STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–17.
Published online in International Academic Press (www.IAPress.org)

The Robust Linear Estimation of Logistic Distribution Parameters in Lifetime Data

Mohammed Abduljabar Ibrahim Hasawy, Taha Hussein Ali*, Bekhal Samad Sedeeq

Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University, Erbil, Iraq

Abstract Outliers or non-normality in lifetime data affect the accuracy of estimated parameters of mathematical distributions, including the logistic distribution, so robust estimators such as the rank regression method are usually used to be less sensitive to outliers or non-normality. In this article, some robust estimation approaches are employed in rank regression estimation to be more robust to outliers. Proposed methods included the robust covariance between the transformed (dependent variable) and the original time values (independent variable) for the Fast Minimum Covariance Determinant method and robust linear regression for weight functions (Bi-square, Andrews, Cauchy, Fair, Huber, Logistic, Talwar, and Welsch) in estimation regression coefficients which can be employed in rank regression to estimate logistic distribution parameters. The efficiency of the traditional method (Rank Regression Method) and the proposed method was compared based on simulation data and real data representing the lifetime of Bacterial Toxemia children's patients at Hefei Children's Teaching Hospital. The study's results showed the efficiency of the proposed methods in handling outliers and the accuracy of estimating the parameters (Location and Scale) of the logistic distribution compared to the traditional method.

Keywords Lifetime Data, Logistic Distribution, Maximum Likelihood Estimation, Outliers, Robust Estimation

DOI: 10.19139/soic-2310-5070-2971

1. Introduction

Here, we introduce a new parameter estimation procedure that takes advantage of rank regression and its robustness against outliers in the setting of lifetime data for estimating the parameters of the two-parameter logistic distribution. Lifetime distributions are fundamental and widely used in survival analysis, reliability engineering, and risk management, where they allow researchers to estimate how long it will take for an event of interest to happen, such as the failure of a mechanical unit or the death of a living being [1]. The logistic distribution is specifically used as it approximates well the normal distribution and allows for a flexible interpretation of the time-to-event. But conventional methods like maximum likelihood estimators (MLE) for obtaining the parameters of these distributions are severely limited in the presence of outlying observations, giving unreliable and biased parameter estimates [2]. On the other hand, the rank regression approach transforms the data into robust centers based on ranks, thus being more resistant to outliers and providing more accurate estimates of the parameters.

This work presents a practical application of the method within the context of the two-parameter logistic model, utilizing lifetime data, and assesses the method's performance in comparison with traditional methods using both simulated and real data. It also sheds light on how outlier management is critical to maintain the quality of estimations and predictions in engineering and medicine.

(Hasan S. Yavuz & İsmail Alayoğlu 2024) Railway maintenance and efficient operation have been important issues for safe rail traffic. When an unexpected malfunction occurs in various components of the train or the

*Correspondence to: Taha Hussein Ali (Email: taha.ali@su.edu.krd). Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University, Erbil, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

railway system, it may result in unscheduled maintenance, which may cause the rail traffic to stop [3]. In this paper, we study the random failure model of some frequently malfunctioning high-speed railway equipment based on the statistical analysis of the real data of failure records in Turkey. Popular distribution functions and parameter estimation methods have been used, considering that the data has a small sample size, and it may contain outliers. In this study, we showed that for the case of a few numbers of failure data, the L-moments method gives effective results when there exists no outlier, and the robust-M method gives effective results when there exists an outlier or outliers [1, 4].

Recent research has highlighted the importance of robust estimation techniques for lifetime and reliability data, particularly in the presence of outliers and small sample sizes. For instance, Yavuz and Alayolu (2024) demonstrated that robust methods outperform classical estimators in railway failure data when contamination is present [3]. In a related context, Ali et al. (2021) proposed an improved non-parametric estimation framework based on universal thresholding and wavelet techniques, showing significant gains in estimation accuracy through extensive simulation studies and real data applications [5]. These contributions collectively emphasize the growing demand for robust and efficient estimation procedures in statistical modeling and reliability analysis. Building on this line of research, the present study integrates FMCD-based robust covariance estimation with rank regression to develop a unified framework for estimating the parameters of the two-parameter logistic distribution.

2. Lifetime data

Age is a commonly used time measure, with baseline age representing the starting time, and individuals existing at the event age or the observation age. Models that use age as a time measure can be adjusted for calendar effects. Some researchers recommend using age, rather than study time, as it may provide less biased estimates. A particular challenge with survival analysis is that only some individuals will experience the event by the end of the study, and therefore, survival times will be unknown for a subset of the study population. This phenomenon is called ”censoring” and can arise in the following ways: the study participant does not experience the relevant outcome, such as relapse or death, by the end of the study; the study participant is lost to follow-up during the study period; or they experience a different event that makes follow-up impossible. These interpolated time intervals underestimate the unknown true time to the event. In most analytical approaches, censoring is assumed to be random or uninformative. There are three main types of censoring: right, left, and interval. If events occur after the end of the study, data are right-censored. Data is censored to the left when an event is observed, but its exact time is unknown. Data is censored to the interval when an event is observed, but participants enter and exit observations, so the exact time of the event is unknown. Most survival analysis methods are designed for right-hand observations, but methods are available for interval and left-hand data. In quality-of-life research, one might be interested in predicting how long an individual will live after a serious medical diagnosis, such as Bacterial Toxemia. Typically, these patients experience an increased mortality rate initially and then a decreased mortality rate later. Statistically, this research question can be answered using event history analysis (also called hazard or survival analysis in biostatistics, duration models in economics, and reliability analysis in engineering). A hazard rate generally describes the conditional probability of an event occurring in each period [3].

3. Logistic distribution

Logistic distribution, also known as the logistic function or logistic curve, is a fundamental concept in statistics and data analysis that describes the probability distribution of a continuous random variable following a logistic function. This distribution is particularly useful in biology, economics, and machine learning, where it models growth processes and binary outcomes. The logistic distribution is characterized by a symmetrical S curve about the mean. It consists of two parameters: the mean and the scale parameter s. The center of the distribution is determined by the mean, and the slope of the curve is determined by the scale parameter. The tails of the logistic distribution are heavier than those of the normal distribution and can thus be used to model extreme data.

The probability density function (PDF) and the cumulative distribution function (CDF) of the logistic distribution are expressed, respectively, as [7]:

The PDF of the Logistic distribution is given by:

$$f(x) = \frac{exp^{-(x-\mu)/s}}{s(1+exp^{-(x-\mu)/s})^2} \tag{1}$$

Where: μ is the location parameter (mean), s is the scale parameter (related to the standard deviation), e is the base of the natural logarithm, x is the variable. Cumulative Distribution Function (CDF): The CDF of the Logistic distribution is:

$$F(x) = \frac{1}{1+exp^{-(x-\mu)/s}} \tag{2}$$

Where x is the value of the random variable, share the scale and μ shape parameters, respectively [8].

Survival Function $S(t)$, and Hazard Function: The survival function is a function that gives the probability that a patient, device, or other object of interest will survive past a certain time. The survival function is also known as the survivor function or reliability function.

Let the lifetime T be a continuous random variable with cumulative hazard function $F(t)$ and hazard function $f(t)$ on the interval $[0, \infty)$. Its survival function or reliability function is [9].

$$S(t) = P(T > t) = \int f(u)\, du = 1 - F(t) \tag{3}$$

4. Detect outliers

Outlier detection is a crucial tool used to improve estimates. By identifying and effectively addressing outliers, researchers can better ensure that their research is more reliable and shows the patterns and trends in their data. We will discuss some statistical methods that can be employed to detect outliers in this section. Such approaches can allow for better-informed research decisions and better overall reliability [10]. These methods include:

4.1. Outliers based on Z-score

The standard score measures the number of standard deviations from the mean of a value. The formula for the Z-score is as follows:

The formula for the Z-score is as follows:

$$Zscore = (x - mean)/std.deviation \tag{4}$$

If the z-score of a data point is more than 3, it indicates that the data point is quite different from the other data points. Such a data point can be an outlier.

A Z-score can be both positive and negative. The farther away from 0, the higher the chance of a given data point being an outlier. Typically, a Z-score greater than 3 is considered extreme [11].

4.2. Outliers based on residuals

This is done by identifying outliers, removing them from the observations, and then using the standard estimates. For example, if we plot the E_{is} (standard residuals) against Yi, outliers can be noticed graphically. Points outside 2 are outliers. These outliers can also be detected by statistical tests, such as [12, 21].

$$E_{is} = \frac{e_i}{\sqrt{MSE}} \tag{5}$$

$i = 1, 2, 3, \ldots, n$; e_i represents the residuals, MSE: mean square error, and Y_i: Outliers

4.3. Interquartile range

We could use the IMR method of identifying outliers to create a "fence" outside Q1 and Q3. Any values that do not fall into this fence are considered outliers. To build this fence, we take 1.5 times the IQR, and then subtract it from Q1 and subtract it from Q3 to make this fence. We now have the minimum and maximum fence posts that we compare to each observation. Outliers are observations with an IQR greater than 1.5 Q1 or more than 1.5 Q3. Minitab defaults to using this to find outliers.

5. Estimation

Parameter estimation is the identification of parameters of a probability distribution that are most representative of the observed data. The distribution determines which estimation method to employ. We will be specifically interested in the estimation process for the parameters of the two-parameter logistic model that is used in this work [13].

5.1. Maximum likelihood estimation

Maximum likelihood is a widely used technique for estimation with applications in many areas, including time series modeling, panel data, discrete data, and even machine learning. Maximum likelihood estimation is a statistical method for estimating the parameters of a model [22, 23]. In maximum likelihood estimation, the parameters are chosen to maximize the likelihood that the assumed model results in the observed data. The equations for the partial derivatives of the log-likelihood function are derived and given next [14].

$$\frac{\partial \Lambda}{\partial \mu} = \sum_{i=1}^{n} ln\left(\frac{x_i - \mu}{s^2}\right) - 2\sum_{i=1}^{n} \frac{e^{-(xi-\mu)/s}}{s(1+e^{-(xi-\mu)/s})}$$

and

$$\frac{\partial \Lambda}{\partial s} = \frac{-n}{s} + \sum_{i=1}^{n} \frac{x_i - \mu}{s^2} - \sum_{i=1}^{n} \frac{e^{-(xi-\mu)/s}}{s^2(1+e^{-(xi-\mu)/s})} + 2\sum_{i=1}^{n} \frac{(xi-\mu)e^{-(xi-\mu)/s}}{s^2(1+e^{-(xi-\mu)/s})^2} = 0$$

Solving the above equations simultaneously, we obtain the estimated parameter values.

5.2. Rank regression

Conducting rank regression (RRY) involves mathematically fitting a straight line to a collection of data points in a manner that minimizes the sum of the squares of the vertical deviations from the points to the line [15]. This procedure is similar to the probability plotting technique. It does, but use the least squares principle to determine the line through the points rather than visual estimation. The first step is to convert our function into a linear distribution for the two-parameter distribution [14].

The form of the Logistic probability paper is based on linearizing the CDF. From the unreliability equation, z can be calculated as a function of the CDF F as follows

$$z = ln(F) - (1 - lnF) \tag{6}$$

or using the equation for z

$$\frac{t-\mu}{s} = ln(F) - (1 - lnF) \tag{7}$$

Then

$$ln(F) - (1 - lnF) = -\frac{\mu}{s} + \frac{1}{s}t \tag{8}$$

Now let

$$y = ln(F) - (1 - lnF) \tag{9}$$

$x = t, a = -\frac{\mu}{s}$, and $b = \frac{1}{s}$.
Which results in the following linear equation:

$$y = a + bx \tag{10}$$

The technique of least squares parameter estimation, which is sometimes referred to as regression analysis, was covered in the article on parameter estimation [16], and the equations for regression on Y were constructed as follows:

$$\hat{a} = \bar{y} - \widehat{b}\bar{x} \tag{11}$$

And

$$\hat{b} = \hat{\rho}\sqrt{\frac{S_y}{S_x}} \tag{12}$$

$$\hat{\rho} = \frac{S_{xy}}{\sqrt{S_x \times S_y}}; S = \begin{pmatrix} S_x & S_{xy} \\ S_{yx} & S_y \end{pmatrix} \tag{13}$$

The sample variances are located on the main diagonal of matrix S, whereas the sample covariances are situated on the **off-diagonal of matrix S. In this case, the equations for y_i and x_i are:**

$$y_i = lnF\left(xi\right) - \left(1 - lnF\left(x_i\right)\right) \tag{14}$$

and

$$x_i = ln\left(x_i\right) \tag{15}$$

The $F(xi)$ values are derived from the median ranks (MR). The Median Ranks approach is used to quantify the unreliability associated with each failure. The median rank represents the genuine probability of failure, $Q_{(xj)}$, at the jth failure among a sample of N units, corresponding to the 50% confidence level [17]. An alternative and simpler approach to determining median ranks involves applying two transformations to the cumulative binomial equation, namely the Beta and F Distribution Approach [18], first to the beta distribution and then to the F distribution.

$$MR = \frac{1}{1 + \left(\left(N - j + 1\right)/j\right) F_{0.50;m;n}} \tag{16}$$

Where $m = 2(N - j + 1)$ and $n = 2j$. $F_{0.50;m;n}$ is the F distribution for a 0.50 point, (m and n degrees of freedom), for failure j from N units.

$$\hat{a} = -\frac{\mu}{s} \qquad \hat{b} = \frac{1}{s}$$

5.3. Robust estimation

Parameter estimation is used in most statistical applications for best practice with the exception of maximum likelihood estimation or least squares regression. These approaches assume that the data presented are of strict quality, such as independent observations, homogeneity of variance, and approximate normality of error terms. Yet such assumptions are rarely made in practice. Real datasets can be erratic in some instances, such as extreme values, recording errors, or really abnormal observations. Even when there are very small outliers, they can have a large influence on the results, leading to biased estimators, too high variances, and false statistical inferences.

As a response to these challenges, solid estimation methods have been developed. Robustness has the central concept of maintaining stable and reliable parameters even in the face of a data deviancy. However, robust procedures mitigate their effects by employing reweighting schemes, influence functions, or alternative measures of central tendency and dispersion, rather than allowing a few anomalous values to dominate the estimation procedure. This makes the estimates more accurately reflect the data structure and prevent contamination [19].

Especially robust methods will be useful when the analysis of lifetime data is concerned. For instance, errors in measurement, incomplete follow-up, or rare cases where variance is not consistent with model assumptions are

often the cause of survival times. This robust estimation not only enables more precise parameter estimations but also ensures better accuracy of subsequent reliability and risk assessments. For these reasons, robust estimation is used for the two-parameter logistic distribution in the present paper using rank regression methods. The next subsections provide a detailed description of two prominent approaches employed in this work: the Fast Minimum Covariance Determinant (FMCD) method and robust linear regression based on different weighting functions [20].

5.3.1. Fast minimum covariance determinant (FMCD) The minimum covariance determinant (MCD) estimator is one of the first highly robust and convergent estimators for multivariate location and dispersion. It is a particularly useful tool to detect outliers because the MCD is highly resistant to extreme observations. It was first developed in 1984, but has been widely used since the efficient Fast-MCD algorithm was introduced in 1999 [25]. Since that time, MCD has been applied to medicine, finance, image analysis, and chemistry. Also, MCD has been used to develop many powerful multivariate techniques such as robust principal component analysis, factor analysis, and multiple regression.

5.3.2. Robust linear regression A robust Rank Regression Estimation is proposed to incorporate the rank regression coefficient estimation approach, where the scale and location parameters of the logistic distribution are estimated. It employs the Rousseuw and Leroy algorithm, the FMCD (Fast Minimum Variance Determinant). This method takes i observations from n (where $n/2 < i <= n$), which includes the least determinant of the variance-covariance matrix. To obtain consistency in the multivariate normal distribution (MND) and to account for bias in small samples, the estimate is the variance-covariance matrix of the i points specified above multiplied by the consistency value and the correction value. On this basis, we obtain the estimates of the mean of the independent and dependent variables (RMx and RMy) and the variance-covariance matrix (RS) resistant to extreme values as follows [26, 27]:

$$RS = \begin{pmatrix} RS_x & RS_{xy} \\ RS_{yx} & RS_y \end{pmatrix} \tag{17}$$

That is, the sample robust variances on the main diagonal of the matrix RS, and sample robust covariances on the off-diagonal of matrix RS; therefore, the robust correlation is [28]:

$$r\hat{\rho} = \frac{RS_{xy}}{\sqrt{RS_x \times RS_y}} \tag{18}$$

Regression coefficients ($r\hat{a}$ and $r\hat{b}$) are computed as [29],

$$r\hat{a} = RM_y - r\hat{b}RM_x \tag{19}$$

$$r\hat{b} = r\hat{\rho}\sqrt{\frac{RS_y}{RS_x}} \tag{20}$$

In our case, the equations for y and x are:

$$y = Ln\left[1 - F\left(t\right)\right] \ \& \ x = t; \tag{21}$$

The $F(t)$ is estimated from the median ranks (Appendix). Once ($r\hat{a}$ and $r\hat{b}$) are obtained, then $\hat{\lambda}$ and $\hat{\gamma}$ for two parameters logistic distribution can easily be obtained from the following equations.

$$\hat{\lambda} = -r\hat{b} \; and \; \hat{\gamma} = r\frac{\hat{a}}{\hat{\lambda}} \tag{22}$$

6. Proposed method

This study proposes a unified robust rank regression framework for estimating the location and scale parameters of the logistic distribution in the presence of outliers. Let the lifetime observations be arranged in ascending order

as $T = t_1, t_2, \ldots, t_n$, where $t_1 \leq t_2 \leq \cdots \leq t_n$. The empirical cumulative distribution function (CDF), used as an approximation to the true CDF, is obtained from the order statistics according to

$$F(t_i) = \frac{1}{1 + \left(\frac{n-i+1}{i}\right) f(i)} \tag{23}$$

where $f(i)$ is determined from the inverse F-distribution. This formulation exploits the relationship between order statistics and the CDF to improve the fit of the assumed distribution [30].

To facilitate regression-based estimation, the logit transformation of the empirical CDF is applied:

$$y_i = \ln\left(\frac{F(t_i)}{1 - F(t_i)}\right) \tag{24}$$

This transformation linearizes the logistic distribution and yields an approximately linear relationship between the transformed response variable y_i and the explanatory variable x_i.

Parameter estimation, then, is done using robust regression methods to reduce outlier influence. In general, the estimate is based on reducing a weighted least squares objective function of the form.

$$\min \sum_{i=1}^{n} w(e_i)[y_i - (a + bx_i)]2,$$

where $e_i = y_i - (a + bx_i)$ denotes the residual and $w(\cdot)$ is a weight function that assigns lower weights to extreme observations.

There are nine robust estimation techniques in this framework, namely: FMCD, Bi-square, Andrews, Cauchy, Fair, Huber, Logistic, Talwar, and Welsch. The FMCD method has strong estimates of the covariance structure because the determinant of the covariance matrix from a small percentage of central observations is tiny; the FMCD is resistant to outliers. All other methods fall under the M-estimators class and only differ in the selection of the weight function and tuning constant, which affects the balance between robustness and efficiency

All robust procedures use the regression coefficients from the rank regression model to estimate the parameters of the logistic distribution in one place. In particular, the scale parameter is estimated as follows:

$$\hat{\sigma} = \frac{1}{\hat{b}} \tag{25}$$

and the location parameter is obtained as

$$\hat{\mu} = -\hat{a}\hat{\sigma} \tag{26}$$

The same formulation is used for all the robust estimation methods to ensure consistency, and each estimation method can be weighted to correct for data contamination. Consequently, the proposed approach provides a flexible and reliable procedure for estimating the parameters of the logistic distribution in the presence of outliers.

The proposed estimators are based on well-established robust rank-regression and M-estimation frameworks. Consequently, their asymptotic properties, including consistency and asymptotic normality, follow from standard results in robust regression theory, while robustness characteristics such as influence functions and breakdown points are inherited from the corresponding M-estimators and FMCD covariance estimators. This study does not consider censored observations and focuses on fully observed data. Extending the proposed robust rank-based estimators to right-censored settings, for example via inverse-probability weighting, is an interesting direction for future research.

7. Simulation study

The simulation could be useful for testing outlier filtering or denoising algorithms, especially when applied to lifetime data. By artificially adding outliers to specific points, you can test how well different methods handle

outliers in a logistic distribution setting and comparison of the efficiency of the estimated parameters (Location and Scale) for the logistic distribution using classical and proposed methods.

Generate (100, 200, and 300) data points from a logistic distribution with a mean of (10 and 15) representing location parameters and a standard deviation of (2 and 5) representing scale parameters. Add outliers to two, four, and six randomly chosen data points by amplifying them by a factor of 5 and shifting them by 5 units for the sample sizes of 100, 200, and 300, respectively.

This study focuses on robustness against outlying observations rather than explicit contamination models. In the simulation design, outliers are introduced as atypical observations without preliminary data-dependent screening, allowing the robust estimators to operate directly on the full sample. In the empirical analysis, Z-score and IQR criteria are used solely for descriptive purposes and do not affect the estimation procedure.

Figure 1 illustrates the first simulation experiment (n = 100) with location (μ = 10) and scale (σ = 2). The artificially introduced outliers are highlighted for visualization purposes. Z-score values and residual magnitudes are reported solely to illustrate the extremeness of these observations and are not used as part of any outlier detection or filtering procedure in the simulation study.

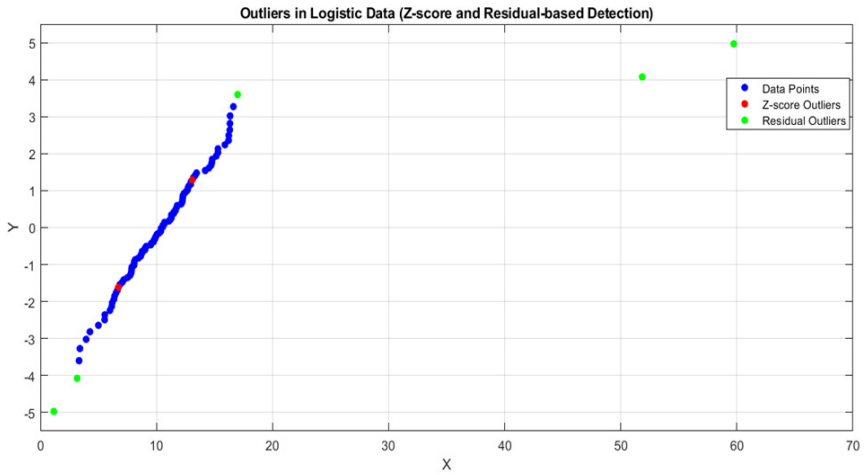

Figure 1. Identified Outliers for the First Experiment Simulation ($n = 100$, $\mu = 10$, and $\sigma = 2$).

Figure 1 shows that blue points represent the original data (x, y), showing the general distribution of data points. Red points represent outliers detected using the Z-score method. Green points represent outliers detected using the residual-based method. The number of outliers based on the Z-score is equal to 2, and the number of outliers based on residuals is equal to 5.

This experiment was repeated (1000) times and the average values from the location (Mean) & scale (standard deviation) parameters logistic distribution were estimated using the traditional (MLE and RRY) and proposed method (RRRY, Bi-square, Andrews, Cauchy, Fair, Huber, Logistic, Talwar, and Welsch), with the calculation of the MSE average for the two parameters shown in Table 1.

Table 1 summarizes the results of a simulation study for different estimation methods used in parameter estimation, specifically focusing on the location and scale estimates and MSE.

Location refers to the estimate of the mean (μ), which is expected to be 10 based on the simulation parameters. The location estimates are close to the true value of 10, with small variations across methods.

There is a significantly larger estimate of location than the RRY method (μ = 10.9052), which suggests that this method is less accurate in estimating the location here. RRRY, Bi-square, Andrews, and Talwar give estimates that are close to the true value of 10. The estimated location using the MLE method is 10.1338, close to the true value.

Table 1. Results of the Simulation Study (n = 100, μ = 10, and σ = 2)

Method	Location	Scale	MSE (Location)	MSE (Scale)
MLE	10.1338	2.6163	0.1419	0.4371
RRY	10.9052	5.8392	1.0095	17.6810
RRRY	10.0822	2.0359	0.1406	0.0696
Bi-square	10.1228	2.1055	0.1412	0.0471
Andrews	10.1226	2.1053	0.1411	0.0471
Cauchy	10.1350	2.1291	0.1450	0.0499
Fair	10.1928	2.4440	0.1654	0.3729
Huber	10.1614	2.2675	0.1536	0.1096
Logistic	10.1631	2.2807	0.1539	0.1179
Talwar	10.1284	2.1157	0.1446	0.0493
Welsch	10.1256	2.1088	0.1422	0.0470

The RRY method (σ = 5.8392) has a much larger scale estimate, suggesting that this method has struggled with data variability or the scale has been overestimated. The RRRY methodology (2.0359), Bi-square (2.1055), Andrews (2.1053), and Cauchy (2.1291) are more or less identical to the scale estimates but do not perform well at the close of 2/2. MLE estimates the scale to be 2.6163, which is somewhat larger than the true value. Fair (2.4440) and Huber (2.2675) give slightly larger scale estimates but are also reasonable over the true value 2.

Location MSE refers to the distance from the estimated location parameter to the true value (10). Smaller MSE values indicate more accurate location estimation. The MSE (Location) of the classical RRY method is 1.0095, which is less than satisfactory for estimating location (mean). MLE (0.1419), RRRY (0.1406), and Bi-square (0.1412) methods have low MSE and thus provide accurate location estimations. The LOCATION values for Cauchy (0.1454) and Fair (0.1644) are slightly higher, but still small.

Scale MSE refers to the relative proximity of the estimated scale to the true value (2). Again, smaller values indicate better accuracy. The RRY method has the highest MSE of 17.6810, indicating that the scaling parameter is not accurate. MSE for RRRY (0.0696), Bi-square (0.0471), Andrews (0.0471), Welsch (0470), and Talwar (0493) is low, meaning they estimate the scale accurately. The MLE method is relatively successful at the MSE (Scale) of 0.4371, which is good, but not quite as good as the low-MSE methods.

For this case, RRRY, Bi-square, and Andrews's methods perform well in both location and scale estimation, with low MSEs for both parameters. MLE is a good method for estimating the location, but it has a relatively higher scale MSE compared to others. The RRY method stands out as the least effective method for both location and scale estimation, with high MSE values for both parameters. The proposed method (Andrews) was the best compared to the rest of the methods, so the Probability density function (PDF) for logistic distribution, cumulative distribution function (CDF), Survival function $S(t)$, and Hazard function $h(t)$ will be explained in Figure 2: Figure 2 explanation of what to expect in terms of behaviour and insights from each plot:

1. Probability Density Function (PDF), the logistic distribution has a symmetric, bell-shaped curve centered around the location parameter of (10.1226). The scale parameter is (2.1053), which controls the spread of the curve. The PDF will show the density of probability around the location μ. As you move away from μ, the density decreases, with values closer to μ having higher density. The PDF typically has tails that extend infinitely in both directions, but due to the nature of the logistic distribution, the tails will taper off rather than approach zero too rapidly. Since the scale parameter is 2.1053, notice that the peak of the PDF is at Location =10.1226, and it starts decreasing gradually in both directions from this point.
2. Cumulative Distribution Function (CDF) is a sigmoidal (shape) curve that starts at 0 and asymptotically approaches 1 as t increases. The CDF gives the cumulative probability of the logistic distribution, i.e., the probability that a random variable T is less than or equal to some value t. The steepest part of the curve occurs at the location parameter μ, and this is where the probability increases most rapidly. For values of t much smaller than μ, the CDF will be close to 0, and for values much larger than μ, it will approach 1. At $t = \mu$, the CDF will be around 0.5, which is the midpoint of the S-curve.

3. Survival Function $(S(t))$ is the complement of the CDF, so it will start at 1 and gradually decrease to 0 as t increases. The survival function S(t) represents the probability that a random variable exceeds the value t. The survival function decreases from 1 (i.e., the probability that $T > t$ is 100% for small t) to 0 (i.e., the probability that $T > t$ is 0 for large t).
4. Hazard Function $(h(t))$ will be a ratio of the PDF to the survival function. Typically, for the logistic distribution, the hazard function will display a "bell-shaped" curve. The hazard function tells us the instantaneous failure rate or risk of an event occurring at a specific value of t, given that it hasn't occurred before. Near $t = \mu$, the hazard function will have its highest point, which reflects the peak rate of events (or failures) happening around the location parameter. The hazard function should decrease as t moves away from μ, both towards the left and right extremes.

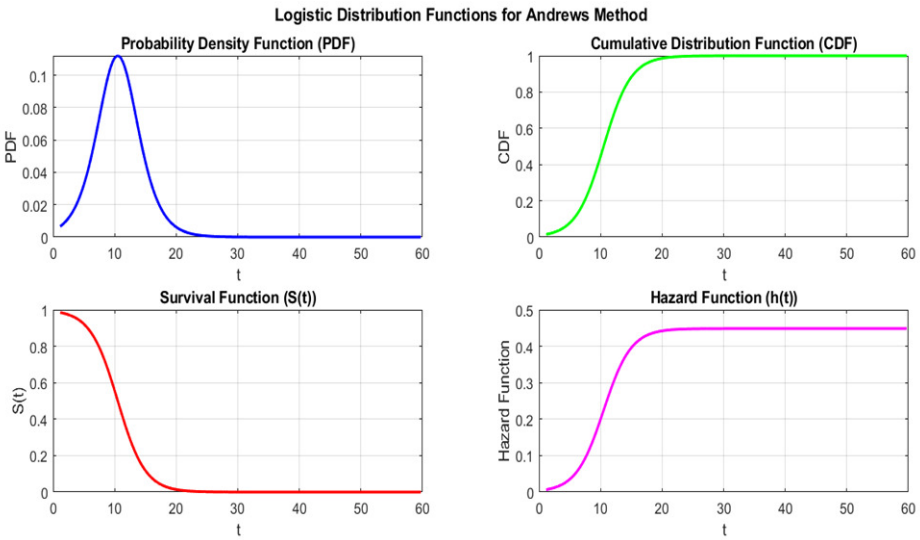

Figure 2. Logistic distribution functions (Andrews method) for simulation study ($n = 100$, $\mu = 10$, and $\sigma = 2$).

This experiment was repeated (1000) times, and the average values from the location & scale parameters of the logistic distribution were estimated using the traditional and proposed methods, with the calculation of the MSE average for the two parameters shown in Table 2.

Table 2. Results of the simulation study (n = 200, μ = 10, and σ = 2)

Method	Location	Scale	MSE (Location)	MSE (Scale)
MLE	10.1445	2.6202	0.0829	0.4145
RRY	10.9105	5.7023	0.9228	15.1402
RRRY	10.0955	2.0467	0.0766	0.0245
Bi-square	10.1294	2.0910	0.0802	0.0256
Andrews	10.1292	2.0909	0.0801	0.0256
Cauchy	10.1422	2.1097	0.0835	0.0283
Fair	10.2056	2.4179	0.1071	0.2421
Huber	10.1697	2.2359	0.0929	0.0752
Logistic	10.1718	2.2484	0.0933	0.0820
Talwar	10.1333	2.0956	0.0815	0.0265
Welsch	10.1324	2.0934	0.0809	0.0257

Table 2 summarizes the results of an MSE for the location. The RRY classical method has the highest MSE of 0.9228, indicating that it performed poorly in estimating the location. MLE (0.0829), RRRY (0.0766), Bi-square (0.0802), Andrews (0.0801), Talwar (0.0815), and Welsch (0.0809) methods show low MSE, indicating good accuracy in estimating the location. The MSE (Location) values for methods like Cauchy (0.0835) and Fair (0.1071) are slightly higher, but still relatively small.

MSE for scale, the RRY method has the largest MSE of 15.1402, indicating poor performance in estimating the scale parameter. RRRY (0.0245), Bi-square (0.0256), Andrews (0.0256), Cauchy (0.0283), Welsch (0.0257), and Talwar (0.0265) methods show low MSE, indicating high accuracy in estimating the scale. The MLE method has a moderate MSE of 0.4145, indicating decent performance but not as good as the low-MSE methods.

For this case, RRRY, Bi-square, and Andrews's methods perform well in both location and scale estimation, with low MSEs for both parameters. MLE is a good method for estimating the location, but it has a relatively higher scale MSE compared to others. The RRY method stands out as the least effective method for both location and scale estimation, with high MSE values for both parameters. The proposed method (RRRY) was the best compared to the rest of the methods.

This experiment was repeated (1000) times, and the average values from the location & scale parameters logistic distribution were estimated using the traditional and proposed methods, with the calculation of the MSE average for the two parameters shown in Table 3.

Table 3. Results of the simulation study (n = 300, μ = 10, and σ = 2)

Method	Location	Scale	MSE (Location)	MSE (Scale)
MLE	10.1300	2.6199	0.0585	0.4044
RRY	10.8907	5.6493	0.8599	14.2459
RRRY	10.0790	2.0467	0.0493	0.0143
Bi-square	10.1122	2.0835	0.0548	0.0193
Andrews	10.1119	2.0833	0.0547	0.0193
Cauchy	10.1259	2.1016	0.0587	0.0215
Fair	10.1905	2.3938	0.0813	0.2048
Huber	10.1551	2.2195	0.0681	0.0614
Logistic	10.1569	2.2312	0.0685	0.0673
Talwar	10.1164	2.0881	0.0566	0.0201
Welsch	10.1151	2.0859	0.0557	0.0194

Table 3 shows that Location estimates, MLE has a location estimate of 10.1300, which is very close to the true value. The RRY has a notably higher location estimate of 10.8907, which is further from the true value, resulting in a high MSE (0.8599). Proposed methods show good performance in estimating the location, with estimates near 10.

Scale estimates, MLE provides a scale estimate of 2.6199, which is higher than the true value. The RRY produces the worst result with a scale estimate of 5.6493 and an MSE of 14.2459, indicating significant bias. Proposed methods provide scale estimates closer to 2, with low MSE values (around 0.02-0.04). The proposed method (RRRY) was the best compared to the rest of the methods.

This experiment was repeated (1000) times, and the average values from the location & scale parameters logistic distribution were estimated using the traditional and proposed methods, with the calculation of the MSE average for the two parameters shown in Table 4.

Table 4. Results of the simulation study (n = 100, μ = 15, and σ = 5)

Method	Location	Scale	MSE (Location)	MSE (Scale)
MLE	15.2773	5.9840	0.8884	1.1035
RRY	16.3131	9.5093	2.9119	31.2473
RRRY	15.1814	5.0813	0.8542	0.2317
Bi-square	15.2729	5.2554	0.8698	0.2879
Andrews	15.2723	5.2545	0.8691	0.2884
Cauchy	15.3008	5.3173	0.8916	0.3077
Fair	15.3887	5.6880	0.9584	0.7625
Huber	15.3466	5.5169	0.9263	0.4871
Logistic	15.3506	5.5361	0.9290	0.5096
Talwar	15.2843	5.2799	0.8871	0.3020
Welsch	15.2788	5.2632	0.8760	0.2872

Table 4 shows that MLE has an MSE of 0.8884, which is relatively high for location estimation compared to the other methods. RRRY, Bi-square, Andrews, and Welsch all have low MSE values (around 0.85-0.88), indicating strong performance in estimating the location parameter. MSE (Scale), The RRY again shows the highest MSE for scale (31.2473), reflecting poor performance in scale estimation. The MLE has a relatively higher MSE for scale (1.1035) compared to the other methods, though still not as high as RRY. RRRY, Bi-square, Andrews, Cauchy, and Welsch show low MSE values for scale (around 0.23-0.31), indicating strong accuracy in estimating scale.

This experiment was repeated (1000) times, and the average values from the location & scale parameters logistic distribution were estimated using the traditional and proposed methods, with the calculation of the MSE average for the two parameters shown in Table 5.

Table 5. Results of the simulation study (n = 200, μ = 15, and σ = 5)

Method	Location	Scale	MSE (Location)	MSE (Scale)
MLE	15.3010	5.8921	0.4783	0.9645
RRY	16.3262	9.3451	2.3452	24.1304
RRRY	15.2120	5.1109	0.4648	0.2135
Bi-square	15.2883	5.2193	0.4784	0.1548
Andrews	15.2878	5.2191	0.4781	0.1547
Cauchy	15.3167	5.2661	0.4953	0.1714
Fair	15.4031	5.5854	0.5642	0.4810
Huber	15.3619	5.4387	0.5323	0.3027
Logistic	15.3654	5.4548	0.5336	0.3185
Talwar	15.2956	5.2296	0.4862	0.1599
Welsch	15.2944	5.2251	0.4815	0.1551

Table 5 shows location estimates. The MLE method gives the closest estimate for the location parameter (15.3010), with an MSE of 0.4783, which is a relatively low value, indicating good accuracy. The RRY method shows the worst performance for location estimation, with a location estimate of 16.3262 and an MSE of 2.3452, which is much higher than other methods.

For scale estimation, RRRY provides the best result with an MSE of 0.2135, indicating it has the least error when estimating scale. RRY once again performs poorly, with an MSE of 24.1304, which is much higher than the others, suggesting that it is not well-suited for scale estimation. The robust methods, such as Bi-square, Andrews, and Huber, show relatively low MSE values for both location and scale compared to the RRY method, highlighting their resilience against outliers in the data. Bi-square and Andrews perform similarly, with MSE values of 0.4784 and 0.4781 for location and 0.1548 and 0.1547 for scale. The proposed method (Andrews) was the best compared to the rest of the methods

This experiment was repeated (1000) times, and the average values from the location & scale parameters logistic distribution were estimated using the traditional and proposed methods, with the calculation of the MSE average for the two parameters shown in Table 6.

Table 6. Results of the simulation study (n = 300, μ = 15, and σ = 5)

Method	Location	Scale	MSE (Location)	MSE (Scale)
MLE	15.2612	5.8995	0.3317	0.9185
RRY	16.2768	9.3037	2.0460	21.9632
RRRY	15.1695	5.1100	0.3009	0.1076
Bi-square	15.2436	5.2059	0.3292	0.1178
Andrews	15.2431	5.2054	0.3289	0.1177
Cauchy	15.2737	5.2498	0.3470	0.1307
Fair	15.3596	5.5447	0.4104	0.3840
Huber	15.3199	5.4099	0.3799	0.2385
Logistic	15.3234	5.4243	0.3820	0.2511
Talwar	15.2532	5.2173	0.3365	0.1228
Welsch	15.2497	5.2112	0.3321	0.1182

Table 6 shows location estimates. The RRRY again provides the most accurate estimate for the location parameter (15.1695), with an MSE of 0.3009. This is a low MSE, indicating that RRRY performs very well in terms of location estimation, though slightly worse than its performance at n = 100 and 200. The RRY method continues to show poor performance in location estimation, with an estimate of 16.2768 and an MSE of 2.0460, which is significantly higher than other methods. Scale estimates, the RRRY stands out for scale estimation, with an MSE of 0.1076, the lowest among all methods. RRY, like the previous table, is a poor scale estimation method with an MSE of 21.9632, the worst among all scale estimation methods. This is consistent with strong methods such as Bi-square, Andrews, and Welsch, and low MSEs for both location and scale. Bi-square and Andrews both have MSEs around 0.3292 and scale MSEs of 0.1178 and 0.1177, respectively, indicating that these methods are extremely robust and efficient. Cauchy, Talwar, Huber, Logistic, and Fair are all well rated, with MSEs in the moderate range for location and scale.

As reported in Tables 1–6, estimates for location and scale parameters are generally more accurate with a greater sample size. The RRRY method provides the best scale estimates for each scenario, while MLE and robust methods such as Bi-square and Andrews are good for location estimation. For both parameters, RRY performs the worst for both parameters in all sample sizes. All proposed techniques were computationally feasible for the sample sizes in this study. All simulations and empirical analyses were conducted using MATLAB. While the implementation code is not publicly available at this stage due to institutional constraints, all algorithmic steps and parameter settings are reported in sufficient detail to ensure reproducibility.

8. Real data

A sample of 130 observations of children under 10 years old and infants, from Hefei Children's Teaching Hospital, who were poisoned, was taken, and the duration of treatment received by the patients from their admission to the hospital until their recovery or death was calculated in hours.

The logistic distribution test was performed for survival time data, and the results are summarized in Table 7.

Table 7. Testing the fit of real data to logistics distribution

Method	Statistic	Critical Value	P-Value
Kolmogorov-Smirnov	0.0917	0.1451	0.226
Anderson-Darling	1.6911	3.9074	——-
Chi-Squared	4.0416	16.812	0.671

Table 7 provides the results of several goodness-of-fit tests to assess how the logistic distribution fits the real survival time data. The tests used are Kolmogorov-Smirnov, Anderson-Darling, and Chi-Squared tests. The significance level is set to 0.01, which means that we will reject the null hypothesis (that the data fits the logistic distribution) if the p-value is less than 0.01. Overall, there is no significant evidence to reject the logistic distribution as a fit for the survival time data. All the test results suggest that the data follows the logistic distribution reasonably well.

Figure 3 shows that red circles on the plot indicate outliers detected by both methods (the Z-score and IQR) after sorting the time in ascending order. These are outliers that significantly differ from most of the data, and the plot will display the entire dataset, with outliers marked in red and blue circles representing the regular data points. There are 16 outliers detected in the data. This means 16 data points are significantly different from the rest.

Table 8 shows that the RRRY method has the highest location parameter (82.89), while the MLE method has the lowest location parameter (56.39). The other methods, RRY, Bi-square, and Talwar, are clustered in a similar range between about 72 and 75. The highest scale parameter, the Talwar method, is 61.69, which means the largest spread in data. The MLE method has a small-scale parameter, 24.99; it can be assumed that there is a tighter distribution around the location parameter.

Talwar has the lowest MSE at 306.59 with the Best Performing Method (Lowest MSE). This suggests that Talwar is best suited for the data, relative to all other methods, in terms of MSE. It captures the relation between location and scale of data in a single, accurate way and reduces errors. RRRY is quite well fitted with an MSE of 372.48, and Bi-square is very well fitted, with MSE values close to each other and significantly superior to others. RRRY is particularly good and robust in that it minimizes MSE without overfitting. Andrews's method performs equally well (MSE = 365.12), but slightly better than Bi-square.

RRY, or Moderate Performers MSE = 1310.60, is a better fit than other more complex methods, but is still behind RRRY, Bi-square, Andrews, and Talwar. The Fair also competes, but is less effective than Bi-square, Andrews, and RRRY (MSE = 375.68). The MLE (MSE = 2164.30) has the highest MSE, suggesting that MLE does not fit the data as well as other methods. It might be more sensitive to outliers or may not be as robust as this dataset, leading to a higher error.

Methods like Cauchy (MSE = 384.84), Logistic (MSE = 385.70), and Welsch (MSE = 379.50) have higher MSEs compared to RRRY, Bi-square, Andrews, and Talwar, but still perform reasonably well in terms of fitting the data, though not as well as the top performers. The proposed method (Talwar) was the best compared to the rest of the methods, so the PDF for logistic distribution, CDF, $S(t)$, and $h(t)$ will be explained in Figure 4.

Table 8. Testing the fit of real data to logistics distribution

Method	Location	Scale	MSE
MLE	56.3873	24.9906	2164.300
RRY	73.0200	59.4872	1310.600
RRRY	82.8872	57.9158	372.4766
Bi-square	72.7820	60.0991	366.3281
Andrews	72.7464	60.1618	365.1162
Cauchy	74.2857	58.8338	384.8366
Fair	75.5820	58.7900	375.6831
Huber	74.9003	57.9259	404.0964
Logistic	74.7978	58.6398	385.6973
Talwar	75.1303	61.6933	306.5941
Welsch	73.0200	59.4872	379.5031

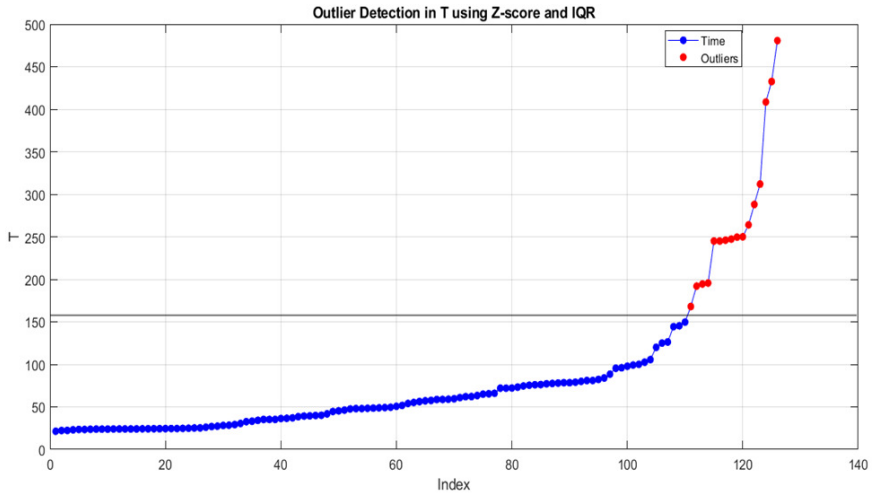

Figure 3. Identified outliers for real data.

Figure 4 will observe the following key points: The PDF curve for Talwar will peak around the location parameter μ, with a symmetrical tail on both sides. The CDF curve will start at 0 and increase to 1, indicating the cumulative probability of survival times being less than or equal to a particular value. The Survival Function will start from 1 (indicating that at time 0, all individuals survive) and decrease as time increases. The failure risk will be expressed in the hazard function, and this usually increases with initial growth, and then plateaus or decreases with increasing expansion. These plots will help illustrate the way in which logistic distribution models the actual data using the Talwar method and how it may be useful in analysis of the survival time.

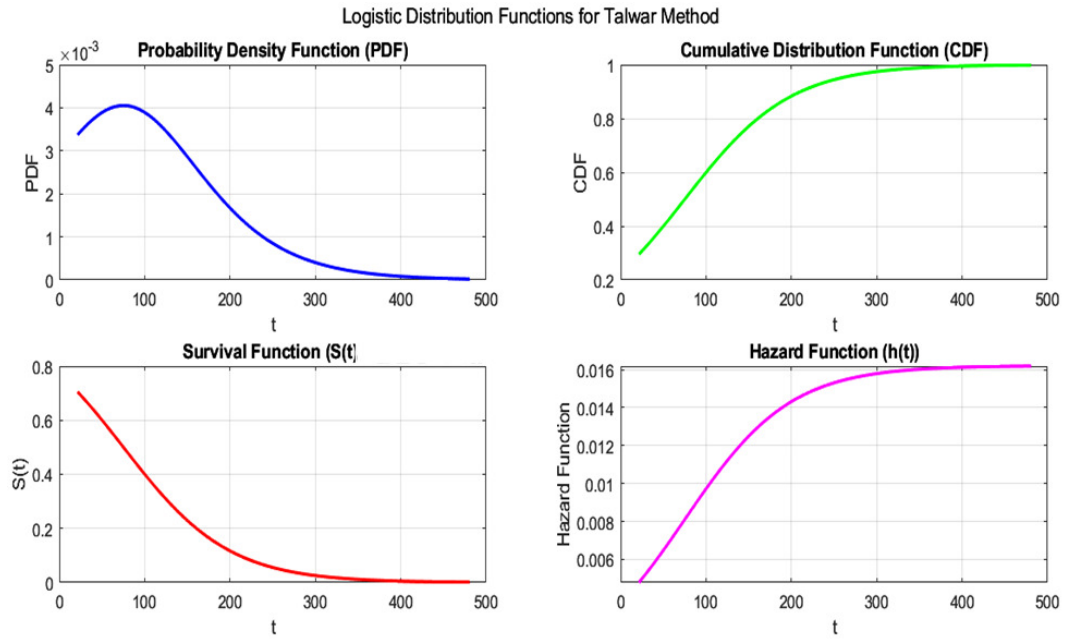

Figure 4. Logistic distribution functions (Talwar method) for real data.

9. Conclusions

1. The proposed techniques effectively address outliers and provide more accurate and robust parameter estimates than previous approaches.
2. The RRRY method is preferred for scale estimation since the RRRY method consistently produces the most accurate scale estimates for all simulations.
3. Strong M-estimation techniques like Bi-square or Andrews for location estimation are preferred, particularly when there are outlying observations.
4. The RRY method consistently produces the lowest results in location estimation as well as scale estimation. It is not, therefore, recommended where outliers are suspected.
5. Increasing sample size increases estimates of both location and scale parameters. This demonstrates the benefits of larger samples even under data irregularities.
6. In the real data application, the Talwar method achieves the lowest mean squared error (MSE) and is therefore recommended when predictive accuracy is the primary objective.
7. The classical MLE approach performs poorly in terms of MSE under outlier contamination, indicating that it should be avoided in applications where data irregularities are expected.

REFERENCES

1. P. J. Rousseeuw, and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, 1987.
2. M. A. M. Safari, N. Masseran, M. H. A. Majid, and R. R. M. Tajuddin, *Robust Estimation of the Three-Parameter Weibull Distribution for Addressing Outliers in Reliability Analysis*, Scientific Reports, vol. 15, no. 1, p. 11516, 2025. https://doi.org/10.1038/s41598-025-96043-1
3. H. S. Yavuz, and İ. Alayoğlu, *Robust lifetime estimation with small sample sized high speed railway ETCS component data*, Nicel Bilimler Dergisi, vol. 6, no. 2, pp. 181–207, 2024. https://doi.org/10.51541/nicel.1571005
4. M. H. Tareq Hasan, T. H. Ali, and N. H. Sedeek Kareem, *Multivariate CUSUM Daubechies Discrete Wavelet Transformation Coefficients Charts for Quality Control*, Passer Journal of Basic and Applied Sciences, vol. 7, no. 1, pp. 533–546, 2025. https://doi.org/10.24271/psr.2025.491277.1834
5. T. H. Ali, H. A. M. Hayawi, and D. S. I. Botani, *Estimation of the bandwidth parameter in Nadaraya–Watson kernel non-parametric regression based on universal threshold level*, Communications in Statistics - Simulation and Computation, vol. 52, no. 4, pp. 1476–1489, 2023. https://doi.org/10.1080/03610918.2021.1884719
6. J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, 1982.
7. A. Ragab and J. Green, *On order statistics from the log-logistic distribution and their properties*, Commun. Statist. Theory Methods, **13**, no. 21, pp. 2713–2724, 1984. DOI: https://doi.org/10.1080/03610928408828855.
8. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 6th ed., Duxbury Press, 2003.
9. D. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text*, Springer, New York, 1996.
10. A. W. Omer, and T. H. Ali, *Dealing with the outlier problem in multivariate linear regression analysis using the Hampel filter*, Kurdistan Journal of Applied Research, vol. 10, no. 1, 2025. https://doi.org/10.24017/science.2025.1.1
11. W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introduction to Probability and Statistics*, 14th ed., Cengage Learning, 2012.
12. M. A. Hasawy, A. J. Ibrahim, T. H. Ali, and B. S. Sedeeq, *Fast MCD-augmented rank regression for robust estimation of Weibull parameters*, Passer Journal of Basic and Applied Sciences, vol. 7, no. 2, pp. 689–697, 2025. https://doi.org/10.24271/psr.2025.511694.2006
13. J. W. Tukey, *Exploratory Data Analysis*, vol. 2, Pearson, London, UK, 1977.
14. H. A. Hayawi, S. M. Azeez, S. O. Babakr, and T. H. Ali, *ARX time series model analysis with wavelets shrinkage (simulation study)*, Pakistan Journal of Statistics, vol. 41, no. 2, pp. 103–116, 2025. https://www.pakjs.com/wp-content/uploads/2025/04/PJS-41201.pdf
15. G. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press, Belmont, CA, 1990.
16. T. A. A. R. S. Al, M. M. Taher, and T. H. Ali, *A novel Dmey wavelet charts for controlling and monitoring the average and variance of quality characteristics*, Statistics, Optimization & Information Computing, vol. 14, no. 6, pp. 3706–3717, 2025. https://doi.org/10.19139/soic-2310-5070-2621
17. W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, vol. 78, John Wiley & Sons, New York, NY, USA, 1998.
18. W. Nelson, *Applied Life Data Analysis*, John Wiley & Sons, New York, NY, USA, 1981.
19. P. J. Huber, *Robust regression: Asymptotics, conjectures and Monte Carlo*, Annals of Statistics, vol. 1, no. 5, pp. 799–821, 1973.
20. D. C. Montgomery, *Introduction to Statistical Process Control*, John Wiley & Sons, Hoboken, 2009.
21. Lukman, A. F., Dawoud, I., Kibria, B. G., Algamal, Z. Y., & Aladeitan, B. *A new ridge-type estimator for the gamma regression model* Scientifica, vol. 1, p.5545356, 2021.
22. Naziyah, A. A., & Algamal, Z. Y. *Jackknifed Liu-type estimator in Poisson regression model* Journal of the Iranian Statistical Society, vol.19, no.1, p. 21-37, 2020.
23. Algamal, Z., & Ali, H. M. *An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression* , Electronic Journal of Applied Statistical Analysis, vol . 10 , no . 1 , 242-256, 2017.

24. D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed., John Wiley & Sons, New York, 2013.
25. P. J. Rousseeuw, and K. V. Driessen, *A fast algorithm for the minimum covariance determinant estimator*, Technometrics, vol. 41, no. 3, pp. 212–223, 1999.
26. D. Botani, N. Kareem, T. H. Ali, and B. Sedeeq, *Optimizing bandwidth parameter estimation for nonparametric regression using fixed-form threshold with Dmey and Coiflet wavelets*, Hacettepe Journal of Mathematics and Statistics, vol. 54, no. 3, pp. 1094–1106, 2025. https://doi.org/10.15672/hujms.1605499
27. T. H. Ali and D. M. Saleh, *Comparison between wavelet Bayesian and Bayesian estimators to remedy contamination in linear regression model*, PalArch's J. Archaeol. Egypt/Egyptol., **18**, no. 10, pp. 3388–3409, 2021.
28. A. W. Omer, and T. H. Ali, *Wavelet analysis for outlier estimation in multivariate linear regression models*, Passer Journal of Basic and Applied Sciences, vol. 7, no. 1, pp. 478–494, 2025. https://doi.org/10.24271/psr.2025.509315.1976
29. T. H. Ali, *Modification of the adaptive Nadaraya–Watson kernel method for nonparametric regression: A simulation study*, Communications in Statistics - Simulation and Computation, vol. 51, no. 2, pp. 391–403, 2022. https://doi.org/10.1080/03610918.2019.1652319
30. R. V. Hogg, J. W. McKean and A. T. Craig, *Introduction to Mathematical Statistics*, 8th ed., Pearson, Boston, 2019.
31. I. I. Elias, and T. H. Ali, *Optimal level and order of the Coiflets wavelet in VAR time series denoise analysis*, Frontiers in Applied Mathematics and Statistics, vol. 11, p. 1526540, 2025. https://doi.org/10.3389/fams.2025.1526540
32. T. H. Ali, B. S. Sedeeq, D. M. Saleh, and A. G. Rahim, *Robust multivariate quality control charts for enhanced variability monitoring*, Quality and Reliability Engineering International, vol. 40, no. 3, pp. 1369–1381, 2024. https://doi.org/10.1002/qre.3472
33. T. H. Ali, A. A. Hamad, S. H. Mahmood, and K. H. Ahmed, *ARIMAX time series analysis for a general budget in the Kurdistan Region of Iraq using wavelet shrinkage*, Communications in Statistics: Case Studies, Data Analysis and Applications, vol. 11, no. 2, pp. 164–188, 2025. https://doi.org/10.1080/23737484.2025.2486984
34. I. I. Elias and T. H. Ali, *VARMA time series model analysis using discrete wavelet transformation coefficients for Coiflets wavelet*, Passer Journal of Basic and Applied Sciences, vol. 7, no. 2, pp. 657–677, 2025. https://doi.org/10.24271/psr.2025.512121.2011
35. H. Hazem Taha, H. A. A. Hayawi, T. H. Ali, and S. R. Ahmed, *Wavelet Daubechies enhanced average chart incorporating classical Shewhart and Bayesian techniques*, Statistics, Optimization & Information Computing, vol. 14, no. 5, pp. 2312–2328, 2025. https://doi.org/10.19139/soic-2310-5070-2742
36. T. H. Ali, H. H. Taha, B. S. Sedeeq, et al., *Novel Hampel filtering and robust regression in control chart applications for regression process monitoring*, Journal of Statistical Theory and Applications, vol. 24, pp. 842–860, 2025. https://doi.org/10.1007/s44199-025-00131-0
37. T. H. Ali, H. H. Taha, B. S. Sedeeq, and H. A. Hayawi, *An innovative hybrid control chart combining wavelet decomposition and support vector machine for effective outlier detection*, Frontiers in Applied Mathematics and Statistics, vol. 11, p. 1682448, 2025. https://doi.org/10.3389/fams.2025.1682448
38. S. H. Mahmood, H. A. A. Hayawi, T. H. Ali, B. S. Sedeeq, and S. H. Ali, *Forecasting the CPI of the Kurdistan Region of Iraq using the combined ARIMA and GARCH model with wavelet analysis*, Journal of Applied Mathematics, vol. 2025, Article ID 2457525, 2025. https://doi.org/10.1155/jama/2457525
39. T. H. Ali, D. Lazgeen Ramadhan, and S. S. Ismael, *Hybrid robust beta regression based on support vector machines and iterative reweighted least squares*, Statistics, Optimization & Information Computing, vol. 15, no. 3, pp. 1513–1527, 2026. http://www.iapress.org/index.php/soic/article/view/2850
40. M. M. Taher, T. A. A.-R. S. Al-Hasso, and T. H. Ali, *Comparative evaluation of classical robust and wavelet enhanced beta regression models for proportional data*, Statistics, Optimization & Information Computing, vol. 15, no. 1, pp. 324–335, 2025. https://doi.org/10.19139/soic-2310-5070-3144
41. T. A.-R. S. Al, H. A. A. Hasso, and T. H. Ali, *Using linear wavelets in analyzing the GARCH model with the simulation*, Pakistan Journal of Statistics, vol. 42, no. 1, pp. 27–44, 2026. https://www.pakjs.com/wp-content/uploads/2026/02/PJS-42103.pdf
42. A. M. Abdulqader, J. F. Salih, and T. H. Ali, *Development of novel quality control charts for individual observations using Haar and Dmey wavelet transforms*, Frontiers in Applied Mathematics and Statistics, vol. 12, 2026. https://doi.org/10.3389/fams.2026.1765200
43. H. Waleed Yaseen, T. Hussein Ali, S. Ramzi Ahmed, and H. Hayawi, *The impact of wavelet-based denoising on beta regression model fit*, Statistics, Optimization & Information Computing, vol. 15, no. 4, pp. 3042–3058, 2026. https://doi.org/10.19139/soic-2310-5070-3344
44. A. W. Omer and T. H. Ali, *Remediating contamination of multivariate linear models using Hampel and wavelets filters*, Communications in Statistics - Theory and Methods, vol. 55, no. 17, pp. 5846–5868, 2026. https://doi.org/10.1080/03610926.2026.2645978
45. T. H. Ali, A. M. Abdulqader, and L. I. Kework, *Structured wavelet-based sparse discriminant analysis in high dimensions*, Statistics & Probability Letters, vol. 237, p. 110793, 2026. https://doi.org/10.1016/j.spl.2026.110793
46. N. S. I. Alsharabi, E. A. Haydier, and T. H. Ali, *A Coiflets-based wavelet decomposition method for improving parameter estimation in combined ARIMA–GARCH models*, Journal of Applied Mathematics, vol. 2026, Article ID 4896128, 11 pages, 2026. https://doi.org/10.1155/jama/4896128
47. M. A. I. Hasawy, T. H. Ali, and B. S. Sedeeq, *Robust rank regression for Weibull parameter estimation using M-estimation and Fast MCD*, Quality and Reliability Engineering International, vol. 42, no. 6, pp. 3486–3502, 2026. https://doi.org/10.1002/qre.70282
48. T. H. Ali, M. T. Hasan, A. W. Omer, and M. W. Awan, *Dual-robust adaptive joint CUSUM monitoring with application to tensile strength control in steel manufacturing*, Quality and Reliability Engineering International, vol. 42, no. 6, pp. 3520–3530, 2026. https://doi.org/10.1002/qre.70281

49. T. H. Ali, D. J. Hassan, and Z. O. Ismael, *Ali exponential–difference wavelet for robust signal denoising and localized anomaly detection under contaminated stochastic environments*, International Journal of Wavelets, Multiresolution and Information Processing, p. 2650029, 2026. https://doi.org/10.1142/S0219691326500293
50. S. S. Ismael, N. S. Ibrahim, and T. H. Ali, *A computational multiresolution nonlinear autoregressive framework based on wavelet decomposition and support vector regression*, Computational and Mathematical Methods, vol. 2026, Article ID 3218150, 19 pages, 2026. https://doi.org/10.1155/cmm4/3218150