Available online at www.garmian.edu.krd

Passer Journal of Basic and Applied Sciences

Passer

Passer 7 (Issue 2) (2025) 689-697

Original Article

Fast MCD-Augmented Rank Regression for Robust Estimation of Weibull Parameters

Mohammed Abduljabar Ibrahim Hasawy [1], Taha Hussein Ali * [1], Bekhal Samad Sedeeq [1]

[1] [1]*Department of Statistics and Informatics, College of Administration and Economics, University of Salahaddin, Erbil 44001, Kurdistan Region, Iraq.*

Received 10 March 2025; revised 19 May 2025;
accepted 09 June 2025; available online 26 July 2025

DOI: 10.24271/psr.2025.511694.2006

ABSTRACT

Outliers hurt the accuracy of life distribution parameters, including the Weibull distribution. Therefore, we suggest employing the fast minimum covariance determinant method in rank regression estimators (Although rank regression provides robust estimates for the shape and scale parameters of the Weibull distribution, it tends to lack efficiency compared to other estimation methods, particularly under ideal model conditions.) to obtain robust estimators for the shape and scale parameters of the Weibull distribution. The proposed method is based on the robust means vector and the robust covariance matrix obtained from the fast minimum covariance determinant method and employed in the rank regression estimation method that depends on the ordinary least squares estimators of the simple linear regression model. The estimated parameters of the Weibull distribution obtained using the proposed technique have been compared with those derived from the conventional maximum likelihood estimation and rank regression, utilizing mean square error as the comparison metric via both simulation and real data. The study's results confirmed the effectiveness of the proposed approach in managing outliers and delivering highly efficient estimators for the shape and scale parameters of the Weibull distribution.

Keywords: Rank Regression, Weibull distribution, Maximum Likelihood Estimation, Robust Estimation, Outliers.

1 Introduction

A lifetime distribution is a statistical distribution that describes the expected time until a significant event occurs, such as the death of an organism, the failure of a mechanical component, or the time required to complete a task. These distributions are widely used in reliability engineering (Scheduling maintenance, estimating the lifespan of systems, and improving product designs), risk management, and survival analysis (time-to-event data in medical studies, as the time until recurrence of the disease or death). Known lifetime distributions include Weibull, exponential, Normal, lognormal, Gamma, and Gompertz distributions [1]. Weibull distribution is a flexible distribution that can model decreasing, constant, or increasing mean time to failure and includes two parameters, shape and scale. The nature of the failure average determines the shape parameter (for example, the increasing failure average for a shape parameter greater than one).

On the other hand, the issue of outliers is seen as a significant challenge that impacts the precision of parameter estimation for life distributions, including the Weibull distribution. Maximum likelihood estimators are ineffective in estimating the parameters of a mathematical distribution when outliers are present, but the rank regression technique, which relies on the linear function of the Weibull distribution parameters, exhibits a degree of robustness to outliers. In this study, the term "rank regression" specifically refers to the regression procedure commonly applied in reliability analysis, where estimated unreliability values (typically obtained through median rank or plotting position formulas) are used to transform lifetime data. A linear regression is then performed on the log-log transformed data to estimate the shape and scale parameters of the Weibull distribution. This usage differs from other interpretations of "rank regression" in statistical literature, which may involve regression on ranked data or residuals.

Boudt et al. (2011) proposed three robust Weibull distribution parameter estimators: the quantile regression, the median/Qn estimator (The Qn estimator is used to estimate the scale (dispersion) of the transformed data robustly), and the repeated median estimator. Derived their breakdown point, asymptotic variance, and influence function. The methods are illustrated on real Weibull distribution data affected by outliers [2].

Coria et al. (2016) proposed estimating both parameters straightforwardly for the Weibull distribution. They analyzed right-censored lifetime data sets with varying sample sizes and censoring percentages to evaluate the effectiveness of the

* Corresponding author
E-mail address taha.ali@su.edu.krd (Instructor).
Peer-reviewed under the responsibility of the University of Garmian.

proposed estimator. demonstrate that the parameter estimator yields a high level of accuracy [3].
Sinha (2019) proposed a robust method for fitting the accelerated failure time model to Weibull distribution data by determining the influence of outliers on both the response variable and associated covariates. The finite-sample properties of the estimators are investigated based on simulated results and real data from breast cancer patients [4].
Roohanizadeh et al. (2022) delineated various estimation methodologies for the two-parameter Weibull distribution utilizing intuitionistic fuzzy lifetime data, encompassing maximum likelihood estimation (employing Newton-Raphson and Expectation-Maximization techniques) and Bayesian estimation (utilizing Tierney and Kadane's approximation for the shape and scale parameters, with Gamma and inverse-Gamma priors, respectively) [5].
The simulation is conducted to identify the most efficient estimator within the intuitionistic fuzzy methodology. Gómez et al. (2024) delineated five parameters of the Weibull distribution, each derived from distinct moments of the mean, quantile, and mode. Furthermore, it explores the meaning of regression results when integrating linear regression models into these parameters. They provided insights into its implications for modelling failure periods and prospective contributions to disciplines requiring reliability and survival analysis [6].
In recent years, considerable attention has been directed toward improving parameter estimation techniques for reliability distributions, particularly in the presence of extreme values or outliers. The three-parameter Weibull distribution has been widely recognized for its flexibility and suitability in modelling lifetime data and failure mechanisms across various industrial and engineering applications. Traditional estimation methods, however, often struggle to maintain accuracy when datasets contain extreme observations, leading to biased parameter estimates and unreliable reliability predictions. In response to this challenge, Safari et al. (2025) introduced the Perturbation Iterative Technique Estimator (PITE) method, a novel technique specifically designed for estimating the parameters of the three-parameter Weibull distribution in the presence of outliers. Their findings demonstrated that the PITE method outperforms conventional estimators in terms of robustness and accuracy, particularly in highly contaminated datasets. This advancement represents a significant contribution to the field of reliability analysis, suggesting that integrating modern, outlier-resistant estimation procedures could substantially improve the validity of reliability modelling and system risk assessment in future studies [7].
In this paper, the proposed method is based on the robust means vector and the robust covariance matrix obtained from the fast minimum covariance determinant method and employed in the rank regression estimation method that depends on the ordinary least squares estimators of the simple linear regression model.

2 Theoretical Aspect

2. 1 Fast Minimum Covariance Determinant

The Fast Minimum Covariance Determinant (FMCD) method, as presented by Rousseeuw and Leroy (1987), is a modification of the minimum covariance determinant (MCD) [8]. This method selects l data points from a total of N, where $N/2 < l \leq N$, to achieve the least possible determinant from the classical variance-covariance matrix. The estimate comprises the mean vector and covariance matrix of the defined l points, adjusted by a consistency value to ensure adherence to the multivariate normal distribution (MND), along with a correction value to address bias in small sample sizes. Based on this, the estimates of the robust mean vector (MR) and the robust variance-covariance matrix (SR) are calculated to be resilient against outliers. These estimators can be configured to achieve a theoretically maximal breakdown point of 50%. This configuration enables the detection of outliers, even when their quantity approaches nearly half of the sample size.
The purpose of MCD is to identify h observations (out of N) that minimize the determinant of the classical covariance. The MCD estimate of location is the mean of these h points, whereas their covariance matrix represents the MCD estimate of dispersion. The resultant breakdown value is equivalent to the minimal volume ellipsoid (MVE); nevertheless, the minimum covariance determinant (MCD) offers significant benefits over the MVE since it exhibits superior statistical efficiency due to its asymptotic normality [9].
A novel method for MCD, called FMCD, has been developed to address such issues. The fundamental concept involves an inequality related to order statistics and determinants, using methods referred to as "selective iteration" or "nested extensions". FMCD generally identifies the precise MCD for small datasets, but for bigger datasets, it yields more accurate answers than current methods and operates much more quickly. FMCD can identify a precise hyperplane that encompasses h or more observations. The algorithm makes the MCD technique a standard instrument for analyzing multivariate data.
Recent studies have reported that using FMCD-enhanced Mahalanobis Distance improves the stability and reliability of outlier detection results, particularly in systems involving complex sensor data and correlated reliability indicators [10]. As modern reliability engineering increasingly relies on real-time multivariate monitoring and predictive analytics, combining robust estimators like FMCD with traditional distance-based techniques represents a practical and effective solution for managing data quality issues and enhancing decision-making in reliability management.

2. 2 Weibull Distribution

The Weibull distribution is a probability distribution often used in reliability engineering and survival analysis to model the time until a specific event occurs. It is named after Wallodi Weibull, who introduced it in the mid-20th century [11]. The distribution is characterized by two parameters: the shape parameter (often denoted as "β") and the scale parameter (often denoted as "η"). The probability density function (PDF) of the Weibull distribution is expressed as [12]:

$$f(x) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^{\beta}} \quad (1)$$

Where:

$f(x) \geq 0,\ x \geq 0,\ \beta \geq 0,\ \eta \geq 0$

2. 3 Estimation Methods

Determining the values of the parameters that best fit the observed data to a probability distribution is essential for estimation [13]. The selection of the estimation technique is contingent upon the distribution you are addressing. This section will concentrate on the techniques employed to estimate the parameters of the two-parameter Weibull distribution as utilized in this study.

2. 3. 1 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a prevalent and robust technique for parameter estimation over several distributions. Maximum Likelihood Estimation (MLE) seeks to identify parameter values that maximize the probability of the observed data [14], [15]. The equations for the partial derivatives of the log-likelihood function are formulated and shown below:

$$\frac{\partial\Lambda}{\partial\beta} = \frac{n}{\beta} + \sum_{i=1}^{n} ln\left(\frac{x_i}{\eta}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\eta}\right)^{\beta} ln\left(\frac{x_i}{\eta}\right) = 0$$

And:

$$\frac{\partial\Lambda}{\partial\eta} = \frac{-\beta}{\eta}.n + \frac{\beta}{\eta}\sum_{i=1}^{n}\left(\frac{x_i}{\eta}\right)^{\beta} = 0$$

The parameter values are estimated by solving the system of equations presented above concurrently.

2. 3. 2 Rank Regression on Y

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. This methodology mirrors the probability plotting method; however, it employs the principle of least squares to ascertain the line through the points rather than relying on visual estimation. The initial step involves transforming our function into a linear format [16]. For the two-parameter Weibull distribution, the cumulative density function is:

$$F(x) = 1 - e^{-\left(\frac{x}{\eta}\right)^{\beta}} \quad (2)$$

Applying the natural logarithm to both sides of the equation results

$$ln[1 - F(x)] = -\left(\frac{x}{\eta}\right)^{\beta}$$

$$ln\{-ln[1 - F(x)]\} = \beta ln\left(\frac{x}{\eta}\right) = -\beta ln(\eta) + \beta ln(x)$$

Now let:

$$y = ln\{-ln[1 - F(x)]\} \quad (3)$$

And:

$$a = \beta\ ln(\eta)\ and\ b = \beta \quad (4)$$

which results in the linear equation of:

$$y = a + bx$$

The technique of least squares parameter estimation, which is sometimes referred to as regression analysis, was covered in the article on parameter estimation [17], [18], and the equations for regression on Y were constructed as follows:

$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} \quad (5)$$

And:

$$\hat{b} = \hat{\rho}\sqrt{\frac{S_{yy}}{S_{xx}}} \quad (6)$$

$$\hat{\rho} = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}} \quad ; \quad S^2 = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}$$

The sample variances are located on the main diagonal of matrix S^2, 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:

$$y_i = ln\{-ln[1 - F(x_i)]\}$$

And:

$$x_i = ln(x_i)$$

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. 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 [19], first to the beta distribution and then to the F distribution.

$$MR = \frac{1}{1 + ((N-j+1)/j)F_{0.50;\,m;\,n}} \quad (7)$$

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

The estimated parameters are:

$$\hat{\beta} = \hat{b} \quad (8)$$

$$\hat{\eta} = e^{-\hat{a}/\hat{b}} \quad (9)$$

2. 3. 3 Proposed Method

To deal with the outlier problem, the proposed method employs the robust mean vector, and the robust covariance matrix is calculated by the Fast Minimum Covariance Determinant (FMCD) method in the rank regression method to get Robust

Rank Regression of dependent Y (RRRY) estimate Weibull distribution parameters through the following.

- Let x = Ln(t), where t represents the failure time, be ranked in ascending order, and x is an (n × 1) vector.

- Compute the cumulative density function $F(x_i)$ from the median ranks (MR) in equation (7).

- $y_i = Ln\{-Ln[1 - F(x_i)]\}$.

- The robust mean vector (MR) and covariance matrix (S^2R) are computed using the Fast Minimum Covariance Determinant (FMCD) method, as described in Section 2.1.

$$\mathrm{MR} = [\bar{\mathrm{x}}\mathrm{R} \quad \bar{\mathrm{y}}\mathrm{R}], \text{ and } \mathrm{S^2R} = \begin{bmatrix} \mathrm{SxxR} & \mathrm{SxyR} \\ \mathrm{SyxR} & \mathrm{SyyR} \end{bmatrix}$$

- The discussion centered around the least squares parameter estimation method applied to the estimation of Weibull distribution parameters, leading to the formulation of the following linear equation:

$$\mathrm{y} = \mathrm{aR} + \mathrm{bRx}$$

$$\hat{\mathrm{b}}\mathrm{R} = \hat{\rho}\mathrm{R}\sqrt{\frac{\mathrm{SyyR}}{\mathrm{SxxR}}} \quad (10)$$

Where:

$$\hat{\rho}\mathrm{R} = \frac{\mathrm{SxyR}}{\sqrt{\mathrm{SxxR} \times \mathrm{SyyR}}}$$

And:

$$\hat{\mathrm{a}}\mathrm{R} = \bar{\mathrm{y}}\mathrm{R} - \hat{\mathrm{b}}\mathrm{R} \times \bar{\mathrm{x}}\,\mathrm{R} \quad (11)$$

- Finally, the proposed estimation for the shape and scale parameters of the Weibull distribution is:

$$\hat{\beta} = \hat{\mathrm{b}}\mathrm{R} \quad (12)$$

$$\hat{\eta} = e^{-\hat{\mathrm{a}}\mathrm{R}/\hat{\mathrm{b}}\mathrm{R}} \quad (13)$$

2. 4 Mean Squared Error

The efficacy of the estimated parameters may be assessed using the mean squared error (MSE), which yields the minimal value for the estimated parameters $\hat{\theta}$ with optimal efficiency, calculable by the following formula:

$$MSE = \frac{\sum_{i=1}^{m}(\hat{\theta}_i - \theta)^2}{m} \quad (14)$$

Where θ represents real parameters and $\theta = [\beta\ \eta]$, and m is the number of observations.

3 Application Aspect

To prove the efficiency of the proposed method (RRRY) in estimating the shape and scale parameters of the Weibull distribution and to compare it with some classical methods (MLE and RRY), simulations were used in addition to real data, as follows:

3. 1 Simulation Study

Data having a Weibull distribution with the shape and scale parameters (for several different values) and for several sample sizes (50, 100, 200, 300, 400, and 500) were generated using a program designed for this purpose in the MATLAB language, in addition, creating a program that calculates the cumulative function of the median ranks at any sample size based on the beta and *F* distribution approach. Also, outliers have been added to the generated data, using a random selection of 5 indices from the generated data and multiplying the corresponding values by 10 to introduce significant outliers.

The initial simulation experiment utilizing Mahalanobis distance values is presented in Figure 1 for the FMCD method, revealing a total of 15 outliers among the 50 generated data points. To compare the proposed method (RRRY) and classical methods (MLE and RRY), the simulation experiments were repeated (1000) times, the parameters of the Weibull distribution were estimated, and the mean of average squared error (MASE) for MSE(Beta) and MSE(Eta) and Error Mean were calculated. The results were summarized in Tables 1-4:

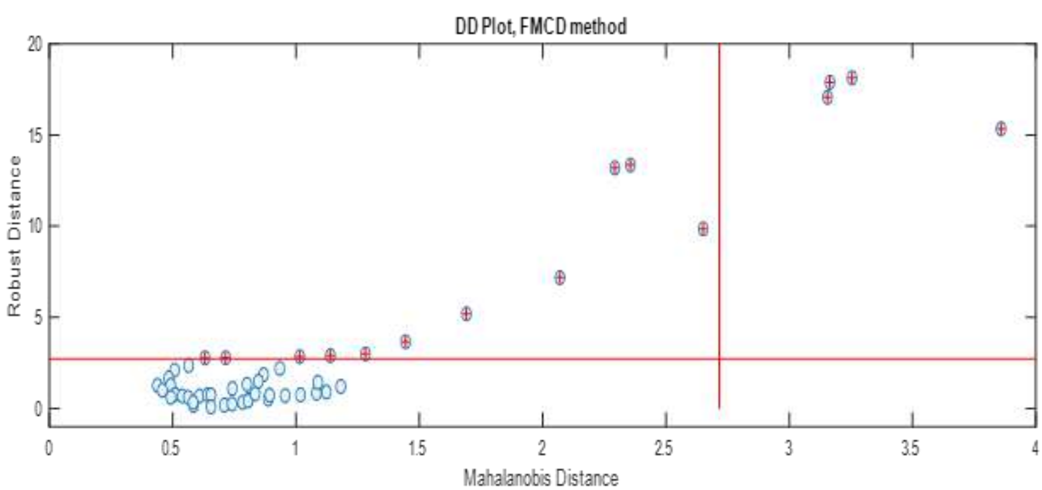

Figure 1: Mahalanobis distance for the first simulation experiment.

Table 1: MASE when beta = 2 & eta = 20

Sample Size	Criterion	MLE		RRY		RRRY	
		Parameters	MASE	Parameters	MASE	Parameters	MASE
50	MSE(Beta)	3.0926	1.1937	2.8266	0.6833	2.3779	0.1428
	MSE(Eta)	31.7751	138.6521	31.1316	123.9135	22.9562	8.7389
	Error Mean		69.9229		62.2984		4.4408
100	MSE(Beta)	2.9757	0.9520	2.5808	0.3373	2.2694	0.0726
	MSE(Eta)	25.9980	35.9755	25.3444	28.5627	21.5991	2.5572
	Error Mean		18.4638		14.4500		1.3149
200	MSE(Beta)	2.8431	0.7108	2.3681	0.1355	2.1761	0.0310
	MSE(Eta)	23.1794	10.1083	22.6881	7.2261	20.9609	0.9234
	Error Mean		5.4096		3.6808		0.4772
300	MSE(Beta)	2.7653	0.5857	2.2733	0.0747	2.1473	0.0217
	MSE(Eta)	22.1867	4.7816	21.8158	3.2973	20.7386	0.5456
	Error Mean		2.6837		1.6860		0.2836
400	MSE(Beta)	2.7087	0.5023	2.2195	0.0482	2.1217	0.0148
	MSE(Eta)	21.6540	2.7358	21.3707	1.8789	20.6312	0.3984
	Error Mean		1.6191		0.9635		0.2066
500	MSE(Beta)	2.6629	0.4394	2.1841	0.0339	2.1127	0.0127
	MSE(Eta)	21.3402	1.7962	21.1173	1.2483	20.5504	0.3029
	Error Mean		1.1178		0.6411		0.1578

Table 2: MASE when beta = 5 & eta = 20

Sample Size	Criterion	MLE		RRY		RRRY	
		Parameters	MASE	Parameters	MASE	Parameters	MASE
50	MSE(Beta)	9.0128	16.1022	8.8148	14.5530	5.9410	0.8855
	MSE(Eta)	34.6977	216.0218	36.3066	265.9051	21.1290	1.2746
	Error Mean		116.0620		140.2290		1.0801
100	MSE(Beta)	8.8735	15.0043	8.4071	11.6083	5.6754	0.4561
	MSE(Eta)	28.2106	67.4139	28.6678	75.1310	20.6265	0.3925
	Error Mean		41.2091		43.3696		0.4243
200	MSE(Beta)	8.7057	13.7319	7.8191	7.9472	5.4406	0.1941
	MSE(Eta)	24.8364	23.3910	24.5978	21.1394	20.3797	0.1442
	Error Mean		18.5615		14.5433		0.1692
300	MSE(Beta)	8.6007	12.9651	7.4111	5.8134	5.3693	0.1364
	MSE(Eta)	23.5902	12.8897	23.1470	9.9037	20.2926	0.0856
	Error Mean		12.9274		7.8586		0.1110
400	MSE(Beta)	8.5212	12.3987	7.1054	4.4329	5.3058	0.0935
	MSE(Eta)	22.8981	8.3988	22.3857	5.6917	20.2504	0.0627
	Error Mean		10.3987		5.0623		0.0781
500	MSE(Beta)	8.4571	11.9515	6.8672	3.4865	5.2827	0.0799
	MSE(Eta)	22.4595	6.0489	21.9247	3.7043	20.2193	0.0481
	Error Mean		9.0002		3.5954		0.0640

Table 3: MASE when beta = 2 & eta = 30

Sample Size	Criterion	MLE		RRY		RRRY	
		Parameters	MASE	Parameters	MASE	Parameters	MASE
50	MSE(Beta)	3.0926	1.1937	2.8266	0.6833	2.3779	0.1428
	MSE(Eta)	47.6626	311.9672	36.6975	278.8053	33.1084	9.6624
	Error Mean		156.5805		139.7443		4.9026
100	MSE(Beta)	2.9757	0.9520	2.5808	0.3373	2.2694	0.0726

	MSE(Eta)	38.9969	80.9450	38.0166	64.2662	32.3987	5.7538
	Error Mean		40.9485		32.3017		2.9132
200	MSE(Beta)	2.8431	0.7108	2.3681	0.1355	2.1761	0.0310
	MSE(Eta)	34.7690	22.7436	34.0322	16.2588	31.4414	2.0777
	Error Mean		11.7272		8.1972		1.0544
300	MSE(Beta)	2.7653	0.5857	2.2733	0.0747	2.1473	0.0217
	MSE(Eta)	33.2800	10.7586	32.7237	7.4188	31.1079	1.2275
	Error Mean		5.6722		3.7468		0.6246
400	MSE(Beta)	2.7087	0.5023	2.2195	0.0482	2.1217	0.0148
	MSE(Eta)	32.4810	6.1555	32.0561	4.2275	30.9468	0.8965
	Error Mean		3.3289		2.1378		0.4556
500	MSE(Beta)	2.6629	0.4394	2.1841	0.0339	2.1127	0.0127
	MSE(Eta)	32.0103	4.0414	31.6759	2.8086	30.8255	0.6815
	Error Mean		2.2404		1.4213		0.3471

Table 4: MASE when beta = 5 & eta = 30

Sample Size	Criterion	MLE		RRY		RRRY	
		Parameters	MASE	Parameters	MASE	Parameters	MASE
50	MSE(Beta)	9.0128	16.1022	8.8148	14.5530	5.9410	0.8855
	MSE(Eta)	52.0465	486.0491	54.4599	598.2864	31.6935	2.8679
	Error Mean		251.0756		306.4197		1.8767
100	MSE(Beta)	8.8735	15.0043	8.4071	11.6083	5.6754	0.4561
	MSE(Eta)	42.3159	151.6813	43.0017	169.0447	30.9398	0.8832
	Error Mean		83.3428		90.3265		0.6696
200	MSE(Beta)	8.7057	13.7319	7.8191	7.9472	5.4406	0.1941
	MSE(Eta)	37.2546	52.6298	36.8966	47.5637	30.5696	0.3244
	Error Mean		33.1809		27.7555		0.2593
300	MSE(Beta)	8.6007	12.9651	7.4111	5.8134	5.3693	0.1364
	MSE(Eta)	35.3853	29.0017	34.7205	22.2834	30.4389	0.1926
	Error Mean		20.9834		14.0484		0.1645
400	MSE(Beta)	8.5212	12.3987	7.1054	4.4329	5.3058	0.0935
	MSE(Eta)	34.3471	18.8972	33.5786	12.8062	30.3756	0.1411
	Error Mean		15.6479		8.6196		0.1173
500	MSE(Beta)	8.4571	11.9515	6.8672	3.4865	5.2827	0.0799
	MSE(Eta)	33.6892	13.6100	32.8870	8.3346	30.3288	0.1081
	Error Mean		12.7807		5.9105		0.0940

4 Discussions

The results of Tables 1-4 show the following:

- The proposed (FMCD) method (RRRY) outperformed the classical techniques (MLE and RRY), depending on the MSE and Error Mean for all simulation cases.

- The accuracy of the shape parameter estimates was better than the scale parameter estimates for the three methods and all simulation cases.

- The accuracy of estimating Weibull distribution parameters increases when the sample size increases.

- The accuracy of estimating the Weibull distribution shape parameter decreases as its assumed value increases for all simulation cases.

- The accuracy of the scale parameter estimation of the Weibull distribution decreases as its assumed value increases for all simulation cases.

- The estimate of the shape parameter remains unaffected by an increase in the scale parameter, whereas an increase influences the estimate of the scale parameter in the shape parameter.

- The Error Mean indicates the superiority of the proposed method (RRRY) over the classical methods (RRY and MLE), while the classical method (RRY) is superior to the classical method (MLE).

Figure 2 shows the Error Mean values for 100 repeated experiments for the three methods. The green line represents the proposed method (RRRY), and the red line represents the classical method (RRY). The blue line represents the classical method (MLE), which shows the efficiency of the proposed method compared to classical methods; the classical method (RRY) outperformed the classical method (MLE).

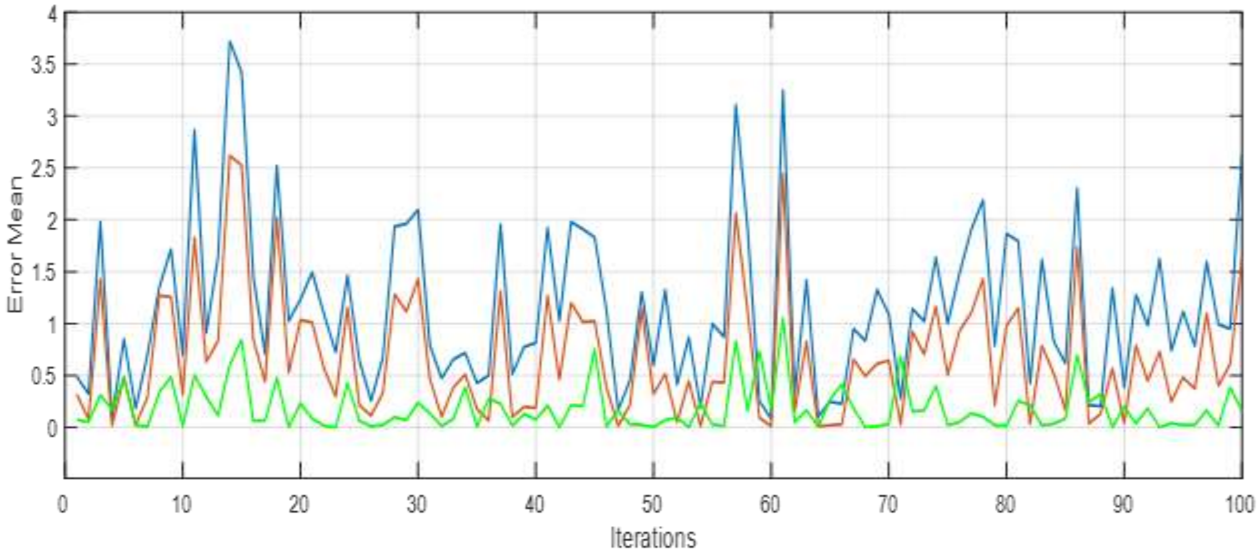

Figure 2: Error Mean for the three methods.

4. 1 Real data

The real data set examined the survival times, measured in days, of guinea pigs infected with virulent tubercle bacilli, as summarized by Bjerkedal (1960) [20]. The data used in this study, along with the MATLAB code for data analysis and implementation of the proposed method, are available from the corresponding author upon reasonable request. Figure 3 presents the Mahalanobis distance values for the FMCD method, revealing that 30 outliers were identified from a total of 74 generated data points.

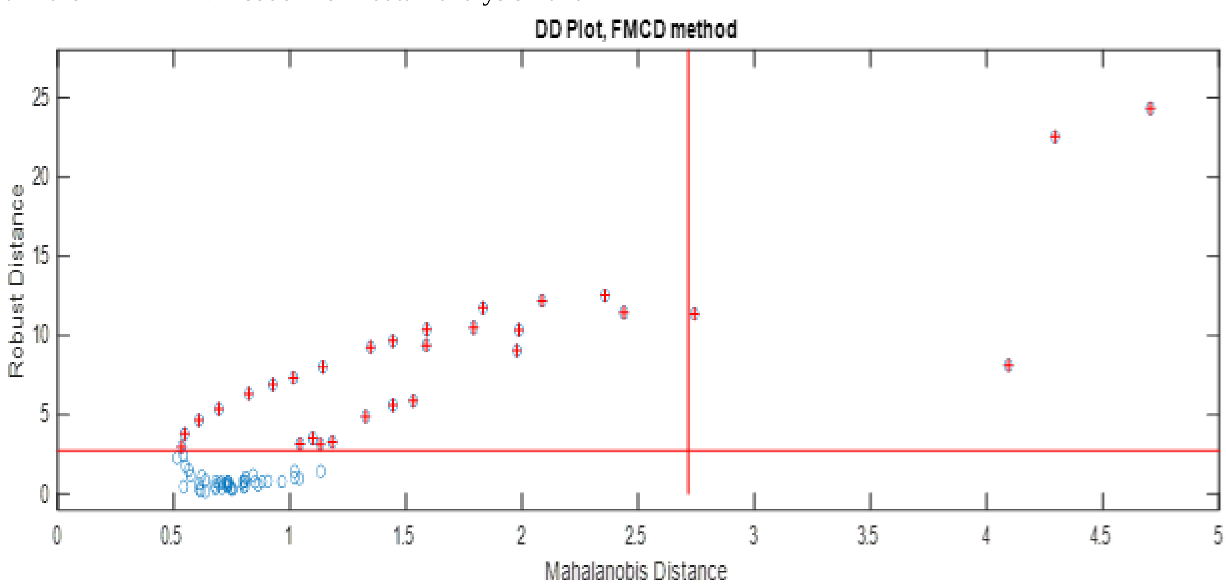

Figure 3: Mahalanobis distance for Real Data.

Following the estimation of the shape and scale parameters of the Weibull distribution using the three methods, these parameters are applied to estimate survival times (expected values). Subsequently, a goodness-of-fit test, specifically the Chi-Square test, is employed to evaluate the efficiency of the estimated models.

This evaluation measures the model's overall adequacy by contrasting the observed data against the predicted values generated by the model. The p-value derived from this test quantifies the statistical significance of the discrepancies between the observed and predicted values. By conducting statistical tests, we can thoroughly assess the validity of the proposed model. The model demonstrating the optimal fit is identified by the lowest

Chi-Square values and non-significant p-values derived from the test statistics, thereby rendering it the most appropriate choice for the specified data set.

Table 5 summarizes the estimation and testing results for the three methods, and shows that the proposed method (RRRY) was better than classical methods (RRY and MLE) because the value of the test statistic was equal to (10.627), which is less than the critical value (12.592) under significance level (0.05) and the degrees of freedom (6), (also, it less than test statistics 23.978 and 13.969 for classical models), and this is confirmed by the p-value (0.101), which was not significant, indicating the efficiency of the proposed model. While the classical models (RRY and MLE) were inefficient, because the value of the test statistic was equal to (23.978 and 13.969) respectively, which is more than the critical value (11.070) under significance level (0.05), and the degrees of freedom (5), and this is confirmed by the p-value (0.000 and 0.016) respectively, which was significant.

Method	Shape parameter (beta)	Scale parameter (eta)	Chi-Square Statistic	p-value	Critical Value	Degrees of freedom
MLE	1.3222	2.2383	23.978	0.000	11.070	5
RRY	1.6635	2.2141	13.969	0.016	11.070	5
RRRY	1.7123	1.9129	10.627	0.101	12.592	6

Figure 4 shows the probability density function of the Weibull distribution using the proposed method (RRRY).

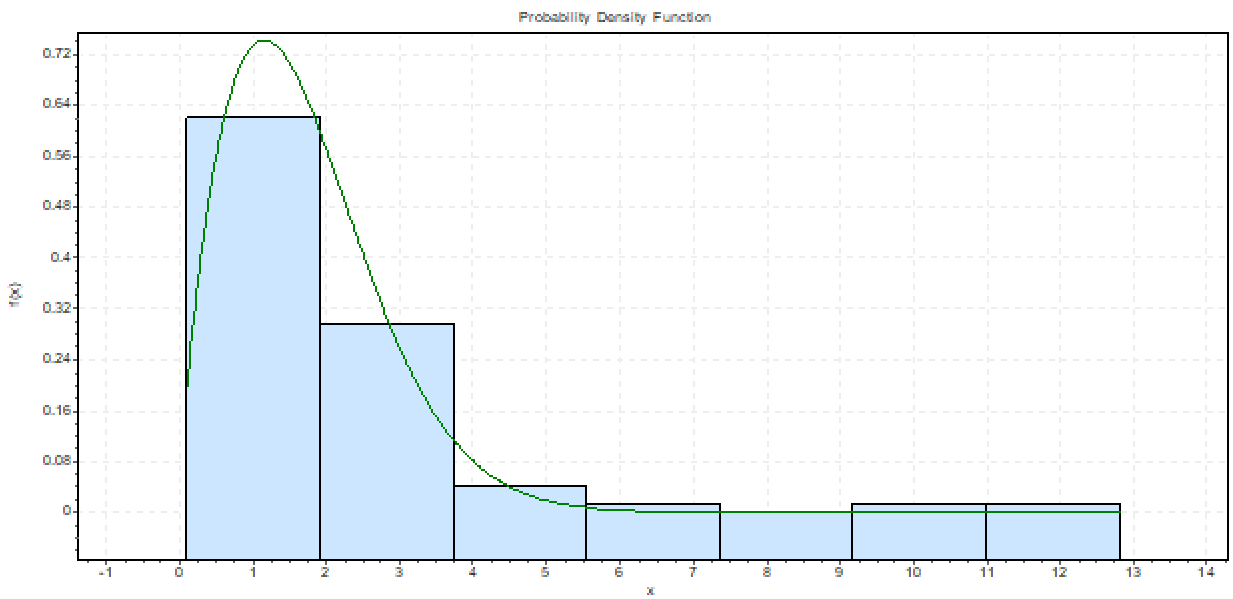

Figure 4: The probability density function of the Weibull distribution.

5 Conclusion & Recommendations

Based on the findings from the simulation study and the analysis of real data, the principal conclusions and recommendations have been distilled as follows:

5. 1 Conclusions

1. The proposed robust regression method based on the Fast Minimum Covariance Determinant (Fast MCD), denoted as RRRY, consistently outperformed the classical estimation methods — Maximum Likelihood Estimation (MLE) and traditional Rank Regression on Y (RRY) — in terms of Mean Squared Error (MSE) and Mean Absolute Error (Error Mean) across all simulation scenarios and in the real data application when outliers were present.
2. Across all simulation scenarios involving outliers, the estimation of the shape parameter of the Weibull distribution exhibited higher accuracy compared to the scale parameter for all three methods.
3. The accuracy of both shape and scale parameter estimates improved as the sample size increased, confirming the positive effect of larger sample sizes on parameter estimation precision in the presence of outliers.
4. The accuracy of estimating the shape parameter tended to decrease as its true assumed value increased, while the estimation accuracy of the scale parameter also decreased as its assumed value increased, under all simulation scenarios.
5. The estimation of the shape parameter remained relatively stable and accurate when the value of the scale parameter

increased, whereas the estimation accuracy of the scale parameter was influenced by changes in the assumed value of the shape parameter.

6. In terms of robustness to outliers, the proposed RRRY method consistently outperformed both classical methods (RRY and MLE). Furthermore, within the classical methods, RRY showed superior performance to MLE based on Error Mean under all simulation scenarios involving outlier contamination.

5. 2 Recommendations

1. Using the robust rank regression method to estimate a two-parameter Weibull distribution when there are outliers.

2. Executing a prospective study employing the robust rank regression technique to estimate the three-parameter Weibull distribution.

3. Conduct a prospective study using the robust rank regression method to estimate the parameters' normal and exponential distribution.

4. Although extensive simulation studies were conducted in this work to thoroughly validate the robustness and performance of the proposed method under various controlled conditions, the real-world application was limited to a classical historical dataset. Future research should consider applying this methodology to contemporary datasets from diverse application areas to further examine its effectiveness under modern data characteristics and reliability contexts.

References

[1] Raza, M. S., Ali, T. H. & Hassan, T. A. Using mixed distribution for gamma and exponential to estimate of survival function (brain stroke). *Polytech. J.* **8**, 1–9 (2018).

[2] Boudt, K., Caliskan, D. & Croux, C. Robust explicit estimators of Weibull parameters. *Metrika* **73**, 187–209, doi: 10.1007/s00184-009-0272-1 (2011).

[3] Coria, V. H., Maximov, S., Rivas-Dávalos, F. & Melchor-Hernández, C. L. Perturbative method for maximum likelihood estimation of the Weibull distribution parameters. *Springerplus* **5**, 1802, doi: 10.1186/s40064-016-3500-y (2016).

[4] Sinha, S. K. Robust estimation in accelerated failure time models. *Lifetime Data Anal.* **25**, 52–78, doi: 10.1007/s10985-018-9421-z (2019).

[5] Roohanizadeh, Z., Baloui Jamkhaneh, E. & Deiri, E. Parameters and reliability estimation for the Weibull distribution based on intuitionistic fuzzy lifetime data. *Complex Intell. Syst.* **8**, 4881–4896, doi: 10.1007/s40747-022-00720-x (2022).

[6] Gómez, Y. M., Gallardo, D. I., Marchant, C., Sánchez, L. & Bourguignon, M. An in-depth review of the Weibull model with a focus on various parameterizations. *Mathematics* **12**, 56, doi: 10.3390/math12010056 (2024).

[7] Safari, M. A. M., Masseran, N., Majid, M. H. A., Musa, M. A. & Kamarulzaman, K. Robust estimation of the three parameter Weibull distribution for addressing outliers in reliability analysis. *Sci. Rep.* **15**, 11516, doi: 10.1038/s41598-025-96043-1 (2025).

[8] Rousseeuw, P. J. & Leroy, A. M. *Robust Regression and Outlier Detection* (John Wiley & Sons, 1987).

[9] Butler, R. W., Davies, P. L. & Jhun, M. Asymptotics for the minimum covariance determinant estimator. *Ann. Stat.* **21**, 1385–1400 (1993).

[10] Ali, T. H., Sedeeq, B. S., Saleh, D. M. & Rahim, A. G. Robust multivariate quality control charts for enhanced variability monitoring. *Qual. Reliab. Eng. Int.* **40**, 1369–1381, doi: 10.1002/qre.3472 (2024).

[11] Omar, C., Ali, T. H. & Hassn, K. Using Bayes weights to remedy the heterogeneity problem of random error variance in linear models. *Iraqi J. Stat. Sci.* **17**, 58–67 (2020).

[12] Murali, K. A. *Life Data Analysis Reference* (ReliaSoft Corporation, 2016).

[13] Elias, I. I. & Ali, T. H. Optimal level and order of the Coiflets wavelet in the VAR time series denoise analysis. *Front. Appl. Math. Stat.* **11**, 1526540, doi: 10.3389/fams.2025.1526540 (2025).

[14] David, J. S. *Reliability Engineering* (Pitman Publishing, 1972).

[15] Hayawi, H. A., Azeez, S. M., Babakr, S. O. & Ali, T. H. ARX time series model analysis with wavelets shrinkage (simulation study). *Pak. J. Statist.* **41**, 103–116 (2025).

[16] Samad Sedeeq, B., Muhammad, Z. A., Ali, I. M. & Ali, T. H. Construction robust -chart and compare it with Hotelling's T^2-chart. *Zanco J. Human Sci.* **28**, 140–157, doi: 10.21271/zjhs.28.1.11 (2024).

[17] Omer, A. W., Sedeeq, B. S. & Ali, T. H. A proposed hybrid method for multivariate linear regression model and multivariate wavelets (simulation study). *Polytech. J. Humanit. Soc. Sci.* **5**, 112–124, doi: 10.25156/ptjhss.v5n1y2024.pp112-124 (2024).

[18] Ali, T. H., Hamad, A. A., Mahmood, S. H. & Ahmed, K. H. ARIMAX time series analysis for a general budget in the Kurdistan Region of Iraq using wavelet shrinkage. *Commun. Stat. Case Stud. Data Anal. Appl.* **1**, 1–25, doi: 10.1080/23737484.2025.2486984 (2025).

[19] Douglas, C. M. *Introduction to Statistical Quality Control*, 6th edn. (John Wiley & Sons, 2009).

[20] Bjerkedal, T. Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *Am. J. Hyg.* **72**, 130–148 (1960).