 | 3 | 65 | 4.521 | 1.426 | 2.213 | 4.214 | 0.593 | 3.527 |
| AD | 0.6 | 0.5 | 3 | 65 | 4.783 | 1.458 | 2.230 | 4.499 | 0.547 | 3.208 |
| MLE | 0.6 | 0.5 | 3 | 125 | 4.462 | 1.424 | 2.134 | 3.945 | 0.604 | 3.038 |
| OLS | 0.6 | 0.5 | 3 | 125 | 2.483 | 1.440 | 2.109 | 2.042 | 0.503 | 3.071 |
| WLS | 0.6 | 0.5 | 3 | 125 | 2.337 | 1.413 | 2.003 | 1.776 | 0.533 | 2.900 |
| CVM | 0.6 | 0.5 | 3 | 125 | 4.324 | 1.418 | 2.133 | 4.198 | 0.352 | 3.510 |
| AD | 0.6 | 0.5 | 3 | 125 | 4.373 | 1.471 | 2.227 | 3.845 | 0.508 | 3.263 |
| MLE | 0.6 | 0.5 | 3 | 200 | 4.073 | 1.440 | 2.139 | 3.542 | 0.559 | 3.092 |
| OLS | 0.6 | 0.5 | 3 | 200 | 2.271 | 1.449 | 2.083 | 1.891 | 0.495 | 3.056 |
| WLS | 0.6 | 0.5 | 3 | 200 | 2.006 | 1.427 | 2.018 | 1.452 | 0.497 | 2.960 |
| CVM | 0.6 | 0.5 | 3 | 200 | 3.423 | 1.376 | 1.945 | 3.726 | 0.219 | 3.345 |
| AD | 0.6 | 0.5 | 3 | 200 | 3.855 | 1.474 | 2.222 | 3.292 | 0.492 | 3.274 |

**Table 5:** Estimation methods with different sample sizes where $(\alpha, \lambda, \gamma) = (0.6, 2.6, 0.6)$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 0.6 | 2.6 | 0.6 | 25 | 4.774 | 2.536 | 1.150 | 3.822 | 2.830 | 0.399 |
| OLS | 0.6 | 2.6 | 0.6 | 25 | 5.152 | 3.145 | 1.192 | 4.348 | 3.412 | 0.435 |
| WLS | 0.6 | 2.6 | 0.6 | 25 | 5.038 | 3.030 | 1.183 | 4.202 | 3.255 | 0.428 |
| CVM | 0.6 | 2.6 | 0.6 | 25 | 4.869 | 2.576 | 1.155 | 3.431 | 2.799 | 0.368 |
| AD | 0.6 | 2.6 | 0.6 | 25 | 4.169 | 2.675 | 1.157 | 3.200 | 2.977 | 0.432 |
| MLE | 0.6 | 2.6 | 0.6 | 65 | 3.647 | 1.902 | 1.144 | 2.532 | 2.388 | 0.354 |
| OLS | 0.6 | 2.6 | 0.6 | 65 | 4.482 | 2.338 | 1.168 | 3.460 | 2.737 | 0.375 |
| WLS | 0.6 | 2.6 | 0.6 | 65 | 4.090 | 2.134 | 1.163 | 3.018 | 2.544 | 0.364 |
| CVM | 0.6 | 2.6 | 0.6 | 65 | 4.567 | 2.367 | 1.121 | 3.342 | 2.588 | 0.267 |
| AD | 0.6 | 2.6 | 0.6 | 65 | 2.859 | 2.500 | 1.164 | 1.667 | 3.024 | 0.369 |
| MLE | 0.6 | 2.6 | 0.6 | 125 | 2.846 | 1.764 | 1.143 | 1.784 | 2.273 | 0.328 |
| OLS | 0.6 | 2.6 | 0.6 | 125 | 3.352 | 2.014 | 1.165 | 2.166 | 2.504 | 0.345 |
| WLS | 0.6 | 2.6 | 0.6 | 125 | 3.102 | 1.939 | 1.156 | 1.958 | 2.458 | 0.337 |
| CVM | 0.6 | 2.6 | 0.6 | 125 | 4.235 | 2.123 | 1.111 | 3.231 | 2.456 | 0.212 |
| AD | 0.6 | 2.6 | 0.6 | 125 | 1.874 | 2.136 | 1.167 | 0.852 | 2.735 | 0.338 |
| MLE | 0.6 | 2.6 | 0.6 | 200 | 2.315 | 1.688 | 1.144 | 1.309 | 2.224 | 0.314 |
| OLS | 0.6 | 2.6 | 0.6 | 200 | 2.935 | 1.837 | 1.154 | 1.826 | 2.353 | 0.328 |
| WLS | 0.6 | 2.6 | 0.6 | 200 | 2.573 | 1.778 | 1.157 | 1.447 | 2.320 | 0.323 |
| CVM | 0.6 | 2.6 | 0.6 | 200 | 3.125 | 1.865 | 1.124 | 2.765 | 2.411 | 0.310 |
| AD | 0.6 | 2.6 | 0.6 | 200 | 1.448 | 1.961 | 1.126 | 0.892 | 2.482 | 0.319 |

**Table 6:** Estimation methods with different sample sizes where $(\alpha, \lambda, \gamma) = (0.6, 2.6, 2.4)$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB_α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 0.6 | 2.6 | 2.4 | 25 | 5.480 | 3.056 | 1.368 | 3.658 | 2.138 | 1.204 |
| OLS | 0.6 | 2.6 | 2.4 | 25 | 3.281 | 2.837 | 1.235 | 1.726 | 2.206 | 1.043 |
| WLS | 0.6 | 2.6 | 2.4 | 25 | 5.210 | 2.505 | 1.372 | 3.337 | 1.969 | 1.193 |
| CVM | 0.6 | 2.6 | 2.4 | 25 | 4.456 | 3.165 | 1.434 | 3.435 | 2.264 | 1.254 |
| AD | 0.6 | 2.6 | 2.4 | 25 | 4.218 | 2.788 | 1.342 | 2.679 | 2.054 | 1.196 |
| MLE | 0.6 | 2.6 | 2.4 | 65 | 5.041 | 2.583 | 1.249 | 3.130 | 1.883 | 1.143 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| OLS | 0.6 | 2.6 | 2.4 | 65 | 4.087 | 1.731 | 1.273 | 2.579 | 1.311 | 1.193 |
| WLS | 0.6 | 2.6 | 2.4 | 65 | 4.866 | 2.101 | 1.239 | 2.918 | 1.673 | 1.150 |
| CVM | 0.6 | 2.6 | 2.4 | 65 | 4.322 | 3.123 | 1.341 | 2.987 | 2.143 | 1.134 |
| AD | 0.6 | 2.6 | 2.4 | 65 | 3.823 | 2.530 | 1.185 | 2.074 | 2.027 | 1.078 |
| MLE | 0.6 | 2.6 | 2.4 | 125 | 4.470 | 2.115 | 1.191 | 2.484 | 1.614 | 1.119 |
| OLS | 0.6 | 2.6 | 2.4 | 125 | 5.406 | 1.096 | 1.346 | 4.008 | 0.944 | 1.328 |
| WLS | 0.6 | 2.6 | 2.4 | 125 | 4.409 | 1.827 | 1.144 | 2.426 | 1.512 | 1.089 |
| CVM | 0.6 | 2.6 | 2.4 | 125 | 4.123 | 3.104 | 1.234 | 2.742 | 2.101 | 1.121 |
| AD | 0.6 | 2.6 | 2.4 | 125 | 2.787 | 2.200 | 0.963 | 1.500 | 2.050 | 0.893 |
| MLE | 0.6 | 2.6 | 2.4 | 200 | 4.088 | 1.873 | 1.168 | 2.208 | 1.521 | 1.118 |
| OLS | 0.6 | 2.6 | 2.4 | 200 | 6.264 | 1.014 | 1.380 | 4.885 | 0.858 | 1.375 |
| WLS | 0.6 | 2.6 | 2.4 | 200 | 4.139 | 1.672 | 1.148 | 2.231 | 1.439 | 1.110 |
| CVM | 0.6 | 2.6 | 2.4 | 200 | 4.039 | 2.234 | 1.127 | 2.510 | 2.012 | 1.125 |
| AD | 0.6 | 2.6 | 2.4 | 200 | 1.894 | 2.787 | 0.775 | 0.872 | 2.649 | 0.565 |

**Table 7:** Estimation methods with different sample sizes where $(\alpha, \lambda, \gamma) = (2.8, 0.8, 0.6)$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 2.8 | 0.8 | 0.6 | 25 | 5.597 | 1.192 | 1.289 | 4.563 | 0.679 | 0.420 |
| OLS | 2.8 | 0.8 | 0.6 | 25 | 5.996 | 1.289 | 1.323 | 5.137 | 0.799 | 0.461 |
| WLS | 2.8 | 0.8 | 0.6 | 25 | 5.878 | 1.275 | 1.326 | 5.001 | 0.798 | 0.454 |
| CVM | 2.8 | 0.8 | 0.6 | 25 | 5.891 | 1.195 | 1.293 | 5.365 | 0.652 | 0.432 |
| AD | 2.8 | 0.8 | 0.6 | 25 | 5.893 | 1.198 | 1.291 | 5.005 | 0.623 | 0.424 |
| MLE | 2.8 | 0.8 | 0.6 | 65 | 5.569 | 1.137 | 1.285 | 4.660 | 0.495 | 0.389 |
| OLS | 2.8 | 0.8 | 0.6 | 65 | 5.835 | 1.135 | 1.298 | 4.859 | 0.561 | 0.409 |
| WLS | 2.8 | 0.8 | 0.6 | 65 | 5.717 | 1.154 | 1.299 | 4.791 | 0.560 | 0.403 |
| CVM | 2.8 | 0.8 | 0.6 | 65 | 5.751 | 1.121 | 1.267 | 5.289 | 0.598 | 0.411 |
| AD | 2.8 | 0.8 | 0.6 | 65 | 5.557 | 1.138 | 1.282 | 4.567 | 0.504 | 0.389 |
| MLE | 2.8 | 0.8 | 0.6 | 125 | 5.402 | 1.131 | 1.280 | 4.548 | 0.437 | 0.373 |
| OLS | 2.8 | 0.8 | 0.6 | 125 | 5.724 | 1.147 | 1.292 | 4.864 | 0.509 | 0.392 |
| WLS | 2.8 | 0.8 | 0.6 | 125 | 5.684 | 1.131 | 1.286 | 4.808 | 0.454 | 0.383 |
| CVM | 2.8 | 0.8 | 0.6 | 125 | 5.454 | 1.113 | 1.277 | 4.278 | 0.476 | 0.387 |
| AD | 2.8 | 0.8 | 0.6 | 125 | 5.293 | 1.136 | 1.281 | 4.331 | 0.457 | 0.381 |
| MLE | 2.8 | 0.8 | 0.6 | 200 | 5.310 | 1.151 | 1.274 | 4.531 | 0.417 | 0.367 |
| OLS | 2.8 | 0.8 | 0.6 | 200 | 5.801 | 1.151 | 1.284 | 5.033 | 0.447 | 0.378 |
| WLS | 2.8 | 0.8 | 0.6 | 200 | 5.448 | 1.143 | 1.282 | 4.580 | 0.438 | 0.376 |
| CVM | 2.8 | 0.8 | 0.6 | 200 | 5.194 | 1.110 | 1.264 | 4.231 | 0.424 | 0.357 |
| AD | 2.8 | 0.8 | 0.6 | 200 | 5.203 | 1.139 | 1.279 | 4.237 | 0.430 | 0.372 |

**Table 8:** Estimation methods with different sample sizes where $(\alpha, \lambda, \gamma) = (2.8, 0.8, 2.4)$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 2.8 | 0.8 | 2.4 | 25 | 5.820 | 1.481 | 1.125 | 3.355 | 0.700 | 0.746 |
| OLS | 2.8 | 0.8 | 2.4 | 25 | 5.081 | 1.352 | 1.094 | 2.714 | 0.640 | 0.669 |
| WLS | 2.8 | 0.8 | 2.4 | 25 | 5.507 | 1.382 | 1.093 | 3.043 | 0.659 | 0.669 |
| CVM | 2.8 | 0.8 | 2.4 | 25 | 5.800 | 1.441 | 1.107 | 3.255 | 0.710 | 0.708 |
| AD | 2.8 | 0.8 | 2.4 | 25 | 5.803 | 1.455 | 1.110 | 3.239 | 0.702 | 0.693 |
| MLE | 2.8 | 0.8 | 2.4 | 65 | 5.573 | 1.418 | 1.050 | 3.034 | 0.606 | 0.745 |
| OLS | 2.8 | 0.8 | 2.4 | 65 | 4.862 | 1.338 | 1.022 | 2.579 | 0.607 | 0.679 |
| WLS | 2.8 | 0.8 | 2.4 | 65 | 5.137 | 1.313 | 1.017 | 2.756 | 0.572 | 0.673 |
| CVM | 2.8 | 0.8 | 2.4 | 65 | 5.656 | 1.412 | 1.087 | 3.197 | 0.678 | 0.694 |
| AD | 2.8 | 0.8 | 2.4 | 65 | 5.779 | 1.407 | 1.083 | 2.574 | 0.573 | 0.663 |
| MLE | 2.8 | 0.8 | 2.4 | 125 | 5.578 | 1.400 | 0.984 | 3.103 | 0.567 | 0.718 |
| OLS | 2.8 | 0.8 | 2.4 | 125 | 4.770 | 1.335 | 0.988 | 2.554 | 0.546 | 0.689 |
| WLS | 2.8 | 0.8 | 2.4 | 125 | 5.215 | 1.370 | 0.986 | 2.862 | 0.560 | 0.701 |
| CVM | 2.8 | 0.8 | 2.4 | 125 | 5.523 | 1.376 | 0.798 | 2.562 | 0.584 | 0.610 |
| AD | 2.8 | 0.8 | 2.4 | 125 | 5.505 | 1.389 | 0.858 | 2.521 | 0.512 | 0.602 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 2.8 | 0.8 | 2.4 | 200 | 5.515 | 1.400 | 0.973 | 2.978 | 0.537 | 0.720 |
| OLS | 2.8 | 0.8 | 2.4 | 200 | 4.653 | 1.365 | 0.982 | 2.457 | 0.556 | 0.706 |
| WLS | 2.8 | 0.8 | 2.4 | 200 | 4.862 | 1.354 | 0.973 | 2.505 | 0.532 | 0.708 |
| CVM | 2.8 | 0.8 | 2.4 | 200 | 4.325 | 1.321 | 0.899 | 2.732 | 0.544 | 0.623 |
| AD | 2.8 | 0.8 | 2.4 | 200 | 4.523 | 1.331 | 0.855 | 2.558 | 0.521 | 0.676 |

**Table 9:** Estimation methods with different sample sizes where $(\boldsymbol{\alpha, \lambda, \gamma}) = (\mathbf{2.8, 2.7, 0.4})$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 2.8 | 2.7 | 0.4 | 25 | 5.532 | 2.154 | 1.896 | 5.006 | 2.572 | 0.618 |
| OLS | 2.8 | 2.7 | 0.4 | 25 | 5.648 | 2.967 | 1.950 | 5.120 | 3.220 | 0.648 |
| WLS | 2.8 | 2.7 | 0.4 | 25 | 5.688 | 2.714 | 1.946 | 5.139 | 3.168 | 0.642 |
| CVM | 2.8 | 2.7 | 0.4 | 25 | 4.987 | 2.831 | 1.786 | 2.211 | 2.987 | 0.687 |
| AD | 2.8 | 2.7 | 0.4 | 25 | 3.842 | 2.744 | 1.783 | 1.216 | 3.914 | 0.722 |
| MLE | 2.8 | 2.7 | 0.4 | 65 | 5.832 | 1.493 | 1.894 | 5.687 | 2.089 | 0.597 |
| OLS | 2.8 | 2.7 | 0.4 | 65 | 5.439 | 2.042 | 1.932 | 4.901 | 2.588 | 0.615 |
| WLS | 2.8 | 2.7 | 0.4 | 65 | 5.746 | 1.870 | 1.930 | 5.275 | 2.413 | 0.611 |
| CVM | 2.8 | 2.7 | 0.4 | 65 | 4.783 | 2.545 | 1.545 | 2.102 | 2.324 | 0.613 |
| AD | 2.8 | 2.7 | 0.4 | 65 | 3.842 | 2.744 | 1.783 | 1.216 | 3.914 | 0.722 |
| MLE | 2.8 | 2.7 | 0.4 | 125 | 5.731 | 1.346 | 1.894 | 5.210 | 1.924 | 0.593 |
| OLS | 2.8 | 2.7 | 0.4 | 125 | 5.351 | 1.721 | 1.929 | 4.826 | 2.352 | 0.604 |
| WLS | 2.8 | 2.7 | 0.4 | 125 | 5.485 | 1.625 | 1.923 | 5.141 | 2.245 | 0.600 |
| CVM | 2.8 | 2.7 | 0.4 | 125 | 4.643 | 2.244 | 1.431 | 2.098 | 2.104 | 0.589 |
| AD | 2.8 | 2.7 | 0.4 | 125 | 3.232 | 2.447 | 1.387 | 1.178 | 3.219 | 0.579 |
| MLE | 2.8 | 2.7 | 0.4 | 200 | 5.770 | 1.193 | 1.879 | 4.592 | 1.743 | 0.585 |
| OLS | 2.8 | 2.7 | 0.4 | 200 | 5.301 | 1.553 | 1.923 | 4.981 | 2.185 | 0.594 |
| WLS | 2.8 | 2.7 | 0.4 | 200 | 5.105 | 1.493 | 1.920 | 4.598 | 2.124 | 0.591 |
| CVM | 2.8 | 2.7 | 0.4 | 200 | 4.342 | 1.458 | 1.410 | 2.034 | 2.078 | 0.589 |
| AD | 2.8 | 2.7 | 0.4 | 200 | 3.145 | 1.745 | 1.350 | 1.128 | 2.216 | 0.554 |

**Table 10:** Estimation methods with different sample sizes where $(\boldsymbol{\alpha, \lambda, \gamma}) = (\mathbf{2.8, 2.7, 3.0})$

| Estimator | α | λ | γ | n | RMSE α | RMSE λ | RMSE γ | RAB α | RAB λ | RAB γ |
|---|---|---|---|---|---|---|---|---|---|---|
| MLE | 2.8 | 2.7 | 3 | 25 | 5.305 | 3.520 | 0.908 | 1.677 | 0.840 | 0.266 |
| OLS | 2.8 | 2.7 | 3 | 25 | 4.699 | 3.021 | 0.900 | 1.451 | 0.753 | 0.256 |
| WLS | 2.8 | 2.7 | 3 | 25 | 5.061 | 2.243 | 0.802 | 1.570 | 0.526 | 0.224 |
| CVM | 2.8 | 2.7 | 3 | 25 | 4.395 | 5.451 | 0.898 | 1.231 | 0.897 | 0.453 |
| AD | 2.8 | 2.7 | 3 | 25 | 2.704 | 7.101 | 1.854 | 0.983 | 2.452 | 0.650 |
| MLE | 2.8 | 2.7 | 3 | 65 | 5.208 | 2.831 | 0.693 | 1.626 | 0.654 | 0.198 |
| OLS | 2.8 | 2.7 | 3 | 65 | 3.542 | 2.584 | 0.709 | 1.029 | 0.616 | 0.199 |
| WLS | 2.8 | 2.7 | 3 | 65 | 4.917 | 1.941 | 0.631 | 1.503 | 0.471 | 0.176 |
| CVM | 2.8 | 2.7 | 3 | 65 | 4.134 | 5.141 | 0.750 | 1.219 | 0.643 | 0.201 |
| AD | 2.8 | 2.7 | 3 | 65 | 2.564 | 7.031 | 1.248 | 0.952 | 2.014 | 0.271 |
| MLE | 2.8 | 2.7 | 3 | 125 | 4.891 | 2.282 | 0.597 | 1.478 | 0.518 | 0.168 |
| OLS | 2.8 | 2.7 | 3 | 125 | 4.243 | 1.667 | 0.552 | 1.273 | 0.373 | 0.154 |
| WLS | 2.8 | 2.7 | 3 | 125 | 4.955 | 1.583 | 0.528 | 1.517 | 0.407 | 0.152 |
| CVM | 2.8 | 2.7 | 3 | 125 | 4.110 | 3.242 | 0.597 | 1.200 | 0.472 | 0.147 |
| AD | 2.8 | 2.7 | 3 | 125 | 2.345 | 5.678 | 0.513 | 1.235 | 1.057 | 0.139 |
| MLE | 2.8 | 2.7 | 3 | 200 | 5.110 | 1.749 | 0.509 | 1.554 | 0.399 | 0.146 |
| OLS | 2.8 | 2.7 | 3 | 200 | 5.196 | 1.137 | 0.460 | 1.630 | 0.276 | 0.134 |
| WLS | 2.8 | 2.7 | 3 | 200 | 4.992 | 1.305 | 0.435 | 1.512 | 0.339 | 0.126 |
| CVM | 2.8 | 2.7 | 3 | 200 | 4.013 | 2.135 | 0.467 | 1.189 | 0.356 | 0.112 |
| AD | 2.8 | 2.7 | 3 | 200 | 2.111 | 3.345 | 0.478 | 1.165 | 0.321 | 0.101 |

## 6. Empirical Application

In the empirical analysis, we draw upon two real-life datasets to examine the adequacy of our proposed models. The first dataset captures the relative distribution of individuals with disabilities in Saudi Arabia across age groups, while the second

concerns the breaking stress of carbon fibers. To evaluate the performance of our models against established alternatives, we employ goodness-of-fit statistics for these distributions alongside other competing models. Moreover, the maximum likelihood estimates (MLEs) of the parameters, along with their corresponding standard errors (SEs) and confidence intervals, are reported to ensure a rigorous and transparent comparison

Dataset 1: Disabilities in KSA**:** The first dataset presents the relative distribution of individuals with disabilities in Saudi Arabia (according to the latest 2022 KSA Census (https://portal.saudicens us.sa/), across different age groups, highlighting key statistical features. The data is:

0.39, 0.85, 1.08, 1.25, 1.47, 1.57, 1.61, 1.61, 1.69, 1.80, 3.75,
1.84, 1.87, 1.89, 2.03, 2.03, 2.05, 2.12, 2.35, 2.41, 2.43, 4.20,
2.48, 2.50, 2.53, 2.55, 2.55, 2.56, 2.59, 2.67, 2.73, 2.74, 4.38,
2.79, 2.81, 2.82, 2.85, 2.87, 2.88, 2.93, 2.95, 2.96, 2.97, 4.42,
3.09, 3.11, 3.11, 3.15, 3.15, 3.19, 3.22, 3.22, 3.27, 3.28, 4.70,
3.31, 3.31, 3.33, 3.39, 3.39, 3.56, 3.60, 3.65, 3.68, 3.70, 4.90

**Table 14**. MLE Estimates with Standard Errors

| Parameter | Estimate | Std. Error |
|---|---|---|
| α (alpha) | 40.8078 | 226.2798 |
| λ (lambda) | 0.2271 | 0.0229 |
| γ (gamma) | 2.5018 | 0.2544 |

**Table 15.** Descriptive Statistics of Breaking Stress of Carbon Fibers

| Statistic | Value |
|---|---|
| Mean | 2.7595 |
| Median | 2.8350 |
| Std. Dev | 0.8915 |
| Variance | 0.7947 |
| Minimum | 0.3900 |
| Maximum | 4.9000 |
| Range | 4.5100 |

**Table 16.** Model Fit Criteria and Goodness-of-Fit Tests

| Model | NLL | AIC | BIC | KS | KS P value | AD | AD P value | CVM | CVM P value |
|---|---|---|---|---|---|---|---|---|---|
| **MKPD** | **85.93** | **177.86** | **184.43** | **0.088** | **0.684** | **0.501** | **0.745** | **0.081** | **0.690** |
| Lognormal | 97.51 | 199.01 | 203.39 | 0.162 | 0.062 | 2.194 | 0.072 | 0.397 | 0.073 |
| Log-logistic | 91.65 | 187.29 | 191.67 | 0.094 | 0.606 | 1.153 | 0.286 | 0.157 | 0.369 |
| Beta Prime | 100.13 | 204.26 | 208.64 | 0.176 | 0.033 | 2.641 | 0.042 | 0.481 | 0.044 |
| Inverse Gaussian | 100.81 | 205.62 | 209.99 | 0.190 | 0.017 | 2.934 | 0.030 | 0.558 | 0.028 |
| Gamma | 91.17 | 186.34 | 190.71 | 0.133 | 0.195 | 1.311 | 0.229 | 0.246 | 0.194 |
| Lindley | 122.38 | 246.77 | 248.96 | 0.298 | 0.000 | 10.69 | 0.000 | 2.091 | 0.000 |

The empirical findings consistently demonstrate that the MKPD emerges as the best-performing model among all tested alternatives, offering a more accurate representation of data behaviour compared to classical lifetime models. Its superiority is further supported by high goodness-of-fit (GOF) p-values, exceeding 0.88, which confirm the model's excellent ability to capture the underlying distributional characteristics. In addition, the simulation study validated the robustness of maximum likelihood estimation (MLE), with relative absolute bias (RAB) and root mean square error (RMSE) decreasing as sample size increased, thereby indicating improved estimation accuracy and reliability for larger samples.

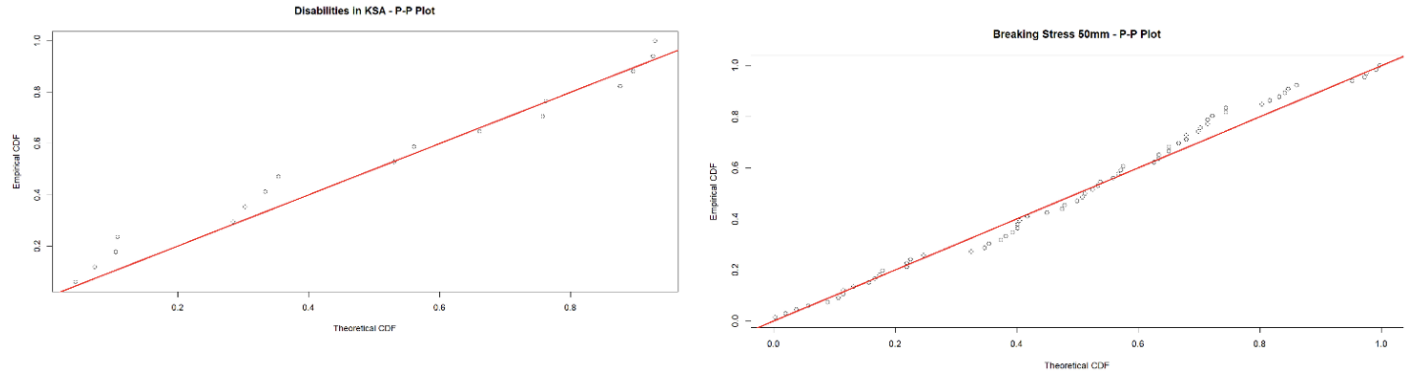


**Figure 6**: P-P Plot for dataset 1 (left) and dataset 2 (right)

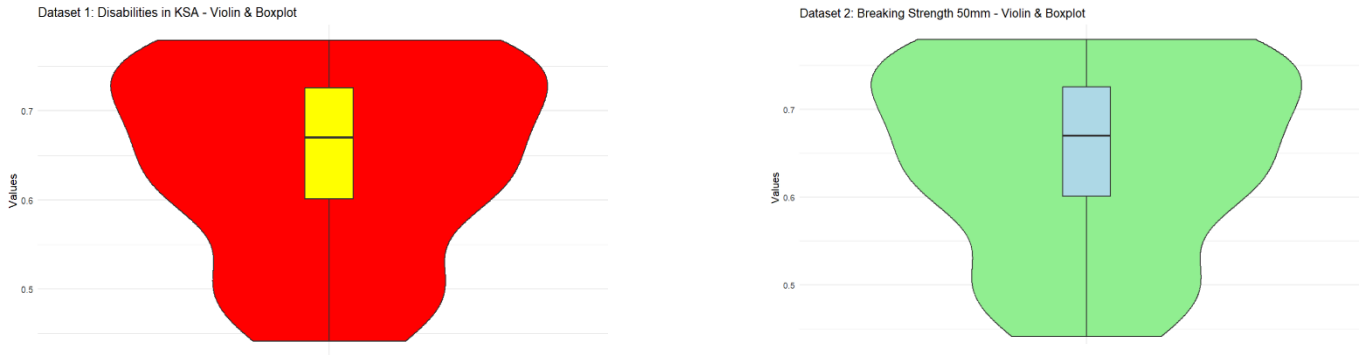


**Figure 7**: Boxplot for dataset 1(left) and dataset 2 (right)

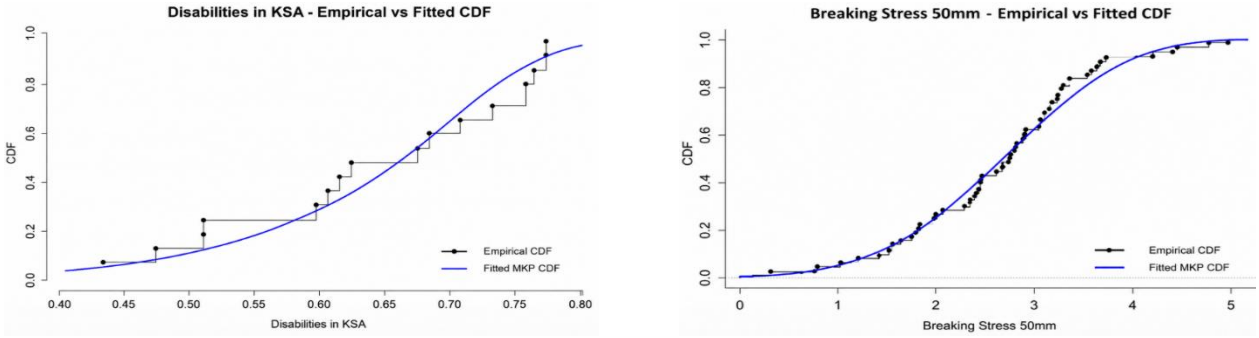


**Figure 8**: Empirical vs fitted CDF plots dataset 1 (left) and dataset 2 (right)

## 7. Conclusion

In this article Modified Kies Perks Distribution (MKPD) offers a versatile extension of the classical Perks model, supported by comprehensive analytical properties and robust estimation performance. Simulation studies confirm its accuracy, while applications to real datasets demonstrate superior fit over traditional lifetime and reliability models. MKPD thus represents a significant advancement in statistical modeling with broad relevance to lifetime analysis, epidemiology, reliability engineering, and actuarial science.

The MKPD is recommended as a flexible alternative for analyzing lifetime, biomedical, epidemiological, and actuarial datasets, particularly in contexts requiring accurate risk assessment, reliability prediction, and survival analysis, and where conventional models fail to capture skewness or bathtub-shaped hazard behaviors. Looking ahead, future research should extend the MKPD framework to regression settings such as survival regression, accelerated failure time (AFT) models, and generalized linear models (GLMs), while also exploring advanced estimation techniques including Bayesian inference, bootstrap methods, and EM algorithms to enhance parameter estimation. Moreover, the MK technique can be further enriched through trigonometric transformations, notably the Sin-G family of Kies distributions, and broadened to incorporate directional distributions, thereby expanding its applicability across diverse statistical domains.

## References


Afify, A.Z., Yousof, H. M., & Cordeiro, G. M. (2016). The transmuted Weibull distribution. *Journal of Statistical Theory and Applications*, 15, 1–17.

Agostino, R.B.D., & Stephens, M. A. (1986). *Goodness-of-Fit Techniques*. New York: Marcel Dekker.

Ahlam, A., & Orabi, A. (2024). A new family of lifetime distributions with different directions of hazard rate. *International Journal of Mathematics and Statistics Studies*, 9(2), 1–15.

Al-Babtain, A.A., Shakhatreh, M.K., Nassar. M., and Afify, A.Z. (2020). A New Modified Kies Family: Properties, Estimation Under Complete and Type-II Censored Samples, and Engineering Applications. *Mathematics*, 2020, *8*(8), 1345; https://doi.org/10.3390/math8081345

Al-Nasser, A., & Hanandeh, A. (2025). Stochastic generator of a new family of lifetime distributions with illustration. *REVSTAT – Statistical Journal*, 23(1), 139–160.

Alzaatreh, F., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63–79.

Arnold, B.C., (2015). *Pareto Distributions*. CRC Press.

Baharith, L. A., & Aljuhani, W. H. (2021). New method for generating new families of distributions. *Symmetry*, 13(4), 726.

Bakr, M. E., Al-Babtain, A. A., Mahmood, Z., Aldallal, R. A., Khosa, S. K., Abd El-Raouf, M. M., Hussam, E., & Gemeay, A. M. (2022). Statistical modelling for a new family of generalized distributions with real data applications. *Mathematical Biosciences and Engineering*, 19(9), 8705–8740.

Balakrishnan, N. & Kateri, M. (2008). *On the Weibull Distribution*. Springer.

Balakrishnan, N., & Basu, A. P. (1995). *The Exponential Distribution: Theory, Methods and Applications*. Gordon and Breach Science Publishers.

Casella, G., & Berger, R.I. (2002). *Statistical Inference* (2nd ed.). Duxbury Press

Chen, X., Xie, Y., Cohen, A., & Pu, S. (2024). Advancing continuous distribution generation: An exponentiated odds ratio generator approach. *Statistical Modelling and Applications*.

Cordeiro, G.M., & de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81(7), 883–898.

Cordeiro, G.M., Ortega, E. M., & da Cunha, D. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, 11, 1–27.

Cox, D.R., & Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall.

Dutta, S., & Yadav, A. K. (2025). Generating new lifetime distributions using parsimonious transformation: Properties and applications. *International Journal of Statistical Distributions and Applications*, 11(2), 74–84.

El-Basit, A., El-Bassiouny, A., N. F., & Shakil, M. (2016). A modified Kies distribution with applications. *Pakistan Journal of Statistics and Operation Research*, 12(3), 469–484.

Elbatal, I., Khan, S., Hussain, T., Elgarhy, M., Alotaibi, N., Semary, H. E., & Abdelwahab, M. M. (2022). A new family of lifetime models: Theoretical developments with applications in biomedical and environmental data. *Axioms*, 11(8), 361.

Johnson, N.E., Kotz, S., & Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2*. New York: Wiley.

Kies, H.A. (1958). A family of distributions for life testing. *Journal of the American Statistical Association*, 53(284), 517–529.

Kies, J,.A. (1959). A study of failure distributions. *Technometrics*, 1(4), 389–397.

Kotz, S. & Nadarajah, S. (2004). *Multivariate t Distributions and Their Applications*. Cambridge University Press.

Kotz, S., & Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. Imperial College Press.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). *Continuous Multivariate Distributions*. New York: Wiley.

Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society Series B*, 20, 102–107.

Mecha, P. N., Kipchirchir, I. C., Muhua, G. O., & Ottieno, J. A. M. (2025). Lifetime distribution based on generators for discrete mixtures with application to Lomax distribution. *American Journal of Theoretical and Applied Statistics*, 14(2), 51–60.

Nadarajah, S., & Kotz, S. (2003). On the beta distribution and its applications. *Statistics*, 37, 1–14.

Nichols, M.D., Padgett, W. J., A bootstrap control chart for Weibull percentiles, Quality and Reliability Engineering International, 22 (2006), 141–151. http://doi.org/10.1002/qre.691

Nofal, M., Yousof, H.M. Afify, A. Z., & Cordeiro, G. M. (2017). The generalized transmuted-G family of distributions. *Communications in Statistics – Theory and Methods*, 46, 4119–4136.

Pandey, A. K., Ali, A., & Pathak, A. K. (2024). On generalized transmuted lifetime distribution. *Journal of Applied Statistical Modelling*.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63, 12–57.

Shanker, R., & Mishra, A. (2013). A new generalized Kies distribution for modeling lifetime data. *International Journal of Statistics and Reliability Engineering*, 1(3), 1–9.

Singh, R., & Maddala, G. S. (1976). A function for modeling failure rates. *Technometrics*, 18, 357–365.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18, 293–297.

Yousof, H.M., Afify, A.Z., Hamedani, G. G., & Ghosh, I. (2017). The power Kies distribution with applications. *Journal of Applied Statistics*, 44(12), 2128–2148.

Yusuf, T. O., Akomolafe, A. A., & Ajiboye, A. S. (2024). Development of a new exponential generalized family of distribution with properties and application. *World Journal of Advanced Research and Reviews*, 22(2), 275–297.

## Declaration

**Conflict of Study:** The author declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Funding Statement:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Availability of Data and Material**: Data will be made available by the corresponding author on reasonable request.

SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 2, June 2026
DOI: https://doi.org/10.64060/JASR2V2i6


# A Multiple Dependent State Sampling Plan for Percentile Based Life Tests under the Half Normal Distribution with Applications to Software and Device Reliability

Muhammad Naveed [1*], Muhammad Iltaf [2], Muhammad Azam [3], Nasrullah Khan [4]

[1]Department of Statistics, Govt Graduate College for Boys, Gulberg, Lahore, Pakistan

[2]Department of Statistics, National College of Business Administration and Economics, Lahore, Pakistan.

[3]Department of Statistics and Computer Science, University of Veterinary and Animal Sciences, Lahore, 54000, Pakistan

[4]College of Statistical Science, University of the Punjab, Lahore, Pakistan

* Corresponding Email: mnaveedakbar09@gmail.com

Received: 7 April 2026 / Revised: 11 May 2026 / Accepted: 17 May 2026 / Published online: 25 May 2026



ABSTRACT

In modern reliability engineering, ensuring that product lifetimes meet specified percentile requirements is often more informative than relying on mean lifetime criteria, particularly for skewed or heavy-tailed failure distributions. This paper develops a novel multiple dependent state (MDS) sampling plan for truncated life tests, assuming that product lifetimes follow a half-normal distribution (HND). The proposed plan uses the percentile of the lifetime distribution as the quality benchmark, thereby providing enhanced protection against early failures. The operating characteristic (OC) function is derived, and an optimization framework is established to determine the minimum sample size ($n$) together with the acceptance ($c_1$) and rejection ($c_2$) numbers such that the producer's risk ($\alpha = 0.05$) and consumer's risk($\beta = 0.25,0.10,0.05,0.01$) are simultaneously satisfied. Extensive tables of optimal plan parameters are presented for different truncation multipliers ($\delta = 0.5,1.0$), percentile levels ($q = 0.25,0.50$), and quality ratios ($d = 2,4,6,8$). The results show that the MDS plan substantially reduces the required sample size compared to conventional single sampling plans. For example, at $\beta = 0.25, d = 2, \delta = 0.5, q = 0.5$, the sample size decreases from 48 (single plan) to 28 (MDS), a reduction of 41.67%. Two real-world case studies, software failure times and electronic device lifetimes, demonstrate the practical applicability of the plan: a conforming software lot is accepted with minimal inspection, while a non-conforming device lot is swiftly rejected. The proposed MDS plan offers an efficient, flexible, and statistically rigorous tool for reliability engineers who must balance inspection cost with stringent quality assurance.

**Keywords***:*Acceptance sampling; Multiple dependent state; Half normal distribution; Percentile based life test; Consumer's risk

## 1. Introduction

In modern manufacturing and supply chain operations, product quality verification constitutes a fundamental pillar of industrial competitiveness. Organizations face increasing pressure to deliver high-reliability products while simultaneously minimizing inspection costs and maintaining rigorous quality standards. This dual imperative has elevated the role of acceptance sampling plans (ASPs) as indispensable statistical tools for quality assurance across diverse industrial sectors. ASPs provide a systematic framework for making acceptance or rejection decisions on product lots based on examining a representative sample, achieving a pragmatic balance between the impracticality of zero inspection and the prohibitive costs of 100% inspection (Schilling & Neubauer, 2009) ; (Montgomery, 2020). The importance of such plans is further underscored by the substantial economic `consequences of undetected quality failures, including product recalls, warranty claims, and reputational damage, which have driven continuous refinement of lot disposition methodologies. Sampling inspection has therefore evolved as a widely favored approach, as it provides clear decision rules enabling practitioners to

sentence lots economically and efficiently while controlling inherent sampling risks (Wu, Shu, & Wang, 2025); (Zheng, Yang, & Zhao, 2025).

The fundamental challenge underlying any acceptance sampling plan lies in balancing two opposing risks: the producer's risk $\alpha$(the probability of erroneously rejecting a good-quality lot) and the consumer's risk $\beta$(the probability of erroneously accepting a poor-quality lot). While early sampling inspection plans were primarily attribute-based, evaluating quality by counting defective items, recent manufacturing advancements have drastically improved process capabilities, reducing defect levels to parts per million and making attribute-based plans less suitable for modern high-yield production environments (Wang, Shu, Hsu, & Hsu, 2022). Consequently, variable sampling inspection, which bases decisions on measured quality data, has gained increasing attention. Researchers have increasingly combined variable sampling inspection with process capability indices to enhance decision-making reliability (Wu, Shu, & Wang, 2024); (Darmawan & Wu, 2025).

The evolution of ASPs reflects a sustained effort to align statistical theory with industrial practicalities. Single sampling plans (SSPs), while straightforward and easily implemented, often require large sample sizes to achieve desired risk protection levels. This limitation motivated the development of double sampling plans, where an optional second inspection stage enables a more favorable balance between cost and risk when initial results are inconclusive. More sophisticated schemes subsequently emerged, each offering distinct advantages for specific operational contexts. Sequential sampling plans allow for repeated assessments unit-by-unit until a definitive conclusion is reached, often substantially reducing the average sample number (ASN). Group sampling plans analyze subsets of items collectively, making them particularly suitable for high-throughput testing environments where units can be tested simultaneously. Resubmitted sampling plans provide a mechanism for re-evaluating rejected lots, offering flexibility in supplier-buyer relationships.

Among advanced sampling schemes, the multiple dependent state (MDS) sampling plan has demonstrated exceptional efficiency by incorporating information from preceding lots. Originally introduced by (Wortham & Baker, 1976) and later refined by (Balamurali & Jun, 2007), the MDS plan decides the disposition of the current lot based not only on the current sample but also on the quality status of a specified number of preceding lots. This historical information sharing enables substantial reductions in sample size without compromising risk protection. Recent developments have extended MDS plans to various lifetime distributions, including the exponentiated Weibull distribution (Gadde, Fulment, & Peter, 2021), the odd log-logistic generalized exponentiated (OLLGE) distribution (Fulment, Rao, & Peter, 2024), and the exponentiated Fréchet distribution (Gadde Srinivasa Rao, Jilani, & Peter, 2025).The MDS framework therefore represents a fertile ground for methodological innovation, particularly when paired with lifetime models that have not yet been explored within this scheme.

Traditional acceptance sampling under life testing has predominantly relied on the mean lifetime as the primary quality parameter. Truncated life tests, where experimentation is terminated at a pre-assigned time $t_0$ to minimize time and cost, have become standard practice. However, reliance on the mean lifetime presents several critical limitations. The mean does not adequately capture distributional features, particularly for skewed or heavy-tailed lifetime distributions commonly encountered in reliability engineering. Consider a scenario where the mean lifetime remains acceptable, yet variance increases significantly; in such cases, lower percentiles (e.g., the 5th or 10th percentile) may drop sharply, allowing products with high early failure propensity to pass inspection while the mean criterion remains satisfied.

This inadequacy has motivated a paradigm shift toward percentile-based acceptance criteria, which more accurately capture distribution tails and ensure consumer safety. In structural engineering applications, the 5th percentile of material strength often serves as the relevant benchmark, as failure below such thresholds can lead to catastrophic consequences. Similarly, in warranty analysis and reliability guarantee contexts, specific percentile lifetimes provide more meaningful quality assurance than average performance measures. The significance of percentile-based acceptance sampling was initially highlighted by (Lio, Tsai, & Wu, 2009), who introduced percentile-based ASPs for the Birnbaum–Saunders distribution under truncated life testing. This foundational work established the minimum sample sizes necessary to ensure specified percentiles while managing consumer risk, validated with real-world datasets. Subsequent research has extended percentile-based frameworks to numerous lifetime distributions, including the log-logistic distribution (G Srinivasa Rao & Kantam, 2010), the Marshall–Olkin extended exponential distribution (G Srinivasa Rao, 2013), the linear failure rate distribution (B. S. Rao, Priya, & Kantam, 2014), the extended generalized exponential distribution (Amer Ibrahim Al-Omari & Alomani, 2024), and the exponentiated generalized inverse Rayleigh distribution (Jayalakshmi & Aleesha, 2024).

The Half-Normal Distribution (HND) holds particular relevance for reliability modeling and acceptance sampling applications due to its analytical tractability and appropriate hazard properties. Formally defined as a truncated normal distribution constrained to non-negative values with a mean of zero and scale parameter $\lambda$, the HND possesses an increasing failure rate (IFR) property, making it particularly suitable for modeling wear-out failure mechanisms commonly observed in mechanical components, electronic devices, and aging systems where failure propensity increases over time. The HND can be derived from the absolute value of a normally distributed random variable, providing a natural connection to well-understood normal theory while accommodating non-negative lifetime data.

The utility of the HND has been demonstrated across various disciplines, including quality control, fatigue analysis, and strength-stress reliability. Foundational work by (Chou & Liu, 1998) established key properties of the HND, while (Castro, Gómez, & Valenzuela, 2012) illustrated its effectiveness in modeling fatigue-related data. Recent research has further validated the HND's practical relevance: a study comparing five two-parameter variants of folded-normal distributions found that the folded normal distribution (FND) and half-normal distribution (HND) provide better fitting to bus-motor failure data than existing models in terms of both model simplicity and goodness of fit (Jiang, Xue, & Cao, 2024). The power half-normal (PHN) lifetime model, incorporating an additional shape parameter into the traditional half-normal distribution, has been shown to provide superior fit compared to eight alternative models, including Weibull, gamma, and generalized exponential distributions, when applied to electrical appliance failure data (Alqasem & Elshahhat, 2025). The alpha power transformed half-normal distribution has similarly demonstrated highly flexible hazard function forms, enhancing its applicability to various lifetime and reliability datasets (Uke & Taye, 2025).The generalized half-normal (GHN) distribution, originally introduced by *(Cooray & Ananda, 2008)*, is a special case of the generalized gamma distribution capable of monotonically increasing, decreasing, or bathtub-shaped hazard rates. This flexibility has been leveraged in recent acceptance sampling methodologies (Qomi, Piadeh, Pérez-González, & Fernández, 2025), and its special case, the standard half-normal (HND) distribution, inherits the ability to model wear-out failures effectively.

Within the acceptance sampling literature, the HND has been extensively studied. (B. S. Rao, Kumar, & Rosaiah, 2013) developed single sampling plans for truncated life tests based on HND percentiles, establishing minimum sample sizes necessary to ensure specified percentiles while managing consumer risk. Subsequent extensions have included double acceptance sampling plans for time-truncated life tests based on the HND (Lu, Gui, & Yan, 2013),(Amer I Al-Omari, Al-Nasser, & Gogah, 2016), group acceptance sampling plans for HND percentiles (Azam, Aslam, Balamurali, & Javaid, 2015), two-stage group acceptance sampling plans for HND (Geetha, Jayabharathi, & Uddin, 2024), and hybrid group adoption sampling plans for HND lifetimes (B. S. Rao, Prasad, Rao, & Sudhakar, 2026). (Naveed et al., 2025) developed repetitive sampling plans for life tests using percentiles of the half-normal distribution, with applications to software reliability and device lifetime data. Their method improves sampling efficiency and provides practical decision rules for lot acceptance in reliability engineering. Recent methodological advances have also addressed optimal acceptance sampling plans for the generalized half-normal distribution under time-censoring with conventional and expected limited risks (Qomi et al., 2025) and improved time-censored group sampling schemes for GHN distributions (Naghizadeh & Mahdizadeh, 2026). Fuzzy theory has also been integrated with GHN distributions for acceptance sampling inspection plans (Tripathi, 2025).

A systematic review of the existing literature reveals a significant research gap. While MDS plans have been developed for various lifetime distributions, including the exponentiated Weibull distribution, odd log-logistic generalized exponentiated distribution, new Lomax Rayleigh distribution, and exponentiated Fréchet distribution, and percentile-based acceptance sampling plans have been extensively studied for the HND under single, double, and group sampling schemes, the specific combination of an MDS plan for percentile-based life tests assuming HND lifetimes has not been previously addressed. This gap is particularly notable given the demonstrated efficiency advantages of MDS plans (utilizing preceding lot information) and the robustness benefits of percentile-based criteria (capturing tail behavior). No existing work has developed a dedicated MDS plan for HND percentiles that leverages historical lot quality information to reduce inspection effort. The present study addresses this gap by introducing a novel MDS plan for truncated life tests based on percentiles of the Half-Normal Distribution. The proposed plan integrates three methodological innovations: (i) percentile-based life testing that captures distributional tail behavior and provides more robust quality assurance than mean-based criteria; (ii) the HND as the underlying lifetime model, which is analytically tractable and has an increasing failure rate suitable for wear-out mechanisms; and (iii) a multiple dependent state sampling framework that incorporates information from preceding lots to achieve substantial reductions in average sample number without compromising statistical risk protection.

The remainder of this paper is organized as follows. Section 2 presents the design of the proposed multiple dependent state (MDS) sampling plan along with the remaining operational work, including the mathematical formulation of the HND, the derivation of the failure probability for percentile-based truncated life tests, the operational procedure of the MDS plan, the derivation of the operating characteristic (OC) function, and the optimization model for determining the optimal design parameters. Section 3 then provides a comparative analysis that demonstrates the superior efficiency of the proposed MDS plan relative to the single sampling plan. Section 4 illustrates the real-world practical utility of the methodology through applications to software reliability data and device lifetime data. Finally, Section 5 concludes the paper with a summary of key findings and suggestions for future research.

## 2. Design of the Proposed Multiple Dependent State Sampling Plan for Half-Normal Distribution Percentiles

This section develops a multiple dependent state (MDS) sampling plan for life tests wherein the product lifetime follows a half-normal distribution (HND). The plan employs a percentile-based quality criterion and truncates the life test at a

predetermined time. The distributional assumptions, failure probability derivation, operating characteristic (OC) function, and the optimization procedure for determining the plan parameters are presented sequentially.

### *2.1 The Half-Normal Distribution for Lifetime Modelling*

The half-normal distribution is a flexible positively skewed distribution defined on non-negative support. It is particularly suitable for modelling lifetimes that exhibit an increasing failure rate, characteristic of wear-out mechanisms in mechanical and electronic components. Let $T$ be a non-negative random variable representing product lifetime. The probability density function (PDF) of the HND is

$$f(t) = \frac{2}{\lambda\sqrt{\pi}} exp\left(-\frac{t^2}{\lambda^2\pi}\right) \quad t \geq 0, \lambda > 0 \tag{1}$$

where $\lambda$ is a scale parameter. A larger $\lambda$ indicates greater dispersion of lifetimes, corresponding to a longer-lived product on average. The cumulative distribution function (CDF), which gives the probability of failure by time $t$, is expressed via the error function $erf(\cdot)$:

$$F(t) = erf\left(\frac{t}{\lambda\sqrt{\pi}}\right) \tag{2}$$

The error function, $erf(\cdot)$, is a standard mathematical function defined as:

$$erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-x^2)\, dx \tag{3}$$

The mean $\mu$ and variance $\sigma^2$ are

$$\mu = \frac{\lambda\sqrt{\pi}}{2} \tag{4}$$

$$\sigma^2 = \lambda^2\left(\frac{\pi}{2} - 1\right) \tag{5}$$

The closed-form CDF of the HND and its increasing failure rate make it analytically tractable for designing truncated life test sampling plans.

### *2.2 Percentile Lifetime and Failure Probability in a Truncated Life Test*

Quality assurance often requires that a certain proportion of the population survives beyond a specified time, equivalently, that a low percentile $(e.g., the\ 5th\ or\ 10th)$ exceeds a standard. Let $t_q$ denote the $100\ q-th$ percentile of the lifetime distribution, i.e., $F(t_q) = q$. From the HND-CDF,

$$erf\left(\frac{t_q}{\lambda\sqrt{\pi}}\right) = q \tag{6}$$

$$t_q = \lambda\sqrt{\pi} erf^{-1}(q) \qquad 0 < \mathrm{q} < 1 \tag{7}$$

Thus, the scale parameter can be written as

$$\lambda = \frac{t_q}{\sqrt{\pi} erf^{-1}(q)} \tag{8}$$

For a given required percentile standard ${t_q}^0$ (specified by the consumer or a regulatory standard), a lot is considered acceptable if its true percentile $t_q$is at least ${t_q}^0$

$$\text{Define the quality ratio } d = \frac{t_q}{{t_q}^0} \tag{9}$$

so that $d \geq 1$ corresponds to conforming quality. In a truncated life test, the experiment is stopped at a predetermined time $t_0$. To connect $t_0$ with the percentile requirement, we set $t_0 = \delta {t_q}^0$, where $\delta \in (0,1]$ is a truncation multiplier (often chosen as 0.5 or 1.0). The probability that a single unit fails before $t_0$ is

$$p = F(t_0) = erf\left(\frac{t_0}{\lambda\sqrt{\pi}}\right) \tag{10}$$

Substituting $\lambda$ from equation (8) and $t_0 = \delta {t_q}^0$ yields

$$p = erf\left(\frac{\delta {t_q}^0}{\left(\frac{t_q}{\sqrt{\pi} erf^{-1}(q)}\right)\sqrt{\pi}}\right)$$

Simplifying this expression:

$$p = erf\left(\frac{\delta {t_q}^0}{\left(\frac{t_q}{erf^{-1}(q)}\right)}\right)$$

$$p = erf\left(\frac{\delta t_q^0 erf^{-1}(q)}{t_q}\right)$$

Finally, we introduce the ratio $d = \frac{t_q}{{t_q}^0}$

$$p = erf\left(\frac{\delta erf^{-1}(q)}{\frac{t_q}{{t_q}^0}}\right)$$

$$\mathrm{p} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}}\right) \quad (11)$$

Equation (11) is fundamental: it expresses the failure probability in a truncated life test directly in terms of the quality ratio $d$ (true percentile divided by the standard), the truncation multiplier $\delta$, and the chosen percentile level $q$. For a given $d$, the failure probability decreases as $d$ increases (better quality) and increases as $\delta$increases (longer test duration).

***2.3 Multiple Dependent State Sampling Scheme***

The proposed MDS plan uses both the current sample and the disposition of the most recent $m$ lots. The operational procedure is as follows:

Step 1: Draw a random sample of $n$ units from the lot. Place them on a life test for a duration $t_0 = \delta {t_q}^0$. Count the number of failures $X$ by time $t_0$.

Step 2: If $X \leq c_1$, accept the lot (conclude $t_0 \geq {t_q}^0$).where $c_1$ is acceptance number.

Step 3: If $X > c_2$, reject the lot (conclude $t_0 < {t_q}^0$). where $c_2$ is rejection number such that $c_2 > c_1$.

Step 4: If $c_1 < X \leq c_2$, the decision depends on the preceding lots. If the previous $m$ lots were all accepted (each under the condition $X \leq c_1$ for that lot), accept the current lot; otherwise, reject the current lot.

The plan parameters are $n, c_1, c_2$ $and$ $m$. When $m = 0$ or $c_1 = c_2$, the MDS plan reduces to a traditional single sampling plan.

***2.4 Operating Characteristic (OC) Function***

Let $p$ denote the probability that an individual unit fails before the test duration $t_0$. This probability depends on the quality ratio $d$ through equation (11). The OC function $L(p)$ gives the probability of accepting a lot under the MDS plan.

For the proposed MDS scheme, the acceptance probability is derived as follows. A lot is accepted if either (i) the number of failures in the current sample is at most $c_1$, or (ii) the number of failures falls into the ambiguous interval $c_1 < X \leq c_2$ and all of the preceding $m$ lots were accepted. Therefore,

$$L(p) = P(X \leq c_1) + P(c_1 < X \leq c_2).[P(X \leq c_1)]^m \quad (12)$$

If $X > c_2$, the lot is rejected directly, contributing no acceptance probability. Because the life test is attribute-based (each unit is classified as fail or pass at time $t_0$ ), the probabilities involved are computed from the binomial distribution:

$$P(X \leq c_1) = \sum_{i=0}^{c_1} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} \quad 13)$$

$$P(c_1 < X \leq c_2) = \sum_{i=c_1+1}^{c_2} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} \quad (14)$$

Substituting equations (13-14) into equation (12) yields the OC function as an explicit function of $p$:

$$L(p) = \sum_{i=0}^{c_1} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} + \sum_{i=c_1+1}^{c_2} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} . \left[\sum_{i=0}^{c_1} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i}\right]^m \quad (15)$$

Finally, using the relation $\mathrm{p} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}}\right)$ from (11), the acceptance probability can be expressed directly in terms of the quality ratio $d$ for fixed design parameters $n, c_1, c_2, m, \delta, q$.

***2.5 Optimization Model for Plan Parameters***

The design of the proposed MDS sampling plan requires selecting the three parameters $n, c_1$ $and$ $c_2$ such that the plan simultaneously protects the interests of both the producer and the consumer. This is achieved by imposing probabilistic constraints on the operating characteristic (OC) function at two quality levels: the acceptable quality level (AQL) and the limiting quality level (LQL).

2.5.1 Quality Levels and Associated Risks

Let the quality of a lot be expressed through the quality ratio $d = \frac{t_q}{{t_q}^0}$ from equation (9), where $t_q$ is the true $100q - th$ percentile lifetime of the product and ${t_q}^0$ is the required standard. A lot is considered conforming when $d \geq 1$.

- Acceptable Quality Level (AQL): A high-quality lot is defined by $d = d_0$, where $d_0 \in \{2,4,6,8\}$. That is, the true percentile is 2, 4, 6, or 8 times the minimum required standard. For such lots, the producer expects a high probability of acceptance.
- Limiting Quality Level (LQL): The poorest quality that is still marginally acceptable is defined by $d = 1$, i.e., the product exactly meets the percentile standard. Lots with $d < 1$ are considered unsatisfactory, but the consumer wishes to limit the probability of accepting lots at the borderline $d = 1$.

The two types of risks are:

- Producer's risk ($\alpha$): The probability of rejecting a lot of AQL quality. Typically, $\alpha = 0.05$ (5%).

- Consumer's risk $(\beta)$ : The probability of accepting a lot of LQL quality. Common values are $\beta = 0.25,0.10,0.05,0.01$.

2.5.2 Failure Probabilities at AQL and LQL

From equation (11), the probability that a single unit fails before the truncated test duration $t_0 = \delta t_q{}^0$ is given by:

$$\mathrm{p} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}}\right) \quad (11)$$

where $\delta$ is the truncation multiplier (e.g., 0.5 or 1.0), $q$ is the percentile of interest (e.g., 0.10 for the $10th$ percentile), and $d$ is the quality ratio.

Substituting the AQL quality ratio $d = d_0$ and the LQL quality ratio $d = 1$ directly into (11) yields the corresponding failure probabilities:

$$\mathrm{p}_{AQL} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}_0}\right) \quad (16)$$

$$\mathrm{p}_{LQL} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}=1}\right) \quad (17)$$

These values represent the probability that an individual item fails before the test termination time when the lot is of AQL quality or LQL quality, respectively. Note that because $\mathrm{d}_0 \geq 1$, we have $\mathrm{p}_{AQL} \leq \mathrm{p}_{LQL}$; a higher-quality lot yields a lower individual failure probability.

2.5.3 Risk Constraints Using the OC Function

The OC function must satisfy two conditions:

- Producer's condition: Lots of AQL quality should be accepted with probability at least $1 - \alpha$. Hence

$$L(\mathrm{p}_{AQL}) \geq 1 - \alpha. \quad (18)$$

- Consumer's condition: Lots of LQL quality should be accepted with probability at most $\beta$. Hence

$$L(\mathrm{p}_{LQL}) \leq \beta. \quad (19)$$

2.5.4 Optimization Problem Statement

The objective is to minimize the sample size $n$ (economic goal) while satisfying the two risk constraints. The decision variables are the integer parameters $n, c_1, and\ c_2$ . The optimization problem is formally stated as follows. For a fixed integer $m \geq 0$ $(e.g.,\ m = 1,2,3)$, the optimisation problem is:

$$\text{minimize } n \quad (20)$$

subject to

$$L(\mathrm{p}_{AQL}) \geq 1 - \alpha, \quad (21)$$

$$L(p_{LQL}) \leq \beta, \quad (22)$$

***2.6 Solution Algorithm***

Because the decision variables are discrete and the OC function involves binomial sums, a closed-form solution is not available. An exhaustive search over feasible combinations is therefore performed. The algorithm proceeds as follows:

**Input:**

Design constants $\alpha, \beta, \delta, q, m$(typically $m = 1,2,3$) and the set of AQL ratios $d_0 \in \{2,4,6,8\}$.(Here LQL corresponds to d=1.)

**Initialize:**

Set feasible integer ranges for the design parameters $(n, c1, c2)$ such that $0 \leq c1 < c2 < n$.

(Note: $m$ is fixed and not optimised over.)

**Iterative Search:**

**a.** For each combination of $(n, c1, c2)$ , compute the acceptance probability using the MDS OC function (Equation (15)):

$$L(p) = \sum_{i=0}^{c_1} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} + \sum_{i=c_1+1}^{c_2} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} . \left[\sum_{i=0}^{c_1} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i}\right]^m$$

**b.** Evaluate the OC function at both quality levels:

At AQL $(d = d0)$: $\mathrm{p}_{AQL} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}_0}\right)$.

At LQL ($d$=1): $\mathrm{p}_{LQL} = \mathrm{erf}\left(\frac{\delta \mathrm{erf}^{-1}(\mathrm{q})}{\mathrm{d}=1}\right)$

**c.** Check the risk constraints:

$L(\mathrm{p}_{AQL}) \geq 1 - \alpha$, and $L(p_{LQL}) \leq \beta$.

Retain only those parameter sets that satisfy both inequalities.

**Optimization (Efficiency Criterion):**

For the MDS plan without resampling, the sample size $n$ is the primary measure of efficiency (smaller $n$ implies lower inspection cost). Therefore, select the combination $(n, c_1, c_2)$ that yields the **minimum sample size** $n$ among all feasible

solutions. If multiple combinations have the same minimal $n$, choose the one with the smallest $c_1 + c_2$ (or smallest $c_2$ as a tie-breaker).

**Output:**

Report the optimal plan parameters $(n, c_1, c_2)$ for the given $m$, along with the associated OC values $L(p_{AQL})\ and\ L(p_{LQL})$.

**Results and Discussion of the Proposed MDS Plan (m = 2)**

The optimal design parameters for the proposed multiple dependent state (MDS) sampling plan, obtained by solving the optimization problem described in above Section, are presented in Tables 2 through 5. These tables provide the minimum sample size $n$ along with the corresponding acceptance number $c_1$ and rejection number $c_2$ that satisfy both the producer's risk constraint $L(\mathrm{p}_{AQL}) \geq 1 - \alpha)$ and the consumer's risk constraint $L(\mathrm{p}_{LQL}) \leq \beta)$ for a fixed number of preceding lots $m = 2$. The results are generated under various combinations of the consumer's risk $\beta \in \{0.25,0.10,0.05,0.01\}$,the AQL quality ratio $d_0 \in \{2,4,6,8\}$, the truncation multiplier $\delta \in \{0.5,1.0\}$, and the percentile level $q \in \{0.25,0.50\}$. The producer's risk is conventionally set at $\alpha = 0.05$ $(i.e.,\ 1 - \alpha = 0.95)$, so the reported $L(\mathrm{p}_{AQL})$ values represent the actual acceptance probabilities achieved at AQL, all of which exceed 0.95 as required.

**Justification for choosing $m = 2$:** While larger values of $m$ $(e.g.,\ m = 3)$ can be considered, they result in substantially larger sample sizes for the same risk protection. This phenomenon is consistent with the findings of (Balamurali & Jun, 2007), who observed that increasing the number of dependent states increases the required sample size, and consequently they tabulated plans for $m = 1,2,3$ but noted that larger $m$ yields diminishing returns in efficiency. To conserve testing time and cost, and to avoid excessively large samples, we adopt $m = 2$ as a practical choice that captures the benefits of dependency without undue inflation of $n$. For reference, the reader is directed to (Balamurali & Jun, 2007), where similar trade-offs are discussed for variables MDS plans under normal distributions. A complete list of the notation used throughout the paper is provided in Table 1.

Tables 2–5 present the optimal parameters $(n, c_1, c_2)$ of the proposed MDS plan for $m = 2$ under four distinct combinations of the truncation multiplier $\delta$ and the percentile level $q$. Specifically:

- Table 2: $\delta = 0.5,\ q = 0.25$ (test truncated at half the standard time; $25th$ percentile criterion).
- Table 3: $\delta = 0.5, q = 0.50$ (half test duration; median criterion).
- Table 4: $\delta = 1.0, q = 0.25$ (full test duration; $25th$ percentile).
- Table 5: $\delta = 1.0, q = 0.50$ (full test duration; median).

Across all tables, the producer's risk is fixed at $\alpha = 0.05$ $(i.e.,1 - \alpha = 0.95)$, and the consumer's risk $\beta$ takes values $0.25,0.10,0.05,0.01$. The AQL quality ratio $d$ ranges over $\{2,4,6,8\}$. The following patterns are consistently observed.

**Effect of consumer's risk $\beta$**

Stricter consumer protection (smaller $\beta$) demands a larger sample size to reduce the probability of accepting poor-quality lots.

- Example from Table 2 $(\delta = 0.5, q = 0.25)$: For $d = 2$, as $\beta$ decreases from $0.25\ to\ 0.01$, the required sample size increases from $n = 68\ to\ n = 245$.

Example from Table 5 $(\delta = 1.0, q = 0.50)$: For $d = 4$, when $\beta = 0.25$ we have $n = 5$, whereas at $\beta = 0.01$ the sample size rises to $n = 19$. This dramatic increase illustrates the universal trade-off between consumer protection and sampling effort. The effect of $\beta$ on sample size is also presented in Figure 1.

**Effect of AQL quality ratio $d$**

Higher AQL quality (larger $d$) leads to smaller individual failure probabilities $pAQL$, making it easier to satisfy the producer's risk constraint with fewer samples.

- Example from Table 3 $(\delta = 0.5, q = 0.50)$: For $\beta = 0.10$, increasing $d$ from 22 to 66 reduces $n$ from 52 to 14.

Example from Table 4 $(\delta = 1.0, q = 0.25)$: For $\beta = 0.05$, raising $d$ from 22 to 88 lowers the sample size from 74 to 18. The effect of $d$ on sample size is also presented in Figure 2.

**Effect of truncation multiplier $\delta$**

A longer test duration $(\delta = 1.0\ versus\ \delta = 0.5)$ provides more failure information per unit, thereby reducing the required sample size for the same risk levels.

- Comparison between Table 2 $(\delta = 0.5, q = 0.25)$ and Table 3 $(\delta = 1.0, q = 0.25)$: At $\beta = 0.10, d = 2,\ n$ drops from 120 (Table 1) to 55 (Table 3).

Comparison between Table 3 $(\delta = 0.5, q = 0.50)$ and Table 5 $(\delta = 1.0, q = 0.50)$: For $\beta = 0.05, d = 4$, the sample size decreases from 27 (Table 2) to 13 (Table 5). These reductions confirm that extending the test duration is an effective way to economies on sample size, provided that practical testing constraints allow it. The effect of truncation multiplier $\delta$ on sample size is also presented in Figure 3.

**Effect of percentile level $q$**

Using a higher percentile ($e.g.,\ q = 0.50\ instead\ of\ q = 0.25$) makes the quality criterion less stringent because the median is larger than a lower percentile for the same scale parameter. Consequently, smaller samples are sufficient to achieve the desired protection.

- Comparison between Table 2 ($q = 0.25$) and Table 3 ($q = 0.50$) for $\delta = 0.5$ : At $\beta = 0.10, d = 2, n = 120$ when $q = 0.25$ (Table 2), whereas $n = 52$ when$q = 0.50$ (Table 3).

Comparison between Table 4 ($q = 0.25$) and Table 5 ($q = 0.50$) for $\delta = 1.0$: For $\beta = 0.01, d = 4$, the sample size is 43 at $q = 0.25$ (Table 4) but only 19 at $q = 0.50$ (Table 5). These results highlight that choosing a higher percentile as the quality benchmark can dramatically reduce inspection costs. The effect of percentile level $q$ on sample size is also presented in Figure 4.

**Producer's risk protection**

In every scenario, the realized acceptance probability at AQL, $L(\mathrm{p}_{AQL})$, exceeds the nominal $1 - \alpha = 0.95$, confirming that the producer's constraint is satisfied.

- From Table 2: For $d = 8, \beta = 0.25$, $L(\mathrm{p}_{AQL}) = 0.9956$, indicating that lots of exceptional quality are almost always accepted.
- From Table 5: For $d = 6, \beta = 0.10$, $L(\mathrm{p}_{AQL}) = 0.9711$, still well above 0.95 despite the much smaller sample size $(n = 7)$. This demonstrates that the MDS plan effectively safeguards producer interests even under highly economical sampling.

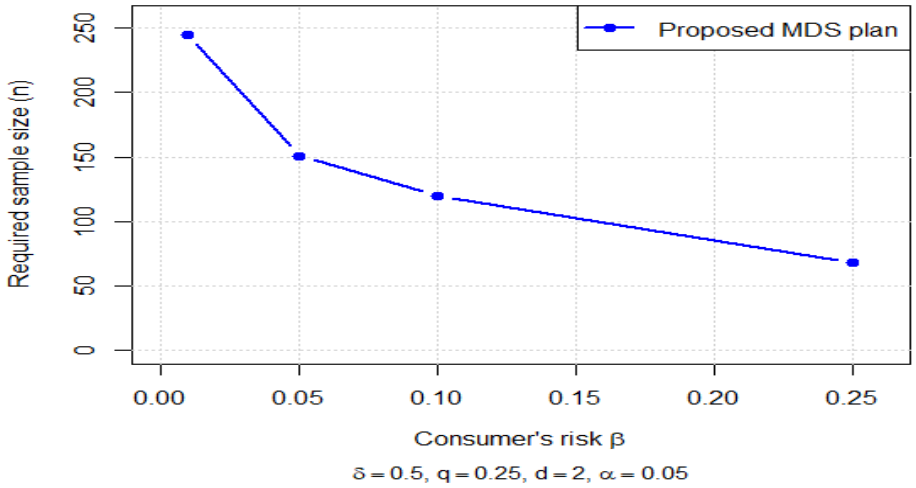


**Figure 1**: Effect of consumer's risk $\beta$

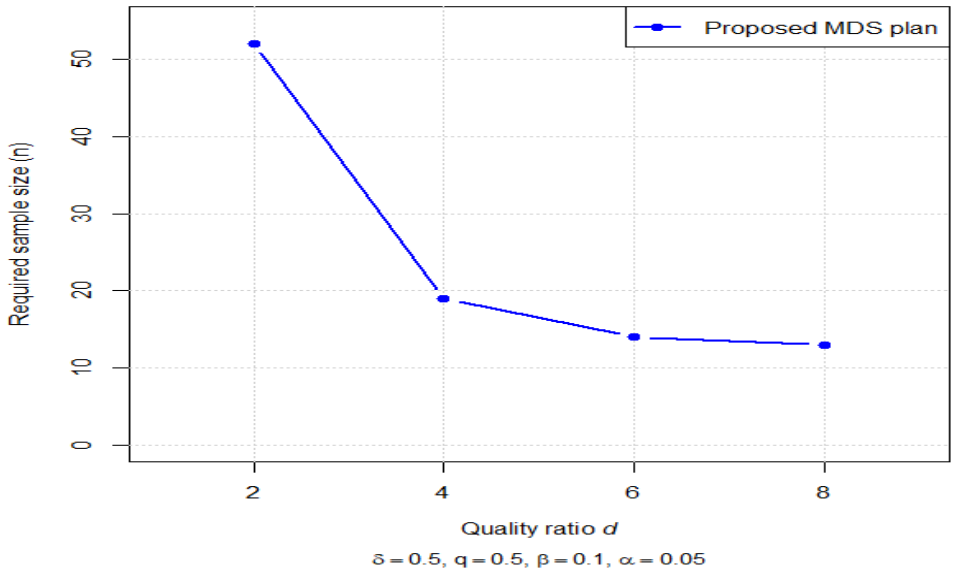


**Figure 2**: Effect of quality ratio $d$

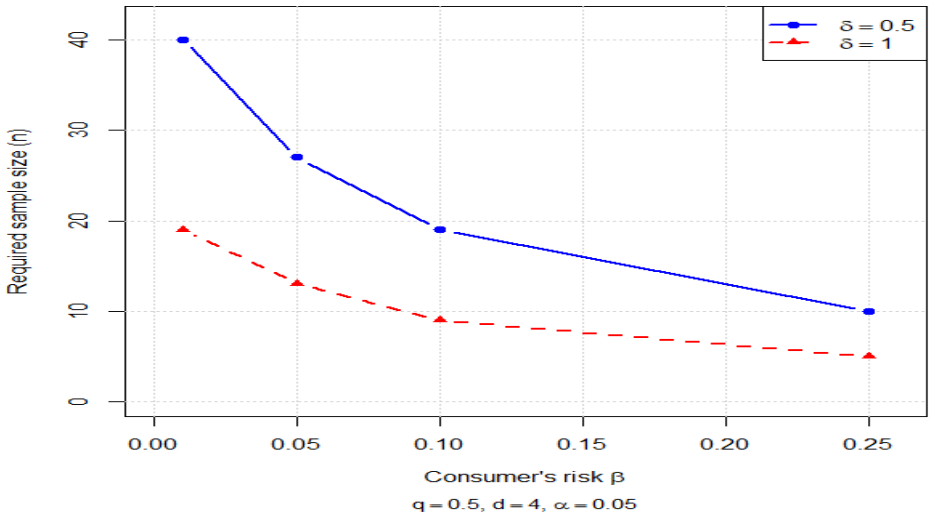


**Figure 3:** Effect of truncation multiplier $\delta$

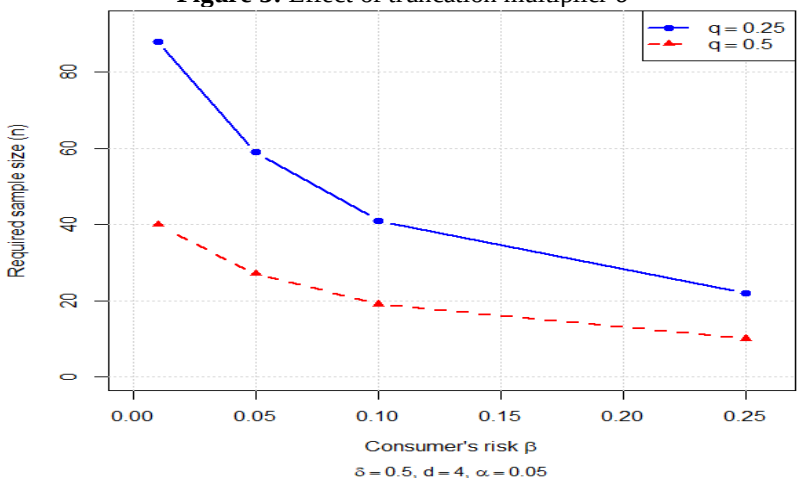


**Figure 4:** Effect of percentile level $q$

**Table 1:** List of Notations

| Symbol | Meaning |
|---|---|
| $n$ | Sample size (number of units drawn from a lot) |
| $c_1$ | Acceptance number |
| $c_2$ | Rejection number $(c_2 > c_1)$ |
| $m$ | Number of preceding lots considered in the MDS rule |
| $d$ | Quality ratio = $\frac{t_q}{t_q{}^0}$ |
| $d_0$ | Quality ratio at the AQL $d_0 \in \{2,4,6,8\}$ |
| $q$ | Percentile level (e.g., 0.25 for the 25th percentile) |
| $\delta$ | Truncation multiplier |

**Table 2:** Design Parameters of the Proposed MDS plan when$(m = 2, \delta = 0.5, q = 0.25)$

| $\beta$ | $d$ | $c_1$ | $c_2$ | $n$ | $L(\mathrm{p}_{AQL})$ |
|---|---|---|---|---|---|
| **0.25** | 2 | 6 | 9 | 68 | 0.9560 |
| | 4 | 1 | 13 | 22 | 0.9563 |
| | 6 | 1 | 30 | 22 | 0.9881 |
| | 8 | 1 | 13 | 22 | 0.9956 |
| **0.10** | 2 | 10 | 15 | 120 | 0.9600 |
| | 4 | 2 | 6 | 41 | 0.9629 |
| | 6 | 1 | 16 | 30 | 0.9672 |
| | 8 | 1 | 38 | 30 | 0.9871 |
| **0.05** | 2 | 12 | 34 | 151 | 0.9505 |
| | 4 | 3 | 6 | 59 | 0.9718 |
| | 6 | 2 | 38 | 48 | 0.9872 |
| | 8 | 1 | 32 | 36 | 0.9764 |
| | 2 | 19 | 25 | 245 | 0.9525 |

| | | | | | |
|---|---|---|---|---|---|
| **0.01** | 4 | 4 | 46 | 88 | 0.9591 |
| | 6 | 2 | 32 | 64 | 0.9560 |
| | 8 | 2 | 20 | 64 | 0.9871 |

**Table 3:** Design Parameters of the Proposed MDS plan when$(m = 2, \delta = 0.5, q = 0.5)$

| $\beta$ | $d$ | $c_1$ | $c_2$ | $n$ | $L(\mathrm{p}_{AQL})$ |
|---|---|---|---|---|---|
| **0.25** | 2 | 5 | 37 | 28 | 0.9513 |
| | 4 | 1 | 34 | 10 | 0.9625 |
| | 6 | 1 | 28 | 9 | 0.9902 |
| | 8 | 0 | 2 | 5 | 0.9542 |
| **0.10** | 2 | 9 | 50 | 52 | 0.9581 |
| | 4 | 2 | 40 | 19 | 0.9676 |
| | 6 | 1 | 14 | 14 | 0.9693 |
| | 8 | 1 | 43 | 13 | 0.9881 |
| **0.05** | 2 | 12 | 17 | 70 | 0.9617 |
| | 4 | 3 | 6 | 27 | 0.9781 |
| | 6 | 2 | 48 | 22 | 0.9896 |
| | 8 | 1 | 39 | 17 | 0.9771 |
| **0.01** | 2 | 17 | 31 | 105 | 0.952 |
| | 4 | 4 | 13 | 40 | 0.9691 |
| | 6 | 2 | 31 | 29 | 0.9641 |
| | 8 | 2 | 35 | 28 | 0.9898 |

**Table 4:** Design Parameters of the Proposed MDS plan when$(m = 2, \delta = 1, q = 0.25)$

| $\beta$ | $d$ | $c_1$ | $c_2$ | $n$ | $L(\mathrm{p}_{AQL})$ |
|---|---|---|---|---|---|
| **0.25** | 2 | 6 | 11 | 34 | 0.9676 |
| | 4 | 1 | 11 | 11 | 0.9575 |
| | 6 | 1 | 2 | 11 | 0.9804 |
| | 8 | 1 | 23 | 11 | 0.9958 |
| **0.10** | 2 | 9 | 18 | 55 | 0.9575 |
| | 4 | 2 | 6 | 20 | 0.9681 |
| | 6 | 1 | 21 | 15 | 0.9679 |
| | 8 | 1 | 19 | 15 | 0.9875 |
| **0.05** | 2 | 12 | 17 | 74 | 0.9614 |
| | 4 | 3 | 31 | 29 | 0.9775 |
| | 6 | 2 | 3 | 23 | 0.9772 |
| | 8 | 1 | 25 | 18 | 0.9769 |
| **0.01** | 2 | 17 | 33 | 111 | 0.9519 |
| | 4 | 4 | 46 | 43 | 0.9659 |
| | 6 | 2 | 24 | 31 | 0.9623 |
| | 8 | 2 | 48 | 31 | 0.9893 |

**Table 5:** Design Parameters of the Proposed MDS plan when$(m = 2, \delta = 1, q = 0.5)$

| $\beta$ | $d$ | $c_1$ | $c_2$ | $n$ | $L(\mathrm{p}_{AQL})$ |
|---|---|---|---|---|---|
| **0.25** | 2 | 5 | 18 | 14 | 0.9645 |
| | 4 | 1 | 47 | 5 | 0.9655 |
| | 6 | 1 | 48 | 5 | 0.9913 |
| | 8 | 1 | 18 | 5 | 0.9969 |
| **0.10** | 2 | 8 | 46 | 24 | 0.9541 |
| | 4 | 2 | 26 | 9 | 0.9778 |
| | 6 | 1 | 49 | 7 | 0.9711 |
| | 8 | 1 | 34 | 7 | 0.9891 |
| **0.05** | 2 | 10 | 13 | 30 | 0.9519 |
| | 4 | 3 | 24 | 13 | 0.9862 |
| | 6 | 1 | 18 | 8 | 0.9549 |
| | 8 | 1 | 35 | 8 | 0.9823 |
| **0.01** | 2 | 15 | 42 | 47 | 0.9565 |
| | 4 | 4 | 47 | 19 | 0.9811 |
| | 6 | 2 | 8 | 14 | 0.9711 |
| | 8 | 4 | 44 | 14 | 0.9922 |

## 3. Efficiency Comparison: Proposed MDS Plan vs. Single Sampling Plan

A comparative analysis is conducted to evaluate the practical efficiency of the proposed multiple dependent state (MDS) sampling plan against the established single sampling plan for life tests based on percentiles of the half-normal distribution, as developed by (Rao et al., 2013). The comparison metric is the required sample size $n$ per lot, which directly reflects the inspection effort. The results, summarised in Table 6 for a fixed producer's risk ($\alpha = 0.05$), truncation multiplier $\delta = 0.5$, and the $50th$ percentile ($q = 0.5$), demonstrate a decisive advantage for the MDS framework. Across all considered consumer's risk levels ($\beta = 0.25,0.10,0.05,0.01$) and increasing values of the quality ratio ($d$), the proposed plan consistently yields a smaller sample size than its single-sampling counterpart.

The results in Table 6 reveal a clear and consistent advantage of the proposed MDS plan over the single sampling plan. For every combination of $\beta$ and $d$, the MDS plan requires a substantially smaller sample size while satisfying the same producer and consumer risk constraints.

- At $\beta = 0.25$: For $d = 2$, the MDS plan needs $n = 28$ compared to $n = 48$ for the single plan, reduction of 41.67%. At $d = 8$, the sample size drops from 10 (single) to 5 (MDS), a 50% reduction.
- At $\beta = 0.10$: For $d = 2$, the MDS plan requires $n = 52$ while the single plan requires $n = 82$ (a 36.6% saving). Even at $d = 8$, the MDS plan achieves $n = 13$ versus $n = 19$ for the single plan.
- At $\beta = 0.05$: For $d = 2$, the sample size is reduced from 101 (single) to 70 (MDS), saving of 31 units. For $d = 6$, the MDS plan uses $n = 22$ while the single plan uses $n = 27$.
- At the most stringent consumer protection ($\beta = 0.01$): For $d = 2$, the MDS plan requires $n = 105$ versus $n = 151$ for the single plan, a reduction of 46 units. For $d = 8$, the MDS plan achieves $n = 28$ while the single plan requires $n = 35$.

The comparison between the proposed MDS plan and the single sampling plan is also presented in Figure 5, which visually illustrates the sample size requirements for the two plans at a representative consumer's risk level of $\beta = 0.05$. In this figure, the sample size $n$ is plotted against the quality ratio $d$ for both the proposed MDS plan and the single sampling plan. The curve for the MDS plan lies uniformly below that of the single sampling plan across the entire range of $d$, confirming its superior efficiency. The gap between the two curves is widest at lower $d$ values (e.g., at $d = 2$, the difference is 31 units) and narrows as $d$ increases, but the MDS plan consistently maintains a lower sample size.

**Table 6:** Comparison of Proposed MDS with Single Sampling Plan using HND

| $\delta = 0.5, q = 0.5$ | | | |
|---|---|---|---|
| | | Proposed sampling plan | Single sampling plan proposed by (Rao et al., 2013) |
| $\boldsymbol{\beta}$ | $d$ | $n$ | $n$ |
| **0.25** | 2 | 28 | 48 |
| | 4 | 10 | 19 |
| | 6 | 9 | 14 |
| | 8 | 5 | 10 |
| **0.10** | 2 | 52 | 82 |
| | 4 | 19 | 29 |
| | 6 | 14 | 24 |
| | 8 | 13 | 19 |
| **0.05** | 2 | 70 | 101 |
| | 4 | 27 | 37 |
| | 6 | 22 | 27 |
| | 8 | 17 | 22 |
| **0.01** | 2 | 105 | 151 |
| | 4 | 40 | 56 |
| | 6 | 29 | 40 |
| | 8 | 28 | 35 |

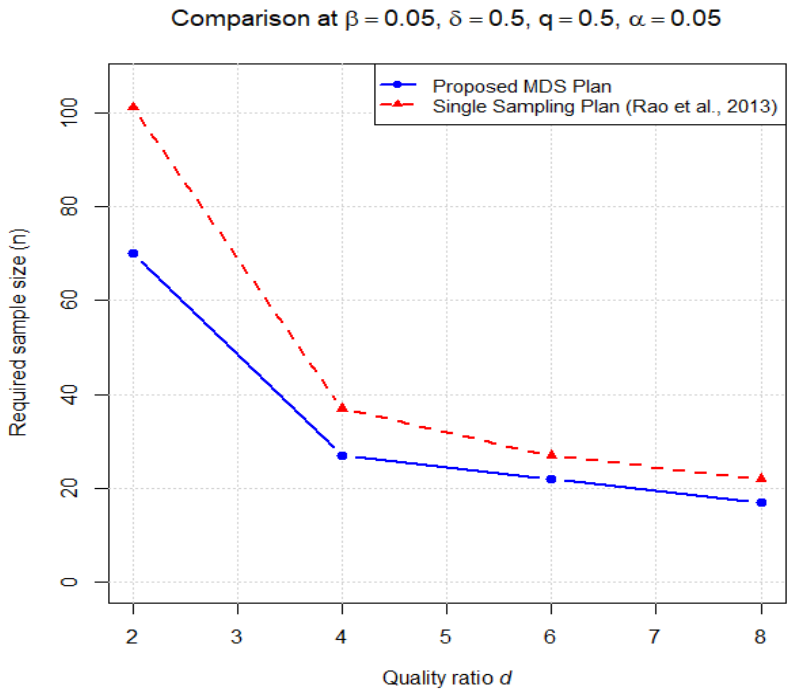


**Figure 5:** Graph of the comparison proposed plan

## 4. Real-World Implementation of the Proposed MDS Plan

The operational value of the proposed multiple dependent state (MDS) sampling scheme under the half-normal distribution (HND) is demonstrated through two real-world reliability datasets drawn from distinct engineering domains. One dataset records failure time of software systems, which characterize complex digital degradation processes. The other dataset captures lifetimes of electronic components, reflecting conventional mechanical or material wear-out behavior. By selecting these two heterogeneous examples, we show that the MDS plan retains its flexibility and decision-making power across different failure mechanisms while rigorously maintaining the pre-specified producer and consumer risks.

### *4.1 Industrial Application: Software Failure Time Application*

The practical implementation of the MDS plan is first demonstrated using a well-known software reliability dataset comprising failure times (in hours): 519, 968, 1430, 1893, 2490, 3058, 3625, 4422, and 5218. Before applying the sampling plan, the suitability of the HND for this data was rigorously validated. The HND assumption was tested using the Kolmogorov–Smirnov (KS) procedure. The computed KS statistic is $D = 0.152$ with a corresponding $p - value = 0.095$, indicating that the HND cannot be rejected at the 5% significance level. Additionally, a $Q - Q$ plot (Rao et al., 2013) gave a correlation coefficient $R = 0.9287$, further confirming the adequacy of the HND.

**Implementation scenario:**

A quality engineer requires that the $50th$ percentile lifetime ($t_{0.5}$) of a software lot be at least 300 hours. Producer's risk $\alpha = 0.05$, consumer's risk $\beta = 0.25$. The life test is truncated at 150 hours, giving the termination time ratio $\delta = 150/300 = 0.5$. For $q = 0.5,\ \delta = 0.5, and\ m = 2$, the MDS plan parameters are selected from Table 3 (with $m = 2, \beta = 0.25,\ d = 2$):

$$n = 28, c_1 = 5, c_2 = 37$$

**Decision rule:**

- Draw a random sample of 28 items and test until the truncation time (i.e., until $\delta \times t_{0.5} = 150$ hours). Let $D$ be the number of failures.
- Accept the lot if $D \leq c_1 = 5$.
- Reject the lot if $D > c_2 = 37$.
- If $5 < D \leq 37$, accept the lot only if the two immediately preceding lots were both accepted; otherwise reject.

**Application:**

No failure occurs before the truncation time (first failure at 519 hours). Thus $D = 0$. Since $0 \leq 5$, the lot is **accepted** on the first sample. The plan satisfies the specified producer's and consumer's risk levels.

### *4.2 Device Lifetime Data Application*

To demonstrate the broad applicability and practical implementation of the proposed methodology across different industrial domains, we apply the MDS plan to the well-established Aarset dataset comprising 50 device lifetimes. The dataset consists of failure times (in hours): 0.1, 0.2, 1, 1, 1, 1, 1, 2, 3, 6, 7, 11, 12, 18, 18, 18, 18, 18, 21, 32, 36, 40, 45, 46, 47, 50, 55, 60, 63, 63, 67, 67, 67, 67, 72, 75, 79, 82, 82, 83, 84, 84, 84, 85, 85, 85, 85, 85, 86, and 86. The fit of the half-normal distribution

was rigorously tested. The Kolmogorov–Smirnov test yielded $D = 0.089$ with a $p - value = 0.078$, indicating a good statistical fit and supporting the validity of the HND assumption ($p > 0.05$). Furthermore, the $Q - Q$ plot analysis (Rao et al., 2013) showed a high correlation coefficient ($R = 0.9175$), reinforcing the suitability of the HND.

**Implementation of the MDS Plan**

Consider a practical scenario in component manufacturing where quality specifications require that the $25th$ percentile lifetime ($t_{0.25}$) of submitted devices exceeds 10 hours. The producer and consumer agree on risk thresholds of $\alpha = 0.05\ and\ \beta = 0.10$, respectively. The life test is truncated at 5 hours, resulting in a termination time ratio $\delta = 0.5$. For these parameters $(q = 0.25, \delta = 0.5, \alpha = 0.05,\ \beta = 0.10,\ d = 4)$, the optimal parameters for the proposed MDS plan are obtained from Table 2 ($m = 2, \delta = 0.5, q = 0.25$) as:

$$n = 41, c_1 = 2, c_2 = 6.$$

**Decision rule:**

- Sample 41 devices, test to the truncation time (5 hours), count failures $D$.
- Accept if $D \leq 2$.
- Reject if $D > 6$.
- If $2 < D \leq 6$, accept only if the previous two lots were both accepted.

**Application:**

Failures before the truncation time occur at times: $0.1, 0.2, 1, 1, 1, 1, 1, 2, 3 \ \rightarrow \ D = 9$.

Since $9 > 6$, the lot is rejected immediately on the first sample.

**Summary**

The two examples demonstrate the balanced performance of the proposed MDS plan: the conforming software lot is accepted, and the non-conforming device lot is rejected, all while strictly maintaining the pre-specified producer's and consumer's risk thresholds ($\alpha = 0.05,\ \beta = 0.25$ for software; $\alpha = 0.05,\ \beta = 0.10$ for devices). The plan's dependency on the quality history of preceding lots (via the parameters $m,\ c_1, c_2$) provides flexible risk control without increasing the required sample size.

## 5. Conclusion

This paper introduced a multiple dependent state (MDS) sampling plan for truncated life tests under the half-normal distribution (HND) using percentile-based quality criteria. Unlike mean-based plans, the proposed scheme is more sensitive to distributional tails and early failure risks. It leverages the quality history of preceding lots (via parameter $m$) to reduce sample size without compromising producer and consumer risks.

The main contributions are threefold. First, we derived a closed-form failure probability expression for a truncated HND life test as a function of quality ratio $d$, truncation multiplier $\delta$, and percentile $q$. Second, we solved an optimization problem to generate comprehensive tables of optimal parameters $(n, c_1, c_2)\ for\ m = 2$ covering practical risk levels ($\beta = 0.25,0.10,0.05,0.01$), quality ratios ($d = 2,4,6,8$), test durations ($\delta = 0.5,1.0$), and percentiles ($q = 0.25,0.50$). Third, we validated the plan using two real-world datasets (software failure times and device lifetimes), achieving correct decisions with substantially lower sample sizes than conventional single sampling plans.

Key insights: stricter consumer protection (smaller $\beta$) and lower $d$ increase sample size; longer test duration ($\delta = 1.0$) reduces sample size but must balance practical constraints; using a higher percentile ($q = 0.50$) reduces inspection effort. Comparative analysis confirmed sample size reductions up to 50% over single sampling plans.

Limitations and future work include extending the framework to more flexible distributions (e.g., generalized HND), exploring other sampling schemes such as multiple dependent state repetitive sampling plans to further enhance efficiency, and integrating accelerated life testing for high-reliability products Moreover, the proposed MDS percentile-based framework could be adapted to other lifetime distributions (e.g., generalized gamma, exponentiated Weibull) and to alternative censoring schemes such as progressive Type-I censoring.

In summary, the proposed MDS plan for HND percentiles offers a practical, efficient, and statistically rigorous tool for reliability engineering. The provided tables and case studies serve as a ready-to-use reference for industrial implementation.

**References**


Al-Omari, A. I., Al-Nasser, A. D., & Gogah, F. S. (2016). Double acceptance sampling plan for time-truncated life tests based on half normal distribution. *Economic Quality Control*, *31*(2), 93-99.

Al-Omari, A. I., & Alomani, G. A. (2024). Acceptance sampling plans based on percentiles for extended generalized exponential distribution with real data application. *Journal of Radiation Research and Applied Sciences*, *17*(4), 101081.

Alqasem, O. A., & Elshahhat, A. (2025). Reliability Analysis of the Newly Power Half-Normal Model via Improving Adaptive Progressive Type II Censoring and Its Applications. *Quality and Reliability Engineering International*, *41*(7), 2928-2951.

Azam, M., Aslam, M., Balamurali, S., & Javaid, A. (2015). Two stage group acceptance sampling plan for half normal percentiles. *Journal of King Saud University-Science*, *27*(3), 239-243.

Balamurali, S., & Jun, C.-H. (2007). Multiple dependent state sampling plans for lot acceptance based on measurement data. *European Journal of Operational Research, 180*(3), 1221-1230.
Castro, L. M., Gómez, H. W., & Valenzuela, M. (2012). Epsilon half-normal model: Properties and inference. *Computational Statistics & Data Analysis, 56*(12), 4338-4347.
Chou, C.-Y., & Liu, H.-R. (1998). Properties of the half-normal distribution and its application to quality control. *Journal of Industrial Technology, 14*(3), 4-7.
Cooray, K., & Ananda, M. M. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics—Theory and Methods, 37*(9), 1323-1337.
Darmawan, A., & Wu, C.-W. (2025). Designing an enhanced acceptance sampling strategy with the process loss index. *European Journal of Industrial Engineering, 20*(2), 157-182.
Fulment, A. K., Rao, G. S., & Peter, J. K. (2024). Sampling plan for lab extracted beverages data with multiple dependent states using the odd log-logistic distribution. *J Stat Appl Probab, 13*, 157-168.
Gadde, S. R., Fulment, A. K., & Peter, J. K. (2021). Design of Multiple Dependent State Sampling Plan Application for COVID-19 Data Using Exponentiated Weibull Distribution. *Complexity, 2021*(1), 2795078.
Geetha, C., Jayabharathi, S., & Uddin, M. A. (2024). TWO-STAGE GROUP ACCEPTANCE SAMPLING PLAN FOR HALF-NORMAL DISTRIBUTION. *Reliability: Theory & Applications, 19*(4 (80)), 835-841.
Jayalakshmi, S., & Aleesha, A. (2024). STUDY ON ACCEPTANCE SAMPLING PLAN BASED ON PERCENTILES FOR EXPONENTIATED GENERALIZED INVERSE RAYLEIGH DISTRIBUTION. *Reliability: Theory & Applications, 19*(2 (78)), 202-208.
Jiang, R., Xue, W., & Cao, Y. (2024). *Five 2-parameter variants of folded-normal distribution and their application in reliability modeling.* Paper presented at the 2024 6th International Conference on System Reliability and Safety Engineering (SRSE).
Lio, Y., Tsai, T.-R., & Wu, S.-J. (2009). Acceptance sampling plans from truncated life tests based on the Birnbaum–Saunders distribution for percentiles. *Communications in Statistics-Simulation and Computation, 39*(1), 119-136.
Lu, X., Gui, W., & Yan, J. (2013). Acceptance sampling plans for half-normal distribution under truncated life tests. *American Journal of Mathematical and Management Sciences, 32*(2), 133-144.
Naghizadeh, M., & Mahdizadeh, Z. (2026). Improved time- censored group sampling schemes based on generalized half-normal distribution. *Caspian Journal of Mathematical Sciences*.
Naveed, M., Iltaf, M., Azam, M., Khan, N., Saeed, M., & Akbar, M. Z. (2025). Repetitive Sampling Plans for Life Tests Based on Percentiles of the Half-Normal Distribution with Applications to Software Reliability and Device Lifetime Data. *SCOPUA Journal of Applied Statistical Research, 1*(3).
Qomi, M. N., Piadeh, M., Pérez-González, C. J., & Fernández, A. J. (2025). Optimal acceptance sampling plans based on generalized half-normal distribution under time-censoring with conventional and expected limited risks. *Communications in Statistics-Simulation and Computation*, 1-20.
Rao, B. S., Kumar, C. S., & Rosaiah, K. (2013). Acceptance sampling plans from life tests based on percentiles of half normal distribution. *Journal of Quality and Reliability Engineering, 2013*(1), 302469.
Rao, B. S., Prasad, J. S. R., Rao, B. V. P., & Sudhakar, M. (2026). Application of Half Normal Distribution in Hybrid Group Adoption Sampling Plan for Life Tests. *Thailand Statistician, 24*(1), 26-33.
Rao, B. S., Priya, M. C., & Kantam, R. (2014). Acceptance sampling plans for percentiles assuming the linear failure rate distribution. *Stochastics and Quality Control, 29*(1), 1.
Rao, G. S. (2013). Acceptance sampling plans from truncated life tests based on the Marshall–Olkin extended exponential distribution for percentiles.
Rao, G. S., Jilani, S., & Peter, J. K. (2025). Designing of multiple dependent state sampling plan for Exponentiated Frechet Distribution. *Research in Statistics, 3*(1), 2531810.
Rao, G. S., & Kantam, R. (2010). Acceptance sampling plans from truncated life tests based on the log-logistic distributions for percentiles.
Schilling, E. G., & Neubauer, D. V. (2009). *Acceptance sampling in quality control*: Chapman and Hall/CRC.
Tripathi, H. (2025). Acceptance sampling inspection plan under fuzzy theory for generalized half-normal distribution. *Life Cycle Reliability and Safety Engineering*, 1-11.
Uke, T., & Taye, A. (2025). Alpha Power Transformed Half Normal Distribution with Its Properties and Applications. *East African Journal of Biophysical and Computational Sciences, 6*(2), 25-37.
Wang, T.-C., Shu, M.-H., Hsu, B.-M., & Hsu, C.-W. (2022). Adjustable variables multiple-dependent-state sampling plans based on a process capability index. *Journal of the Operational Research Society, 73*(12), 2626-2639.
Wortham, A., & Baker, R. (1976). Multiple deferred state sampling inspection. *The International Journal of Production Research, 14*(6), 719-731.
Wu, C.-W., Shu, M.-H., & Wang, T.-C. (2024). Designing a variables two-plan sampling system with adjustable acceptance criteria for lot disposition. *Quality Engineering, 36*(3), 521-533.
Wu, C.-W., Shu, M.-H., & Wang, T.-C. (2025). Optimal design of variables switch-based sampling scheme for verifying Weibull distributed product lifetimes. *Journal of Applied Statistics, 52*(1), 119-134.
Zheng, H., Yang, J., & Zhao, Y. (2025). Accelerated acceptance sampling plan for degraded products based on inverse Gaussian process considering the acceleration factor uncertainty. *Computers & Industrial Engineering, 203*, 111052.

## Declaration

**Conflict of Study:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Funding Statement:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author Contribution Statement**: All the authors worked together as a cell group and designed the research; Analyzed and interpreted the data, and wrote the paper.

**Availability of Data and Material**: Data will be made available by the corresponding author on reasonable request.

## Author(s) Bio / Authors' Note

***Muhammad Naveed*** *is an Assistant Professor in the Higher Education Department, Punjab, Pakistan. His research interests include statistical quality control and related areas. Email:* mnaveedakbar09@gmail.com

***Muhammad Iltaf*** *earned his MPhil in Statistics from the National College of Business Administration and Economics, Lahore. His area of interest is statistical quality control. Email:* ranaaltaf123@gmail.com

***Muhammad Azam*** *earned his Master's degree in Statistics from the Islamia University of Bahawalpur (Pakistan) in 1996, graduating with distinction as a Gold Medalist. He completed his MPhil in Statistics at Quaid-i-Azam University, Islamabad, in 2006, and his PhD at the University of Innsbruck, Austria, in 2010. With over 28 years of teaching and research experience, Dr. Azam began his academic career as a Lecturer in Statistics with the Punjab Education Department, where he served for 13 years. In 2010, he joined Forman Christian College (A Chartered University), Lahore, as an Assistant Professor, and in 2015, he moved to the University of Veterinary and Animal Sciences (UVAS), Lahore, as an Associate Professor. He was promoted to Professor of Statistics in January 2018. He has twice served as Dean, Faculty of Life Sciences Business Management, UVAS Lahore—from March 2018 to March 2021, and again from June 2022 to June 2025—showcasing his leadership in academic administration. Dr. Azam has published over 110 research papers, primarily in high impact international journals. His research interests include survey sampling, statistical quality control, decision trees, and ensemble classifiers. He has successfully supervised four PhD and twenty-six MPhil scholars, contributing significantly to the advancement of statistical research and education. Email:* mazam@uvas.edu.pk

***Nasrullah Khan*** *is an Associate Professor in the Department of Statistical Sciences at the University of the Punjab, Lahore, Pakistan. His research interests span statistical modeling, machine learning, climate data analysis, and quality control. He has contributed in methodological and applied statistics, focusing on the integration of classical and modern data-driven approaches for real-world problem solving. Email:* nasrullah.stat@pu.edu.pk