AIMS Mathematics, 10(11): 25639–25666.
DOI: 10.3934/math.20251136
Received: 17 August 2025
Revised: 03 October 2025
Accepted: 17 October 2025
Published: 06 November 2025

https://www.aimspress.com/journal/Math

Research article

An improvement in predictive modeling techniques with application to pivotal quantity and least square method

M. H. Harpy[1,*], O. M. Khaled[2], Mahmoud El-Morshedy[1] and K. S. Khalil[3]

[1] Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[2] Department of Mathematics, Faculty of Science, Port said University, Port Said, Egypt
[3] Department of basic science, Faculty of computer and informatics, Suez Canal University, Ismailia 41522, Egypt

* **Correspondence:** Email: mahmoud_haussieny@science.suez.edu.eg.

Abstract: In this research, we develop a new method for predicting order data. Our approach involves selecting the best-fitting distribution through different tests, estimating its parameters, and constructing prediction intervals that leverage observed and predicted data. In this method, we entered the predicted data one by one, along with the observed data. At each step, we found a suitable distribution and then estimated its parameters and applied the prediction method, such as pivotal quantity and modified least square with cumulative hazard function. We implemented the new method using the R programming language and conducted comparative analyses against several established methods across datasets, encompassing health insurance coverage, glass fiber strength, and COVID-19 recovery rates. The results demonstrated this method's superior performance, particularly in terms of Mean square error (MSE) and coefficient of variation (CV), as well as its ability to predict more data and outperform traditional methods in most scenarios. This method has the ability to obtain a large number of predicted observations to reach about 150% to 200% of the real observations, as explained through a simulation study and real data.

Keywords: order statistics; predictive model; predictive interval; point predictor
Mathematics Subject Classification: 62-08, 62G30, 62E15

1. Introduction

Prediction of future observations is an issue of major importance in many applications of reliability theory. Point prediction is an essential tool for predicting future observations and is widely used in reliability theory and lifetime problems. For many years within reliability theory, especially in life

testing experiments, researchers have focused largely on point forecasting using a limited number of methods such as the least square method, maximum likelihood predictor, unbiased optimal predictor, conditional median predictor, and Bayesian predictor. A new method for establishing a point predictor for the future order statistics in the case of the sample size as a random variable (RV) was proposed by Barakat et al. [1].

In the field of reliability theory, the point and interval predictions for ordered RVs are used widely, and industrial applications and survival studies for predicting the number of items may be defective during the future production process. An example for this is that, the point predictor is very useful to choose a suitable censoring scheme in the experiments of lifetime testing. In such experiments, we can test all items (n items, say) and then we wait until getting a reasonable (with respect to their cost and time) number of items that failed ($r < n$, say). After that, and based on the observed failure times, we can predict the other failure times of the survivor items. Based on this information, we can choose the suitable censoring scheme, (e.g., Type I or Type II censoring) and the time or the sth (say $s > r$) future failed item, at which the test must be terminated. The reader can review several papers that entail the problems of prediction involving order statistics, such as Barakat et al. [2] and Patel [3].

Morever, Aly et al. [4] proposed a novel least squares approach for estimation and prediction that relies on cumulative hazard function.

The pivotal quantities when used in conjunction with order statistics gives a powerful statistical technique to construct predictive intervals (PIs). Two pivotal statistics were developed by Barakat et al. [2] to construct prediction intervals of future observations in the case of fixed sample size; and if the sample size is a positive integer valued RV, independent of the observations, they also derived explicit forms for the distribution functions (CDFs) of the used pivotal statistics. Moreover, Khaled et al. [5] did the same work when the mixture of two gamma distributions was used, and in the case of the mixture of two beta distributions.

In this paper, we introduce a new method for predicting future observations. Let $X_1, X_2, ..., X_n$ be n observations. First, we select the best-fitting distribution (some of fitting distributions will be present in Section 3) through different tests for these observations. Then, we estimate its parameters and construct prediction intervals, considering the new predicted value as Y_1. This process is repeated for each new prediction: To predict Y_2, we re-fit the model to the extended dataset $X_1, X_2, \ldots, X_n, Y_1$, updating the distribution and its parameters. Similarly, we use the dataset $X_1, X_2, \ldots, X_n, Y_1, Y_2$ to predict Y_3, and so on. This method is more accurate and enables us to predict up to 100% of the data. However, in this paper, we propose that the lower bound of the next interval should be the estimated value of the last observation. Model selection was performed using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) by fitting data using minimum AIC and BIC for all distributions (see Arijit Chakrabarti [6]). In Section 3, we give comprehensive simulation studies for the new method. Finally, in Section 4, five real data sets are analyzed for illustrative and comparison purposes.

2. Auxiliary results

In this section, we employ two auxiliary methods to enhance the suggested methodology. First, we apply the approach of Barakat et al. [2], which utilizes a pivotal quantity for constructing confidence intervals. Second, we incorporate the technique described by Aly et al. [4], which employs the least

squares method to predict future observational data points. The theorems underpinning the proposed method, along with the corresponding R code, are presented below.

2.1. Pivotal quantity

The proposed method relies on the results of Barakat et al. [2], which is re-coined in the following lemmas:

Lemma 2.1. *Let $X_i,\ i = 1,\ 2,\ ...,\ n$ be a random sample (i.i.d) of size n have an absolutely continuous distribution function F and let $X_{1:n} < X_{2:n} < ... < X_{n:n}$ denote the corresponding order statistics. Moreover, let $X_i^* = -\log(1 - F(X_i)),\ i = 1, 2, ..., n$, and $X_{i:n}^* = -\log(1 - F(X_{i:n})),\ i = 1, 2, ..., n$. Then the normalized spacing*

$$Z_i = (n - i + 1)(X_{i:n}^* - X_{i-1:n}^*),\ i = 1, 2, ..., n,\ (X_{0:n} \equiv F^{-1}(0+)) \tag{2.1}$$

are i.i.d RVs, from the standard exponential distribution $EXP(1)$.

Barakat et al. [2] (specially, Lemma 2.1, p. 2) suggested the pivotal quantity $U_{r,s;n} = \frac{X_{s:n}^* - X_{r:n}^*}{X_{r:n}^*}$ and $V_{r,s;n} = \frac{X_{s:n}^* - X_{r:n}^*}{T_{r:n}},\ r < s$, where $X_{i:n}^* = -\log(1 - F(X_{i:n})),\ i = 1, 2, ..., n.$ and $T_{r:n} = \sum_{j=1}^{r} Z_j$ represents the total time in a life test and $Z_j,\ j = 1, 2, ..., r$, are defined by (2.1) to construct the PI for $X_{s:n}$, as well as $X_{s:N}$, where N is a positive integer-valued RV independent of the basic RVs $X_i,\ i = 1,\ 2,\ ...,\ n$. The following lemma (Theorem 2.5 in Barakat et al. [2], p. 5) explains the previous result.

Lemma 2.2. *For a positive integer-valued RV N, which is independent of $X_i,\ i = 1,\ 2,\ ,..., n$, the CDF $F_{V_{r,s;N}}$ of the pivotal quantity $V_{r,s;N}$ is given by*

$$F_{V_{r,s;N}}(v) = 1 - \frac{1}{(s - r - 1)!\ P(N \geq s)} \sum_{n=s}^{\infty} \sum_{i=0}^{s-r-1} (-1)^i \binom{s - r - 1}{i} \frac{(n - r)!}{(n - s)! \eta_i(n, s)} \times [1 + v\eta_i(n, s)]^{-r} P(N = n), \quad v \geq 0, \tag{2.2}$$

where $\eta_k(n, t) = n + k + 1 - t$. Hence, the equation $F_{V_{r,s;N}}(v) = 1 - \alpha$ can be solved numerically to get the quantile value v for any given α.

Additionally, $[x_{r:N}, F^{-1}(1 - e^{-v\ t_{r:N}} \overline{F}(x_{r:N}))]$ is $(1 - \alpha)100\%$ PI for $X_{s:N}$. Moreover, if N is a fixed integer (i.e., $P(N = n) = 1$), $n \geq s$, then

$$F_{V_{r,s;n}}(v) = 1 - B_{r,s}(n) \sum_{i=0}^{s-r-1} (-1)^i \binom{s - r - 1}{i} [\eta_i(n, s)(1 + v\eta_i(n, s))]^{-1}, \quad V \geq 0, \tag{2.3}$$

where $B_{r,s}(n) = (n - 1)![(s - r - 1)!(n - s)!]^{-1}$.

We apply [2, Theorem 2.5] to the suggested method for a sample from a family of distributions like wiebull, exponential, and gamma. In this paper, we investigate the three issues for the usage of this method through an extensive simulation study that is carried out using R-Package:

1. Find the quantile function for the distribution.
2. Create a confidence interval for the given observation using the quantile function and propose a multi-point predictor for fixed sample sizes. The confidence interval will provide a range of potential values for the observation, while the point predictor will offer a precise estimate of its most probable value.
3. Generate parameter estimates for the distribution, considering each predictive data point.

2.2. Modified least squares with cumulative Hazard function

In the field of estimation and prediction for generalized ordered statistics, Aly et al. [4] introduced a novel application of the cumulative hazard function suitable for any continuous distribution. The innovative method employed makes use of the Rényi representation and is adaptable for complete data and type II right-censored data. Rigorous simulation experiments were conducted to validate the efficiency of the proposed procedures. The methods were also compared against the maximum likelihood and ordinary weighted least squares (WLS) methods, with evaluations based on the root mean squared error (RMSE) and Pitman's measure of closeness (PMC). The practicality of the proposed methods was further substantiated through the analysis of two real data sets. Moreover, we apply the suggested method to this real data and compare the results.

An approximate modified weighted least square estimate of Θ based on the first r observed GOSs $\widetilde{x}_r = (x_1, x_2, ..., x_r)$ for $r \leq n$ can be obtain by minimizing the function

$$\mathcal{WL}^*_{H,r}(\Theta|\widetilde{x}_r) = \sum_{i=1}^{r} w^*_i (H(x_i;\Theta) - \mu^*_{i:n})^2 + (n-r) w^*_r (H(x_r;\Theta) - \mu^*_{r:n})^2, \tag{2.4}$$

with respect to Θ, where $\widetilde{x}_r = (x_1, x_2, ..., x_r)$ are observation values of the GOSs model. Approximate modified weighted least square estimates of $x_{s:n}$ based on $(x_1, x_2, ..., x_r)$ are derived by minimaizing the predictive weighted least square function

$$\mathcal{PWL}^*_{H,r,s}(\Theta, x_s|\widetilde{x}_r) = \sum_{i=1}^{r} w^*_i (H(x_i;\Theta) - \mu^*_{i:n})^2 + (s-r+1) w^*_r (H(x_r;\Theta) - \mu^*_{r:n})^2 + (n-s+1) w^*_s (H(x_s;\Theta) - \mu^*_{s:n})^2. \tag{2.5}$$

2.3. Exploring distributions in simulation: Methods and applications

We present a concise review of Weibull and Modified Kies-Exponential and Gompertez distributions, were used in Aly et al. [4] as an application of prediction observation using the modified least square method with the hazard function.

2.3.1. Weibull distribution

The Weibull distribution is a continuous probability distribution that is widely used in statistics, engineering, and scientific fields to model the distribution of lifetimes, failure times, or survival times for processes, products, and systems. It is named after Wallodi Weibull, a Swedish engineer, who introduced it in the mid-20th century.

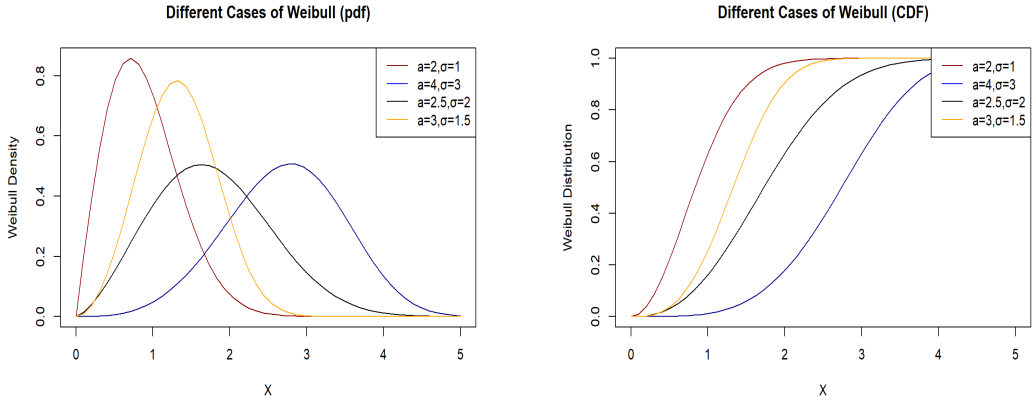
The Weibull distribution is characterized by two parameters: The shape parameter (a) and the scale parameter (σ). These parameters define the shape and scale of the distribution, making it a flexible model for a wide range of real-world applications. Rahmouni and Ziedan [7] introduced the Weibull-generalized shifted geometric (WGSG) distribution, including its properties, estimation methods, and applications, showing improvements over classical lifetime models in handling challenging hazard behaviors. This distribution was used by Magdy et al. [8] to fit data sets that consist of voltage levels at which failures occur in a certain type of electrical cable, the test involved 20 specimens. The probability density function (pdf) of the Weibull distribution is given by:

$$f(x;a,\sigma)=\begin{cases}\frac{a}{\sigma}\left(\frac{x}{\sigma}\right)^{a-1}e^{-(x/\sigma)^a}, & x\geq 0,\\ 0, & x<0,\end{cases}$$

where, x is the RV representing the time to failure or survival time. Parameter a is the shape parameter, which determines the shape of the distribution curve; if $a>1$, the distribution is right-skewed (increasing hazard rate); if $a<1$, it is left-skewed (decreasing hazard rate) and when $a=1$, it reduces to the exponential distribution. Parameter σ is the scale parameter, which represents the characteristic life of the distribution, the point at which about 63.2% of the units have failed (see also Rinne [9] and Karolczuk and Palin-Luc [10]). The Weibull CDF is given by:

$$F(x;a,\sigma)=\begin{cases}1-e^{-(x/\sigma)^a}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

Figure 1(a) shows the Weibull pdf at different cases of the shape and scale parameters. Figure 1(b) shows the Weibull CDF at the same values of these parameters.



(a) pdf of weibull (b) CDF of weibull

Figure 1. The plot of pdf and CDF of weibull distribution.

We apply the method of pivotal quantity for some data set from the weibull distribution to find a PI. Moreover, due to Lemma 2.2, let $\mathcal{V}_{r,s+t-1,n}$ be the upper bound of PI, such that

$$\mathcal{V}_{r,s+t-1,n}=Q(1-e^{-vt_{r:n}}\overline{F}(X_{r:n};a;\sigma)). \tag{2.6}$$

Remark 2.1. *In this paper, we suggest a new technique to choose a predicting point, such that: If* $x_i\leq 1,\forall\ i=1,\ 2,...,\ r$ *we choose the predicting point as the average of (median, harmonic mean, geometric mean, third decile) for the PI.*

Otherwise, if $x_i\geq 1,\forall\ i=1,\ 2,...,\ r$, *we choose the predicting point as the third decile of the PI.*

It is worth mentioning that the method we use for predicting the parameters is the maximum likelihood.

2.3.2. The modified Kies-exponential (MKE) distribution

It is known that the exponential distribution has an important property, which is the lack of memory, making it one of the important classical distributions, in addition to it is analytically tractable distribution. However, it has some limited applications because of its fixed hazard rate and unimoda pdf. For that, several extensions of the exponential distribution are considered to increase its flexibility and applicability. One of these extensions is the modified Kies-exponential (MKE) (see, Babtain et al. [11] and Aly et al. [4]). MKE distribution has many applications in various fields such as reliability engineering: Modeling time-to-failure of components or systems with non-constant failure rates, survival analysis: Analyzing time until an event of interest occurs, accounting for different hazard functions, and queuing theory: Modeling interarrival times or service times in queuing systems with complex arrival patterns.

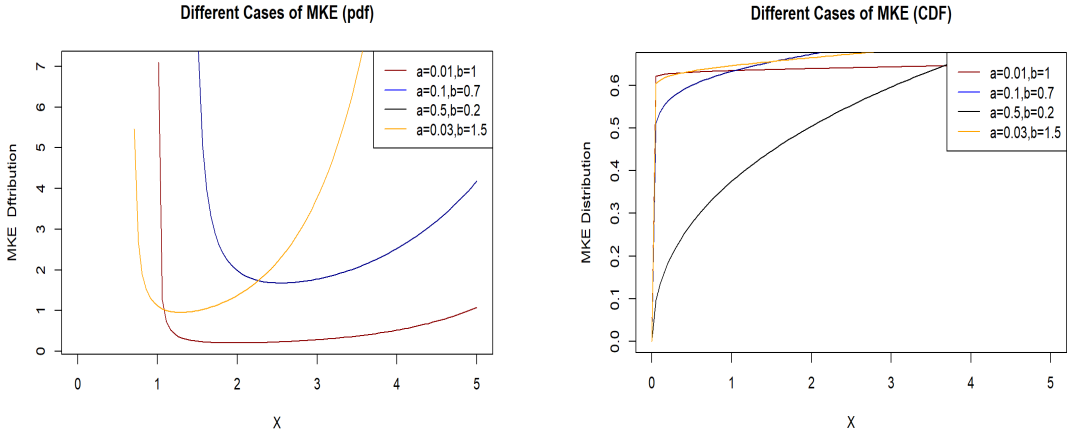
The pdf and CDF of the MKE distribution are given by:

$$g(x;a,b) = a\,b\,e^{-(e^{bx}-1)^a}\,e^{bx}\,(e^{bx}-1)^{a-1},$$

$$G(x;a,b) = 1 - e^{-(e^{bx}-1)^a}, \quad x > 0,$$

where a is the shape parameter and b is the scale parameter.

Figure 2(a) shows the MKE pdf at different cases of the shape and scale parameters. Figure 2(b) shows the MKE CDF at the same values of these parameters.



(a) pdf of MKE (b) CDF of MKE

Figure 2. The plot of pdf and CDF of MKE distribution.

2.3.3. Gompertz-Makeham distribution

Aly et al. [4] used the data reported by Hoel et al. [12] to apply their method, and the adequate distribution for fitting these data was the Gompertz-Makeham distribution. For the applications of this distribution, it can be used in actuarial science: Modeling human mortality rates for life insurance and pension schemes, demography: Analyzing population dynamics and life expectancy trends, and biostatistics: Studying survival analysis and disease progression. Another recent work is the discrete Gompertz-Makeham distribution for multidisciplinary data analysis by Elshahhat et

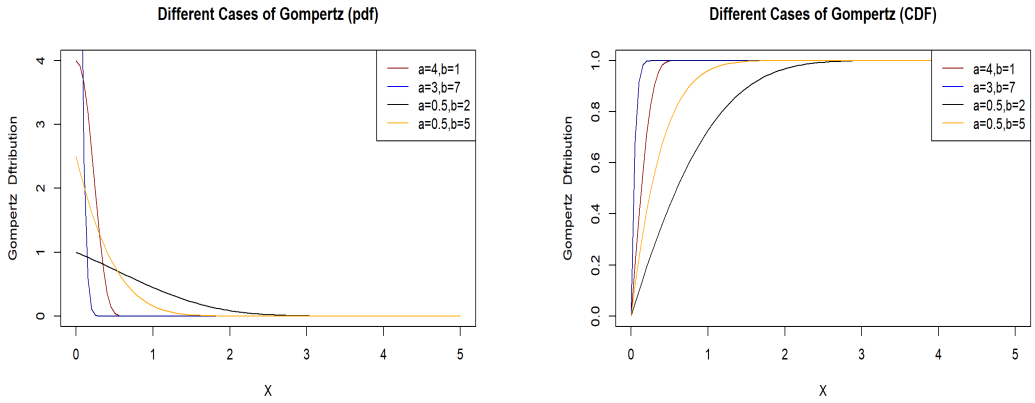
al. [13], which develops a discrete analogue of the Gompertz-Makeham distribution, studies its quantiles, order statistics, skewness, kurtosis, and estimates its parameters using likelihood-based and Bayesian methods under censored data, showing its usefulness compared to discrete lifetime models encompassing Gompertz-Makeham distribution for multidisciplinary data analysis. The pdf and CDF of this distribution are defined as:

$$f(x; a, b) = a\, b\, e^{b(1-e^{ax})}\, e^{ax}, \quad x \geq 0,$$

$$F(x; a, b) = 1 - e^{b(1-e^{ax})}, \quad x \geq 0,$$

where a is the scale parameter and b is the shape parameter.

Figure 3(a) shows the Gompertz-Makeham pdf at different cases of the scale and shape parameters. Figure 3(b) shows the Gompertz-Makeham CDF at the same values of these parameters.



(a) pdf of Gompertz-Makeham (b) CDF of Gompertz-Makeham

Figure 3. The plot of pdf and CDF of Gompertz-Makeham distribution.

3. Major results

In this section, we apply the new methodology to various families of distributions, including Weibull, gamma, and exponential distributions. We conduct simulations on data derived from the Weibull distribution with varying parameters, treating it as a special case. Moreover, we will apply the new method using the same data used before by Magdy et al. [8], which applies a pivotal quantity, enabling us to perform a comparative analysis demonstrating the efficacy of the proposed method. Initially, we present a concise review of the Weibull distribution. Additionally, we present a concise review of Modified Kies-Exponential and Gompertez distributions, which were used in Aly et al. [4] as application of prediction observation using modified least square method with hazard function. Furthermore, we employ the maximum likelihood estimation method for parameter estimation of each observation, which is built into the R programming environment. The details of our algorithm and the results are displayed below.

3.1. Simulation study

In this section, we apply the suggested methodology to simulate data from a Weibull distribution. The dataset consists of 24 observations, with a planned interruption after the 12th observation. Our objective is to utilize the first 12 observations to make predictions for the remaining 12. We apply the method 1000 times, each time with a different data from the same distribution as the previous ones. Ultimately, we calculate the average of the 1000 observations. In the simulation study, the scale and shape parameters of the Weibull distribution are randomly chosen across runs. This strategy is adopted to generate a wide range of scenarios and to evaluate the robustness of the proposed prediction method under varying parameter settings.

Algorithm I: Pivotal quantity for a fixed sample

1. Determine the PIs, by choosing the value of confidence interval α. By solving the equation $F_{V_{r,s;N}}(v) = 1 - \alpha$, we get the quantile value u for a given α, (see Eq (2.3)),
2. fit new data and select a new distribution with minimum BIC and AIC,
3. estimate parameter for the data from x_1 to x_r,
4. use Eq (2.6) to evaluate the upper bound of interval,
5. compute the point predictions, $\hat{x}_{s:n},\ s = r + 1$,
6. rearrange data from x_1 to $\hat{x}_s$ to predict the next observation $\hat{x}_{s+1}$,
7. repeat steps 1–6, 1000 times,
8. compute the averages of $x_{s:n}$, $\hat{x}_{s:n}$ and PI over 1000,
9. compute the MSE and CV of the point predictor $\hat{x}_s$.

Algorithm II: Pivotal quantity for a random sample

1. Determine the PIs, by choosing the value of confidence interval α. By solving the equation $F_{V_{r,s;N}}(v) = 1 - \alpha$, we get the quantile value u for a given α, Eq (2.3),
2. generate a random integer from $B(l, p) = Bin(40, 0.8)$, and Poisson $P(\lambda) = (50)$ distributions say n_z,
3. generate an ordered random sample of size n_z from F,
4. solve the nonlinear equations $F_{V_{r,s:n}}(v) = 1 - \alpha$ by using Eq (2.2) in Lemma 2.2 after replacing ∞ by 40 at the Binomial case and 50 Poisson case,
5. estimate the parameter for simulation data from $x_{1:n}$ to $x_{r:n}$,
6. use Eq (2.6) to evaluate the upper bound of the interval,
7. compute the point predictions, $\hat{x}_{s:n},\ s = r + 1$,
8. replace $x_{r:n}$ by $\hat{x}_{s:n}$ to predict the next observation,
9. obtain a new parameter estimate using fresh data ranging from $x_{1:n}$ to $\hat{x}_{s:n}$,
10. repeat steps 1–9, 1000 times,
11. compute the averages of $x_{s:n}$, $\hat{x}_{s:n}$, and PI over 1000,
12. compute the MSE and CV of the point predictor $\hat{x}_{s:n}$.

Algorithm III: The least square method

1. Select a data set and fit based on minimum AIC and BIC,
2. estimate the parameter for a distribution using Eq (2.4),
3. use Eq (2.5) to compute the point predictions $\hat{x}_{s:n}\ \ s = r + 1$,

4. replace $x_{r:n}$ by $\hat{x}_{s:n}$ to predict the next observation,
5. refit the dataset by incorporating the prior predicted values,
6. obtain a new parameter estimate using fresh data ranging from $x_{1:n}$ to $x_{r:n}$,
7. compute the MSE and CV of the point predictor $\hat{x}_{s:n}$.

The algorithm is employed in a simulation study conducted using the R programming language (version 4.4.2 (2024-10-31)), and the outcomes of the simulation are presented in the following tables.

Remark 3.1. *The CV values reported in the tables are calculated based on the predicted values $\hat{x}_{s:n}$ presented in each table. They represent the variability of the predicted values and should not be interpreted as direct indicators of the predictive performance of the proposed methods. Relatively high CV values therefore reflect the dispersion of the predicted sequence, than the limitations of the methods.*

3.2. Simulation pivotal quantity

In the following two tables, we apply the proposed method and achieve remarkable results. This approach enables us to predict 100% of the obtained data, in contrast to previously used methods, which predict at most 35% of the data. This is because the previous methods consider the minimum prediction horizon as the last data point, which remains constant across all prediction intervals for future data. Furthermore, we apply the method to predict more than 100% of the data we have in the following tables, extending our predictions beyond our existing data-set.

In Tables 1 and 2, we observe the robustness of the employed methodology in prediction. In the fourth column, representing the difference between the actual and predicted observations, we note an exceptionally low error percentage. Additionally, an overall MSE is computed for the data in Table 1, yielding a remarkably small value.

Table 1. Predict 12 future observation 100% prediction from weibull $a = 4, \sigma = 3$.

r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI
12	2.759	2.730	0.029	(2.679, 2.841)
13	2.842	2.773	0.069	(2.730, 2.865
14	2.926	2.819	0.107	(2.773, 2.918)
15	3.011	2.867	0.144	(2.819, 2.971)
16	3.096	2.917	0.179	(2.867, 3.026)
17	3.183	2.970	0.212	(2.917, 3.086)
18	3.274	3.027	0.247	(2.970, 3.151)
19	3.375	3.089	0.286	(3.027, 3.225)
20	3.485	3.158	0.327	(3.089, 3.311)
21	3.612	3.239	0.373	(3.158, 3.421)
22	3.771	3.342	0.429	(3.239, 3.579)
23	3.991	3.495	0.496	(3.342, 3.877)
MSE=0.0778, CV=0.0748				

Table 2. Predict 24 future observation 200% prediction from weibull $a = 2, \sigma = 15$.

r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI
12	9.734	9.534	0.200	(9.237, 10.225)
13	10.202	9.742	0.461	(9.534, 10.200)
14	10.677	9.959	0.718	(9.742, 10.440)
15	11.146	10.178	0.968	(9.959, 10.663)
16	11.609	10.398	1.210	(10.178, 10.884)
17	12.097	10.619	1.478	(10.398, 11.105)
18	12.591	10.841	1.751	(10.619, 11.330)
19	13.079	11.065	2.015	(10.841, 11.557)
20	13.603	11.291	2.312	(11.065, 11.789)
21	14.127	11.522	2.605	(11.291, 12.027)
22	14.652	11.757	2.895	(11.522, 12.272)
23	15.223	11.997	3.226	(11.757, 12.527)
24	15.788	12.245	3.543	(11.997, 12.788)
25	16.387	12.501	3.885	(12.245, 13.066)
26	17.027	12.769	4.258	(12.501, 13.360)
27	17.707	13.051	4.656	(12.769, 13.675)
28	18.431	13.350	5.081	(13.051, 14.016)
29	19.217	13.673	5.544	(13.350, 14.393)
30	20.117	14.025	6.091	(13.673, 14.819)
31	21.078	14.419	6.659	(14.025, 15.316)
32	22.199	14.872	7.326	(14.419, 15.920)
33	23.604	15.419	8.185	(14.872, 16.715)
34	25.409	16.133	9.276	(15.419, 17.909)
35	27.931	17.241	10.690	(16.133, 20.319)
MSE=23.74, CV=0.166				

Random data is generated in Tables 1 and 2. In the first table, the data follow a Weibull distribution with parameters $a = 4$ and $\sigma = 3$ and the method is applied to predict 100% of the available data. The results showcase a highly accurate prediction.

In Table 2, random data is again generated, this time with parameters $a = 2$ and $\sigma = 15$ from the same Weibull distribution. Attempting to predict 200% of the acquired data, the method is applied, and the MSE is computed, providing insights into the predictive performance.

This approach enables us to predict 100% of the obtained data, in contrast to previously used methods, which predict at most 35% of the data. This is because the previous methods consider the minimum forecast horizon as the last data point, which remains constant across all forecasting intervals for future data. Furthermore, we apply the method to predict more than 100% of the data we have in the following tables, extending our predictions beyond our existing data-set.

Notably, the relatively large MSE (close to 200%) obtained in this case is mainly due to the high value of the scale parameter of the Weibull distribution (15), while the final biase yields a scale around 10.6. To verify this effect, we re-generate data with a smaller scale parameter (5), and the resulting MSE is considerably lower. This indicates that the scale parameter has a strong impact on the

performance of the method (see Table 3).

Table 3. Predict 24 future observation 200% prediction from weibull $a = 2, \sigma = 5$.

r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI
12	3.245	3.178	0.067	(3.079, 3.408)
13	3.401	3.247	0.154	(3.178, 3.400)
14	3.559	3.320	0.239	(3.247, 3.480)
15	3.715	3.393	0.323	(3.320, 3.554)
16	3.870	3.466	0.403	(3.393, 3.628)
17	4.032	3.540	0.493	(3.466, 3.702)
18	4.197	3.614	0.583	(3.540, 3.777)
19	4.360	3.688	0.672	(3.614, 3.852)
20	4.534	3.764	0.771	(3.688, 3.930)
11	4.709	3.841	0.868	(3.764, 4.009)
22	4.884	3.919	0.965	(3.841, 4.091)
23	5.074	3.999	1.075	(3.919, 4.176)
24	5.263	4.082	1.181	(3.999, 4.263)
25	5.462	4.167	1.295	(4.082, 4.355)
26	5.676	4.256	1.419	(4.167, 4.453)
27	5.902	4.350	1.552	(4.256, 4.558)
28	6.144	4.450	1.694	(4.350, 4.672)
29	6.406	4.558	1.848	(4.450, 4.798)
30	6.706	4.675	2.030	(4.558, 4.940)
31	7.026	4.806	2.220	(4.675, 5.105)
32	7.400	4.957	2.442	(4.806, 5.307)
33	7.868	5.140	2.728	(4.957, 5.572)
34	8.470	5.378	3.092	(5.140, 5.970)
35	9.310	5.747	3.563	(5.378, 6.773)
MSE=2.638				

In Table 4, the suggested method is applied under the scenario where the sample size is an RV following either a two-sided binomial or a Poisson distribution with parameter $B(l, p) = (40, 0.8)$ and $P(\lambda) = (50)$ simultaneously. Random data is generated from a Weibull distribution with parameters $a = 4, \sigma = 7$. The method is employed, and results are obtained by predicting a single point, taking into consideration that the sample size varies randomly.

Table 4. $r = 12$ and predict 100% of data weibull $a = 4, \sigma = 7$.

binomial distribution $B(l, p) = (40, 0.8)$					Poisson $P(\lambda) = (50)$ distribution				
r	$x_{s:N}$	$\hat{x}_{s:N}$	biase	PI	r	$x_{s:N}$	$\hat{x}_{s:N}$	biase	PI
12	5.841	5.697	0.144	(5.688, 5.706)	12	5.132	5.010	0.122	(5.007, 5.014)
13	5.992	5.702	0.290	(5.697, 5.707)	13	5.247	5.013	0.234	(5.010, 5.016)
14	6.141	5.707	0.434	(5.702, 5.712)	14	5.365	5.016	0.349	(5.013, 5.019)
15	6.281	5.712	0.570	(5.707, 5.716)	15	5.472	5.019	0.453	(5.016, 5.021)
16	6.424	5.716	0.708	(5.712, 5.721)	16	5.570	5.021	0.549	(5.019, 5.023)
17	6.572	5.721	0.851	(5.716, 5.725)	17	5.674	5.023	0.651	(5.021, 5.025)
18	6.715	5.725	0.990	(5.721, 5.729)	18	5.778	5.025	0.753	(5.023, 5.027)
19	6.858	5.729	1.129	(5.725, 5.733)	19	5.877	5.026	0.851	(5.025, 5.028)
20	6.999	5.733	1.266	(5.729, 5.737)	20	5.983	5.028	0.954	(5.026, 5.030)
21	7.149	5.737	1.412	(5.733, 5.741)	21	6.077	5.029	1.047	(5.028, 5.030)
22	7.304	5.741	1.563	(5.737, 5.745)	22	6.167	5.030	1.136	(5.029, 5.031)
23	7.471	5.745	1.726	(5.741, 5.749)	23	6.258	5.031	1.226	(5.030, 5.033)
MSE= 1.093 ,CV= 0.002					MSE= 0.601 , CV=0.001				

When predicting future observations for the data, given the fluctuating sample size, it becomes uncertain whether the sample size includes these observations or not. This introduces an additional layer of complexity in interpreting and understanding the predictive outcomes.

3.3. Simulation modifies least square with cumulative hazard function

In the following table (Table 5), random data is generated using the Weibull distribution with specified parameters $a = 3$ and $\sigma = 5$. We apply our technique using the modified least squares method that we mention algorithm III.

Table 5. Predict 60 future observation 150% prediction from weibull $a = 3, \sigma = 5$.

r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	70	5.3728	5.2184	0.1544
40	4.0224	3.9826	0.0398	71	5.3807	5.2645	0.1162
41	4.1152	4.0236	0.0916	72	5.4475	5.3114	0.1360
42	4.1653	4.0643	0.1010	73	5.4908	5.3591	0.1317
43	4.2060	4.1048	0.1012	74	5.5962	5.4077	0.1885
44	4.2374	4.1451	0.0923	75	5.6096	5.4572	0.1523
45	4.3054	4.1852	0.1202	76	5.6678	5.5078	0.1599
46	4.3760	4.2252	0.1508	77	5.6756	5.5596	0.1160
47	4.5196	4.2651	0.2545	78	5.7408	5.6126	0.1282
48	4.5302	4.3050	0.2252	79	5.7622	5.6670	0.0952
49	4.5667	4.3448	0.2219	80	5.8203	5.7230	0.0973
50	4.5702	4.3845	0.1857	81	5.9356	5.7806	0.1550
51	4.6103	4.4243	0.1860	82	6.0169	5.8402	0.1767
52	4.6244	4.4641	0.1603	83	6.1720	5.9018	0.2702
53	4.6451	4.5040	0.1411	84	6.2107	5.9659	0.2449
54	4.6723	4.5439	0.1284	85	6.2427	6.0326	0.2102
55	4.6825	4.5840	0.0985	86	6.2729	6.1023	0.1706
56	4.7039	4.6242	0.0797	87	6.3360	6.1755	0.1605
57	4.7789	4.6646	0.1143	88	6.3614	6.2528	0.1085
58	4.7925	4.7052	0.0873	89	6.4075	6.3349	0.0727
59	4.7961	4.7460	0.0501	90	6.5000	6.4225	0.0775
60	4.8169	4.7870	0.0299	91	6.5748	6.5170	0.0578
61	4.9276	4.8283	0.0993	92	6.5754	6.6199	0.0445
62	4.9462	4.8699	0.0763	93	6.6528	6.7334	0.0807
63	4.9703	4.9119	0.0583	94	6.6652	6.8609	0.1957
64	4.9956	4.9543	0.0414	95	7.2770	7.0074	0.2696
65	5.0734	4.9970	0.0764	96	7.2817	7.1817	0.1000
66	5.1117	5.0402	0.0715	97	7.3439	7.4009	0.0569
67	5.1847	5.0839	0.1008	98	7.7365	7.7055	0.0310
68	5.2206	5.1282	0.0924	99	9.7339	8.2472	1.4867
69	5.2311	5.1730	0.0581	MSE= 0.0554 , CV= 0.1855			

4. Application

In our research endeavor, we embark on a novel approach by utilizing a distinct dataset to enrich the new methodological framework. Recognizing the significance of diverse data sources in refining analytical techniques, we deliberately select an alternative dataset to broaden the scope of our investigation. This strategic decision enables us to explore unique perspectives and uncover latent patterns that might have remained unnoticed with conventional datasets.

Furthermore, our study extends beyond mere methodological innovation; it encompasses a comprehensive comparative analysis. To elucidate the efficacy and robustness of the proposed method,

we juxtapose its performance against an established benchmark on the same dataset. This comparative examination serves as a litmus test, offering insights into the strengths and limitations of various methodologies under similar conditions.

In summary, our research endeavor not only introduces a novel methodological approach but also contributes to the broader discourse by juxtaposing it with existing methodologies on a shared dataset. Through this multifaceted exploration, we strive to advance the frontiers of knowledge in our field and pave the way for more informed decision-making and robust analytical practices.

4.1. Comparison with alternative prediction methods

In this subsection, extend our analysis to compare the proposed pivotal approach with the method of Valiollahi et al. [14] (e.g., maximum likelihood predictor (MLP), best unbiased predictor (BUP), conditional median predictor (CMP), and Bayesian predictor (BP)). Table 6 presents a side by side comparison of the exact observed values, the predictions obtained from the pivotal method, and those obtained from the alternative method under the Type-II HCS scheme with $r = 20$. The results show that the pivotal method produces prediction points and intervals that are consistent with the actual data and comparable to the alternative approach. Importantly, the width of the prediction intervals obtained by the suggested method remains competitive, reflecting efficiency and robustness. This comparison highlights that the improvement in pivotal method provides an accurate prediction of future order statistics while retaining computational simplicity. Therefore, by incorporating both methods, the reader can see the accuracy and reliability of the proposed pivotal technique. The real data set analyzed are available in Valiollahi et al. [14], and was originally reported by Lawless [15].

Table 6. Comparison of point predictions, biases, and 95% PIs under Type-II HCS $r = 20$.

	Exact	MLP		BUP		CMP		BP		Proposed Method (Pivotal results)	
r	value	Pred	Bias	Pred	Bias	Pred	Bias	Pred	Bias	$\hat{x}_{s:n}$	Bias
$r = 13$	6.479	6.369	0.110	6.464	0.015	6.436	0.043	6.521	0.042	6.4364	0.043
$r = 15$	6.515	6.557	0.042	6.653	0.138	6.626	0.111	6.852	0.337	6.5671	0.052
$r = 17$	6.538	6.745	0.207	6.843	0.305	6.819	0.281	7.162	0.624	6.6939	0.156
$r = 20$	6.725	7.030	0.305	7.156	0.431	7.112	0.387	7.802	1.077	6.8770	0.152
MSE			0.0374		0.0745		0.0607		0.4162		0.0130

4.2. Voltage stress data in a laboratory experiment

We use real data from Magdy et al. [8] to improve the proposed method the data were given by Lawless [15, p. 189]. Non-parametric plots for voltage stress data in a laboratory experiment are displayed in Figure 4. It consists of voltage levels at which failures occur in a certain type of electrical cable insulation (Type 1 insulation) when specimens are subjected to an increasing voltage stress in a laboratory experiment. The test involves 20 specimens, and the failure voltages in kilo-volts per millimeter are shown in Table 7.

Table 7. Voltage stress data in a laboratory experiment.

32.0	35.4	36.2	39.8	41.2	43.3	45.5	46.0	46.2	46.4
46.5	46.8	47.3	47.3	47.6	49.2	50.4	50.9	52.4	56.3

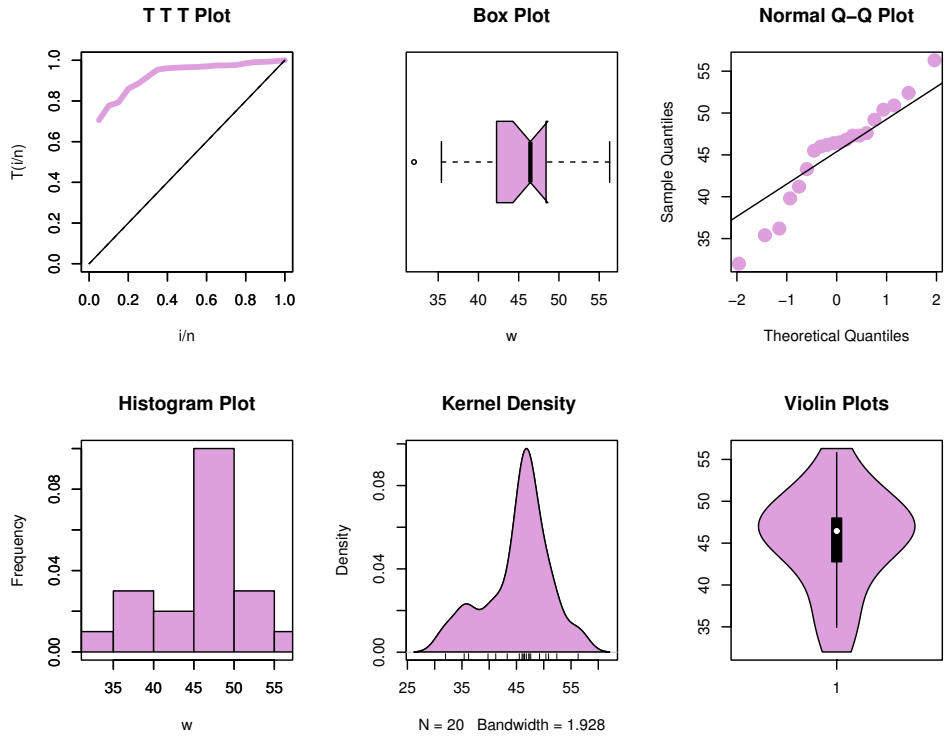


Figure 4. Non-parametric visualization plots for voltage stress data in a laboratory experiment.

In Table 8, we apply the suggested method with choosing $\hat{x}$ in pivotal quantity to be the first third interval, and the mean square error is lower than the mean square error if it is evaluated as in [8]. Further more, Table 7 shows that by applying the methods of pivotal quantity and modified least squares, the method of pivotal quantity is the best because it has the lowest MSE.

Table 8. Specimens and the failure voltages initial weibull:$a = 9.1973, \sigma = 47.7383$.

Pivotal quantity					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
9	46.40	46.82	0.42	(46.20, 48.15)	9	46.40	47.50	1.10
10	46.50	47.41	0.91	(46.82, 48.65)	10	46.50	48.71	2.21
11	46.80	47.97	1.17	(47.41, 49.16)	11	46.80	49.86	3.06
12	47.29	48.52	1.23	(47.97, 49.67)	12	47.30	51.00	3.70
13	47.31	49.06	1.75	(48.52, 50.21)	13	47.30	52.14	4.84
14	47.60	49.62	2.02	(49.06, 50.78)	14	47.60	53.32	5.72
15	49.20	50.20	1.00	(49.62, 51.41)	15	49.20	54.57	5.37
16	50.40	50.82	0.42	(50.20, 52.14)	16	50.40	55.94	5.54
17	50.90	51.53	0.63	(50.82, 53.03)	17	50.90	57.53	6.63
18	52.40	52.40	0.00	(51.53, 54.29)	18	52.40	52.53	0.13
19	56.30	53.69	2.61	(52.40, 56.58)	19	56.30	62.07	5.77
MSE=1.7672, CV=0.0411					MSE=20.2249, CV=0.075			

4.3. *Strength of glass fibers*

Smith et al. [16] discuss the application of Bayesian methods to the Weibull distribution and highlight that the challenges faced are related to the interpretation of results rather than computational issues. The authors introduce a paper structured as a case study, focusing on two sets of experimental data representing the strength of glass fibers with lengths of 1.5 cm (63 observation, see Table 9) and 15 cm (46 observation, see Table 10). Non-parametric plots for the strength of glass fibers with lengths of 1.5 cm and 15 cm are displayed in Figures 5 and 6. The data comes from the National Physical Laboratory in England. Wu et al. [17] use a 63 data set with length 1.5 cm. In this paper, we apply the proposed method to predict the future observation of real data with two lengths, 1.5 cm and 15 cm. The results are displayed in Tables 11 and 12.

Table 9. Strength of glass fibers with lengths of 1.5 cm.

0.55	0.93	1.25	1.36	1.49	1.52	1.58	1.61	1.64	1.68	1.73	1.81	2.00	0.74
1.04	1.27	1.39	1.49	1.53	1.59	1.61	1.66	1.68	1.76	1.82	2.01	0.77	1.11
1.28	1.42	1.50	1.54	1.60	1.62	1.66	1.69	1.76	1.84	2.24	0.81	1.13	1.29
1.48	1.50	1.55	1.61	1.62	1.66	1.70	1.77	1.84	0.84	1.24	1.30	1.48	1.51
1.55	1.61	1.63	1.67	1.70	1.78	1.89							

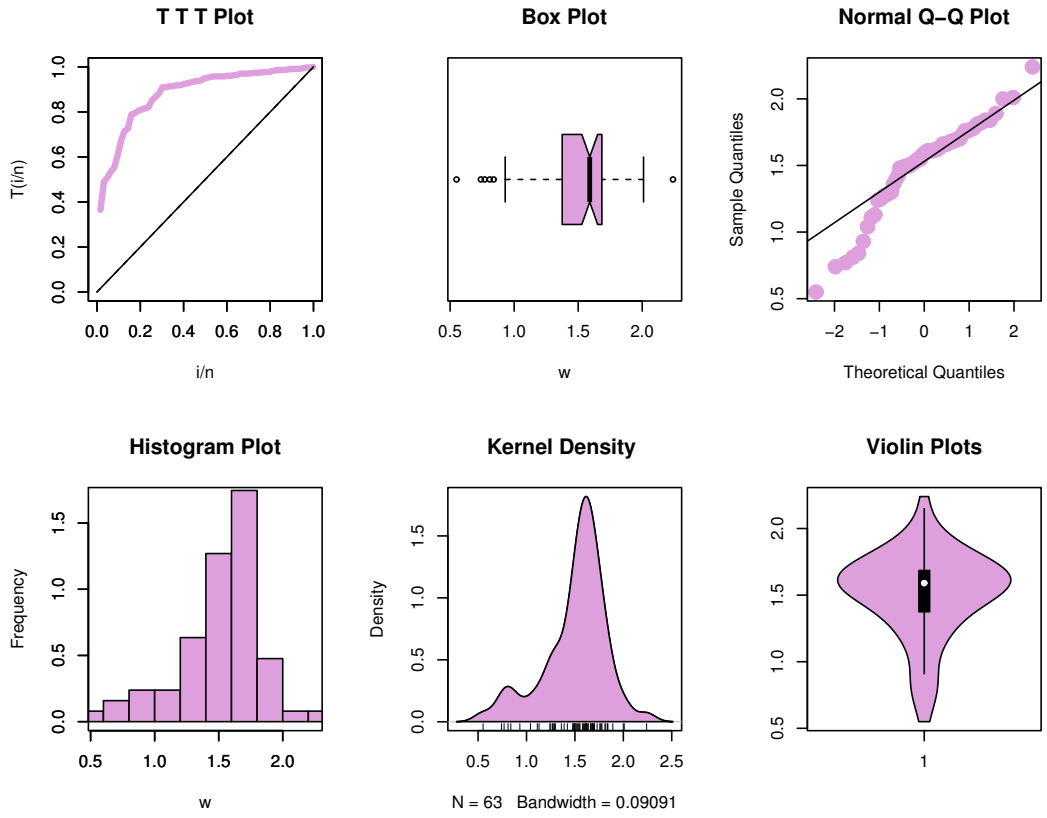


Figure 5. Non-parametric visualization plots for the strength of glass fibers with a lengths of 1.5 cm.

Table 10. Strength of glass fibers with lengths of 15 cm.

0.37	0.40	0.70	0.75	0.80	0.81	0.83	0.86	0.92	0.92
0.94	0.95	0.98	1.03	1.06	1.06	1.08	1.09	1.10	1.10
1.13	1.14	1.15	1.17	1.20	1.20	1.21	1.22	1.25	1.28
1.28	1.29	1.29	1.30	1.35	1.35	1.37	1.37	1.38	1.40
1.40	1.42	1.43	1.51	1.53	1.61				

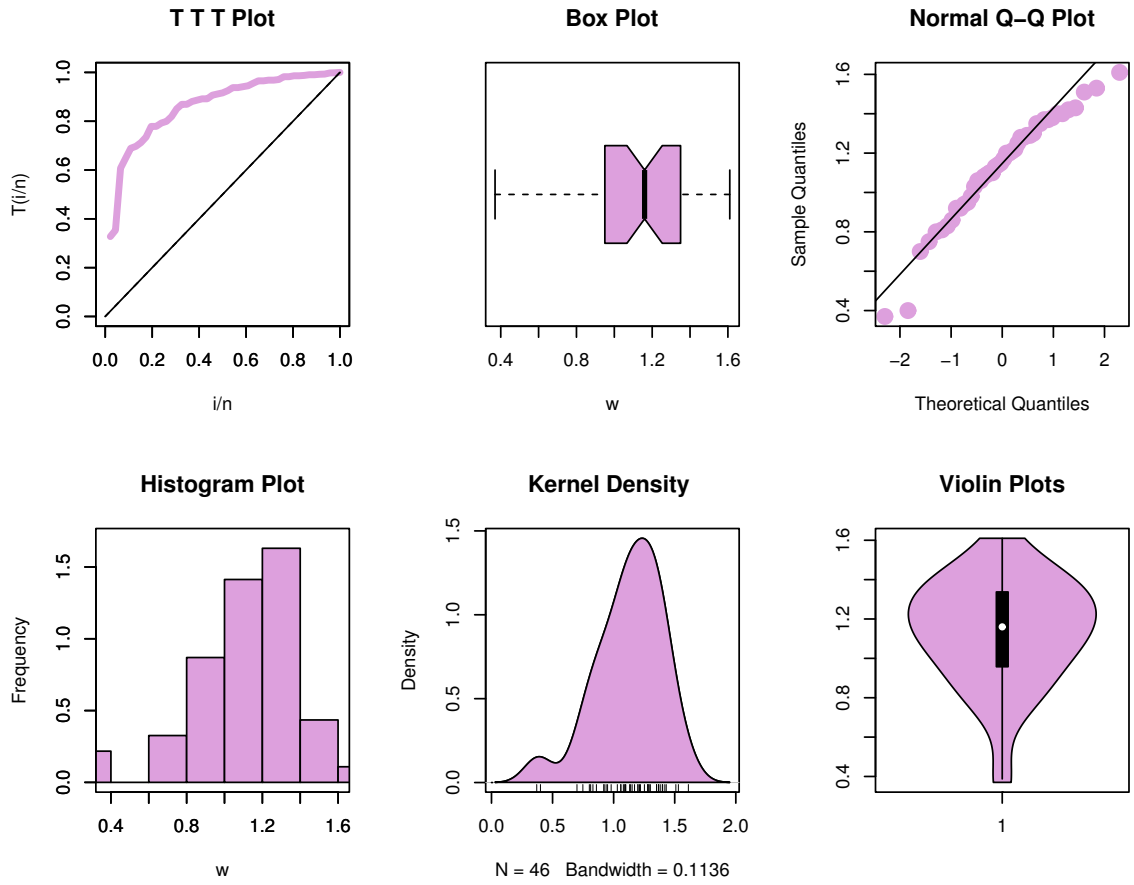


Figure 6. Non-parametric visualization plots for Strength of glass fibers with lengths of 15 cm data.

Table 11. Strength of glass fibers with lengths of 15 cm.

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
23	1.17	1.16	0.01	(1.15, 1.19)	23	1.17	1.13	0.04
24	1.20	1.17	0.03	(1.16, 1.20)	24	1.20	1.14	0.06
25	1.20	1.18	0.02	(1.17, 1.21)	25	1.20	1.15	0.05
26	1.21	1.20	0.01	(1.18, 1.22)	26	1.21	1.13	0.08
27	1.22	1.21	0.01	(1.20, 1.23)	27	1.22	1.14	0.08
28	1.25	1.22	0.03	(1.21, 1.24)	28	1.25	1.15	0.10
29	1.28	1.23	0.05	(1.22, 1.25)	29	1.28	1.13	0.15
30	1.28	1.24	0.04	(1.23, 1.26)	30	1.28	1.14	0.14
31	1.29	1.25	0.04	(1.24, 1.27)	31	1.29	1.15	0.14
32	1.29	1.26	0.03	(1.25, 1.29)	32	1.29	1.13	0.16
33	1.30	1.27	0.03	(1.26, 1.30)	33	1.30	1.14	0.16
34	1.35	1.29	0.06	(1.27, 1.31)	34	1.35	1.15	0.20
35	1.35	1.30	0.05	(1.29, 1.33)	35	1.35	1.13	0.22
36	1.37	1.31	0.06	(1.30, 1.34)	36	1.37	1.14	0.23
37	1.37	1.33	0.04	(1.31, 1.35)	37	1.37	1.15	0.22
38	1.38	1.34	0.04	(1.33, 1.37)	38	1.38	1.13	0.25
39	1.40	1.36	0.04	(1.34, 1.39)	39	1.40	1.14	0.26
40	1.40	1.37	0.03	(1.36, 1.41)	40	1.40	1.15	0.25
41	1.42	1.39	0.03	(1.37, 1.43)	41	1.42	1.13	0.29
42	1.43	1.41	0.02	(1.39, 1.46)	42	1.43	1.14	0.29
43	1.51	1.44	0.07	(1.41, 1.50)	43	1.51	1.15	0.36
44	1.53	1.48	0.05	(1.44, 1.56)	44	1.53	1.13	0.40
45	1.61	1.53	0.08	(1.48, 1.67)	45	1.61	1.14	0.47
MSE=0.0017, CV=0.076					MSE=0.0523, CV=0.0070			

Table 12. Strength of glass fibers with lengths of 1.5 cm.

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
30	1.58	1.56	0.02	(1.55, 1.59)	30	1.58	1.54	0.04
31	1.59	1.57	0.02	(1.56, 1.60)	31	1.59	1.55	0.04
32	1.60	1.59	0.01	(1.57, 1.61)	32	1.60	1.55	0.05
33	1.61	1.60	0.01	(1.59, 1.62)	33	1.61	1.54	0.07
34	1.61	1.61	0.00	(1.60, 1.63)	34	1.61	1.55	0.06
35	1.61	1.62	0.01	(1.61, 1.64)	35	1.61	1.55	0.06
36	1.61	1.63	0.02	(1.62, 1.65)	36	1.61	1.54	0.07
37	1.62	1.64	0.02	(1.63, 1.66)	37	1.62	1.55	0.07
38	1.62	1.65	0.03	(1.64, 1.67)	38	1.62	1.55	0.07
39	1.63	1.66	0.03	(1.65, 1.68)	39	1.63	1.54	0.09
40	1.64	1.67	0.03	(1.66, 1.70)	40	1.64	1.55	0.09
41	1.66	1.68	0.02	(1.67, 1.71)	41	1.66	1.55	0.11
42	1.66	1.70	0.04	(1.68, 1.72)	42	1.66	1.54	0.12
43	1.66	1.71	0.05	(1.70, 1.73)	43	1.66	1.55	0.11
44	1.67	1.72	0.05	(1.71, 1.74)	44	1.67	1.55	0.12
45	1.68	1.73	0.05	(1.72, 1.75)	45	1.68	1.54	0.14
46	1.68	1.74	0.06	(1.73, 1.76)	46	1.68	1.55	0.13
47	1.69	1.75	0.06	(1.74, 1.78)	47	1.69	1.55	0.14
48	1.70	1.76	0.06	(1.75, 1.79)	48	1.70	1.54	0.16
49	1.70	1.78	0.08	(1.76, 1.80)	49	1.70	1.55	0.15
50	1.73	1.79	0.06	(1.78, 1.81)	50	1.73	1.55	0.18
51	1.76	1.80	0.04	(1.79, 1.83)	51	1.76	1.54	0.22
52	1.76	1.82	0.06	(1.80, 1.84)	52	1.76	1.55	0.21
53	1.77	1.83	0.06	(1.82, 1.86)	53	1.77	1.55	0.22
54	1.78	1.85	0.07	(1.83, 1.88)	54	1.78	1.54	0.24
55	1.81	1.86	0.05	(1.85, 1.90)	55	1.81	1.55	0.26
56	1.82	1.88	0.06	(1.86, 1.92)	56	1.82	1.55	0.27
57	1.84	1.90	0.06	(1.88, 1.94)	57	1.84	1.54	0.30
58	1.84	1.92	0.08	(1.90, 1.97)	58	1.84	1.55	0.29
59	1.89	1.95	0.06	(1.92, 2.01)	59	1.89	1.55	0.34
60	2.00	1.98	0.02	(1.95, 2.05)	60	2.00	1.54	0.46
61	2.01	2.02	0.01	(1.98, 2.12)	61	2.01	1.55	0.46
62	2.24	2.09	0.15	(2.02, 2.25)	62	2.24	1.55	0.69
MSE=0.0027, CV=0.0769					MSE=0.053, CV=0.003			

Tables 11 and 12 show that by applying the methods of pivotal quantity and modified least squares, the method of pivotal quantity is the best because it has the lowest MSE.

4.4. Recovery rate of Covid-19

The study uses two data sets: 38 observations from France (Table 13) and 25 from Turkey (Table 14). These data consist of the daily ratio of total recoveries to cumulative confirmed cases, as well as cumulative death counts. Ahmad et al. [18] and Khaled et al. [5] used these data to predict a future observation; we will use the same data to apply our new method and try to predict 100% from existing data. Non-parametric plots for the recovery rate of Covid-19 in France and Turkey are shown in Figures 7 and 8.

Table 13. Recovery rate of Covid-19 in France.

0.195	0.2338	0.2368	0.1073	0.1592	0.2784	0.0689	0.1791
0.1121	0.1865	0.2631	0.0716	0.1411	0.1477	0.1874	0.0853
0.0922	0.1711	0.1962	0.2146	0.1041	0.1524	0.1811	0.0643
0.2698	0.1245	0.176	0.2363	0.0712	0.1361	0.1386	0.3316
0.077	0.1367	0.1549	0.2178	0.0951	0.1346		

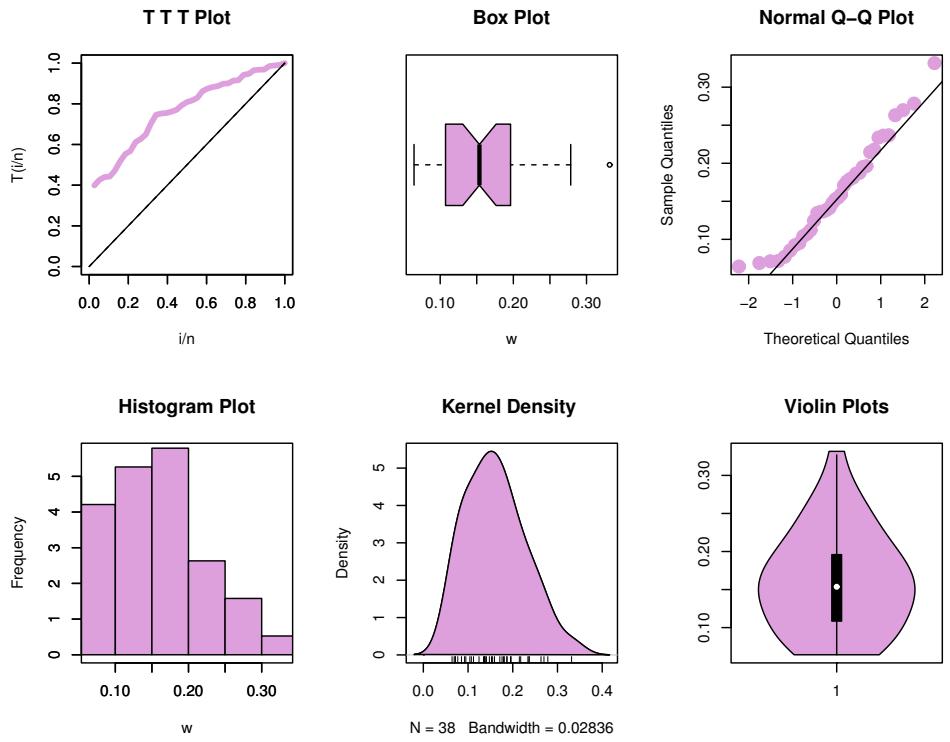


Figure 7. Non-parametric visualization plots for the recovery rate of Covid-19 in France.

Table 14. Recovery rate of Covid-19 in Turkey.

0.0074	0.0095	0.0113	0.015	0.018	0.0212	0.0229	0.0231	0.0328
0.0385	0.0439	0.0464	0.0483	0.0507	0.0515	0.0568	0.0605	0.0648
0.0737	0.0818	0.0955	0.1099	0.127	0.1388	0.1476		

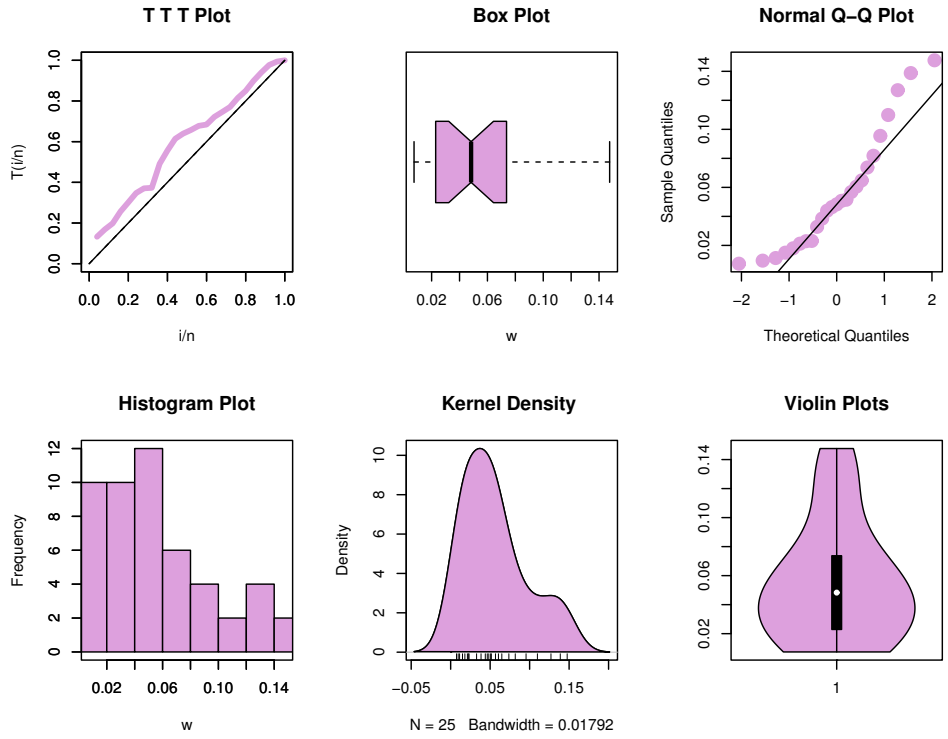


Figure 8. Non-parametric visualization plots for the recovery rate of Covid-19 in Turkey.

Table 15 shows that the pivotal quantity is better than the modified least square method. In Table 16, $\hat{x}_s$ is obtained by taking the average of the four ways mentioned above. From MSE, we see that the modified least square method is better than pivotal quantity, as it has a minimum value: MSE=0.00020.

Table 15. Predcting 100% of the recovery rate of Covid-19 in Turkey.

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
12	0.0483	0.0518	0.0035	(0.0464, 0.0585)	12	0.0483	0.0507	0.0024
13	0.0507	0.0576	0.0069	(0.0518, 0.0648)	13	0.0507	0.0553	0.0046
14	0.0515	0.0620	0.0105	(0.0576, 0.0672)	14	0.0515	0.0602	0.0087
15	0.0568	0.0668	0.0100	(0.0620, 0.0725)	15	0.0568	0.0656	0.0088
16	0.0605	0.0722	0.0117	(0.0668, 0.0784)	16	0.0605	0.0715	0.0110
17	0.0648	0.0780	0.0132	(0.0722, 0.0849)	17	0.0648	0.0780	0.0132
18	0.0737	0.0846	0.0109	(0.0780, 0.0924)	18	0.0737	0.0853	0.0116
19	0.0818	0.0921	0.0103	(0.0846, 0.1011)	19	0.0818	0.0937	0.0119
20	0.0955	0.1010	0.0055	(0.0921, 0.1116)	20	0.0955	0.1286	0.0331
21	0.1099	0.1117	0.0018	(0.1010, 0.1250)	21	0.1099	0.1325	0.0226
22	0.1270	0.1255	0.0015	(0.1117, 0.1433)	22	0.1270	0.1450	0.0180
23	0.1388	0.1452	0.0064	(0.1255, 0.1722)	23	0.1388	0.1687	0.0299
24	0.1476	0.1797	0.0321	(0.1452, 0.2358)	24	0.1476	0.2168	0.0692
MSE= 0.00014 , CV=0.382					MSE= 0.00064, CV=0.4681			

Table 16. Predcting 100% of the recovery rate of Covid-19 in France.

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
19	0.1549	0.1561	0.0012	(0.1524, 0.1600)	19	0.1549	0.1564	0.0015
20	0.1592	0.1598	0.0006	(0.1561, 0.1637)	20	0.1592	0.1605	0.0013
21	0.1711	0.1636	0.0075	(0.1598, 0.1676)	21	0.1711	0.1646	0.0065
22	0.1760	0.1675	0.0085	(0.1636, 0.1715)	22	0.1760	0.1688	0.0072
23	0.1791	0.1714	0.0077	(0.1675, 0.1756)	23	0.1791	0.1731	0.0060
24	0.1811	0.1755	0.0056	(0.1714, 0.1797)	24	0.1811	0.1775	0.0036
25	0.1865	0.1797	0.0068	(0.1755, 0.1841)	25	0.1865	0.1820	0.0045
26	0.1874	0.1840	0.0034	(0.1797, 0.1886)	26	0.1874	0.1867	0.0007
27	0.1950	0.1886	0.0064	(0.1840, 0.1934)	27	0.1950	0.1916	0.0034
28	0.1962	0.1934	0.0028	(0.1886, 0.1985)	28	0.1962	0.1968	0.0006
29	0.2146	0.1985	0.0161	(0.1934, 0.2039)	29	0.2146	0.2022	0.0124
30	0.2178	0.2040	0.0138	(0.1985, 0.2099)	30	0.2178	0.2081	0.0097
31	0.2338	0.2101	0.0237	(0.2040, 0.2165)	31	0.2338	0.2144	0.0194
32	0.2363	0.2168	0.0195	(0.2101, 0.2239)	32	0.2363	0.2214	0.0149
33	0.2368	0.2244	0.0124	(0.2168, 0.2326)	33	0.2368	0.2293	0.0075
34	0.2631	0.2334	0.0297	(0.2244, 0.2431)	34	0.2631	0.2385	0.0246
35	0.2698	0.2446	0.0252	(0.2334, 0.2569)	35	0.2698	0.2498	0.0200
36	0.2784	0.2598	0.0186	(0.2446, 0.2771)	36	0.2784	0.2653	0.0131
37	0.3316	0.2853	0.0463	(0.2598, 0.3164)	37	0.3316	0.2922	0.0394
MSE=0.0003. CV=0.173					MSE= 0.00020 , CV=0.1794			

4.5. Employer-sponsored health insurance (ESI)

This dataset contains data on employer-sponsored health insurance (ESI) coverage among private-sector workers in the USA from 1979 to 2019. It includes demographic breakdowns such as race, gender, education level, and recent graduation status. The data is available at: https://www.kaggle.com/datasets/asaniczka/health-insurance-coverage-in-the-usa-1979-2019/data.

We predict 21 future observations (percentage of female workers with ESI coverage).

Table 17 shows that by applying the methods of pivotal quantity and modified least squares, the method of pivotal quantity is the best because it has the lowest MSE.

Table 17. Percentage of female workers with ESI coverage: 41 observations used to predict 21 future observations.

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
20	53.00	53.23	0.23	(53.00, 53.46)	20	53.00	53.15	0.15
21	53.20	53.46	0.26	(53.23, 53.70)	21	53.20	53.32	0.12
22	53.30	53.70	0.40	(53.46, 53.94)	22	53.30	53.48	0.18
23	53.40	53.94	0.54	(53.70, 54.19)	23	53.40	53.66	0.26
24	53.40	54.19	0.79	(53.94, 54.44)	24	53.40	53.85	0.45
25	53.50	54.45	0.95	(54.19, 54.71)	25	53.50	54.04	0.54
26	53.60	54.71	1.11	(54.45, 54.98)	26	53.60	54.04	0.44
27	53.70	54.99	1.29	(54.71, 55.27)	27	53.70	54.04	0.34
28	54.00	55.28	1.28	(54.99, 55.57)	28	54.00	54.04	0.04
29	54.30	55.58	1.28	(55.28, 55.89)	29	54.30	54.04	0.26
30	54.80	55.90	1.10	(55.58, 56.23)	30	54.80	54.04	0.76
31	54.90	56.24	1.34	(55.90, 56.59)	31	54.90	54.04	0.86
32	55.30	56.61	1.31	(56.24, 56.98)	32	55.30	54.04	1.26
33	59.00	57.01	1.99	(56.61, 57.41)	33	59.00	54.04	4.96
34	59.40	57.45	1.95	(57.01, 57.89)	34	59.40	54.04	5.36
35	59.80	57.94	1.86	(57.45, 58.44)	35	59.80	54.94	4.86
36	60.10	58.51	1.59	(57.94, 59.09)	36	60.10	55.26	4.84
37	60.80	59.18	1.62	(58.51, 59.87)	37	60.80	55.26	5.54
38	60.90	60.03	0.87	(59.18, 60.91)	38	60.90	55.26	5.64
39	61.10	61.21	0.11	(60.03, 62.45)	39	61.10	56.04	5.06
40	61.30	63.29	1.99	(61.21, 65.51)	40	61.30	56.04	5.26
MSE=1.62561. CV=0.0468					MSE=10.48 , CV=0.0147			

4.6. Reticulum cell sarcoma and windscreen failures

We utilize the data on windscreen failures, which include 84 observed failure times for a specific model windscreen. Following Aly et al. [4], we use Eq (2.4) to estimate the parameter and Eq (2.5) to predict the point for every next predictive point. As such, we consider the last predictive point and re-estimate parameters. To do this, we apply the proposed method to predict 150% of the data and compare it with the work of Aly et al. [4]. The results are shown in Table 19. The second dataset reported and analyzed by Hoel. [12] and Azm et al. [19], involves male mice receiving a radiation dose of 300 roentgen at an age of 5–6 weeks. Each mouse's cause of death was identified through autopsy, classified as thymic lymphoma, reticulum cell sarcoma, or other causes. For the purpose of our analysis, reticulum cell sarcoma is designated as cause 1, while the other two causes are merged to form cause 2. We focus on the observations attributed to cause 1 (reticulum cell sarcoma), which includes 38 data points (Table 18).

Table 18. Reticulum cell sarcoma.

317	318	399	495	525	536	549	552	554	557	558	571	586
594	596	605	612	621	628	631	636	643	647	648	649	661
663	666	670	695	697	700	705	712	713	738	748	753	

Non-parametric plots for the reticulum cell sarcoma are shown in Figure 9.

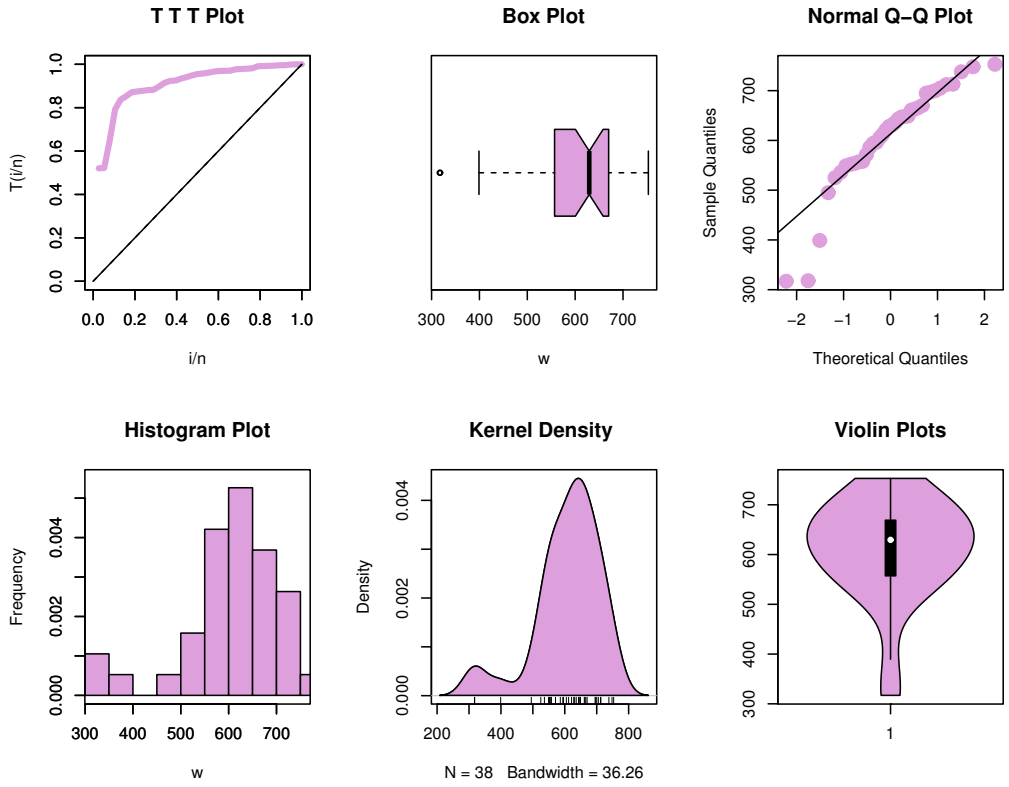


Figure 9. Non-parametric visualization plots for reticulum cell sarcoma data.

Aly et al. [4] studied this data, and we applay the methodology for the same data results shown in Table 20 demonstrates that the Makeham-Gompertz distribution is adequate for fitting reticulum cell sarcomas.

Table 19 shows that pivotal quantity is the best because it has the lowest MSE, and Table 20 shows that modified least squares is the best.

Table 19. At r = 34, there are 150% prediction windscreen failures

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
34	2.154	2.169	0.015	(2.135, 2.204)	34	2.154	2.167	0.013
35	2.190	2.203	0.013	(2.169, 2.238)	35	2.190	2.199	0.009
36	2.194	2.237	0.043	(2.203, 2.272)	36	2.194	2.230	0.036
37	2.223	2.270	0.047	(2.237, 2.303)	37	2.223	2.260	0.037
38	2.224	2.302	0.078	(2.270, 2.336)	38	2.224	2.289	0.065
39	2.229	2.335	0.106	(2.302, 2.368)	39	2.229	2.318	0.089
40	2.300	2.367	0.067	(2.335, 2.400)	40	2.300	2.347	0.047
41	2.324	2.399	0.075	(2.367, 2.433)	41	2.324	2.375	0.051
42	2.385	2.432	0.047	(2.399, 2.465)	42	2.385	2.403	0.018
43	2.481	2.464	0.017	(2.432, 2.497)	43	2.481	2.431	0.050
44	2.610	2.496	0.114	(2.464, 2.529)	44	2.610	2.459	0.151
45	2.625	2.528	0.097	(2.496, 2.561)	45	2.625	2.487	0.138
46	2.632	2.560	0.072	(2.528, 2.593)	46	2.632	2.514	0.118
47	2.646	2.592	0.054	(2.560, 2.625)	47	2.646	2.542	0.104
48	2.661	2.624	0.037	(2.592, 2.657)	48	2.661	2.569	0.092
49	2.688	2.657	0.031	(2.624, 2.691)	49	2.688	2.597	0.091
50	2.823	2.690	0.133	(2.657, 2.724)	50	2.823	2.624	0.199
51	2.890	2.723	0.167	(2.690, 2.757)	51	2.890	2.652	0.238
52	2.902	2.756	0.146	(2.723, 2.790)	52	2.902	2.680	0.222
53	2.934	2.789	0.145	(2.756, 2.823)	53	2.934	2.708	0.226
54	2.962	2.823	0.139	(2.789, 2.858)	54	2.962	2.737	0.225
55	2.964	2.857	0.107	(2.823, 2.893)	55	2.964	2.765	0.199
56	3.000	2.893	0.107	(2.857, 2.929)	56	3.000	2.794	0.206
57	3.103	2.929	0.174	(2.893, 2.965)	57	3.103	2.823	0.280
58	3.114	2.964	0.150	(2.929, 3.001)	58	3.114	2.852	0.262
59	3.117	3.001	0.116	(2.964, 3.038)	59	3.117	2.882	0.235
60	3.166	3.038	0.128	(3.001, 3.076)	60	3.166	2.913	0.253
61	3.344	3.076	0.268	(3.038, 3.115)	61	3.344	2.944	0.400
62	3.376	3.115	0.261	(3.076, 3.155)	62	3.376	2.975	0.401
63	3.443	3.155	0.288	(3.115, 3.195)	63	3.443	3.007	0.436
64	3.467	3.196	0.271	(3.155, 3.239)	64	3.467	3.040	0.427
65	3.478	3.239	0.239	(3.196, 3.283)	65	3.478	3.073	0.405
66	3.578	3.282	0.296	(3.239, 3.327)	66	3.578	3.107	0.471
67	3.595	3.328	0.267	(3.282, 3.374)	67	3.595	3.143	0.452
68	3.699	3.375	0.324	(3.328, 3.423)	68	3.699	3.179	0.520
69	3.779	3.424	0.355	(3.375, 3.474)	69	3.779	4.179	0.400
70	3.924	3.475	0.449	(3.424, 3.528)	70	3.924	3.984	0.060
71	4.035	3.529	0.506	(3.475, 3.585)	71	4.035	3.851	0.184
72	4.121	3.586	0.535	(3.529, 3.645)	72	4.121	3.787	0.334
73	4.167	3.647	0.520	(3.586, 3.710)	73	4.167	3.782	0.385
74	4.240	3.712	0.528	(3.647, 3.779)	74	4.240	4.863	0.623
75	4.255	3.781	0.474	(3.712, 3.853)	75	4.255	4.527	0.272
76	4.278	3.857	0.421	(3.781, 3.937)	76	4.278	4.350	0.072
77	4.305	3.941	0.364	(3.857, 4.029)	77	4.305	4.325	0.020
78	4.376	4.035	0.341	(3.941, 4.134)	78	4.376	4.390	0.014
79	4.449	4.144	0.305	(4.035, 4.258)	79	4.449	4.497	0.048
80	4.485	4.272	0.213	(4.144, 4.408)	80	4.485	4.631	0.146
81	4.570	4.431	0.139	(4.272, 4.603)	81	4.570	4.799	0.229
82	4.602	4.648	0.046	(4.431, 4.886)	82	4.602	5.029	0.427
83	4.663	5.006	0.343	(4.648, 5.424)	83	4.663	5.425	0.762
MSE=0.064 ,CV=0.218					MSE=0.0803,CV= 0.276			

Table 20. At $r = 15$, there are 150% prediction of reticulum cell sarcoma.

Pivotal quantity method					Modified least square method			
r	$x_{s:n}$	$\hat{x}_{s:n}$	biase	PI	r	$x_{s:n}$	$\hat{x}_{s:n}$	biase
15	605	600	5	(596, 610)	15	605	603	2
16	612	605	7	(600, 613)	16	612	611	1
17	621	609	12	(605, 617)	17	621	617	4
18	628	613	15	(609, 621)	18	628	624	4
19	631	616	15	(613, 624)	19	631	630	1
20	636	620	16	(616, 628)	20	636	636	0
21	643	624	19	(620, 631)	21	643	642	1
22	647	627	20	(624, 635)	22	647	647	0
23	648	631	17	(627, 638)	23	648	653	5
24	649	634	15	(631, 642)	24	649	658	9
25	661	638	23	(634, 645)	25	661	664	3
26	663	642	21	(638, 649)	26	663	669	6
27	666	645	21	(642, 653)	27	666	674	8
28	670	649	21	(645, 657)	28	670	680	10
29	695	653	42	(649, 661)	29	695	686	9
30	697	657	40	(653, 665)	30	697	692	5
31	700	661	39	(657, 670)	31	700	698	2
32	705	666	39	(661, 675)	32	705	705	0
33	712	671	41	(666, 681)	33	712	712	0
34	713	676	37	(671, 688)	34	713	721	8
35	738	683	55	(676, 697)	35	738	731	7
36	748	692	56	(683, 709)	36	748	744	4
37	753	704	49	(692, 732)	37	753	764	11
MSE=960.7, CV=0.043					MSE=31.198 , CV=0.063			

5. Conclusions

We have developed a prediction of future observations for the dataset using new techniques. This requires proper data testing to determine the optimal distribution based on statistical tests. Considering historical data predictions alongside predictive future observations, we emphasize the importance of not relying solely on a specific limit from distributions. Instead, we incorporate statistical parameters to determine the best-fitting distribution for the expected dataset. This comprehensive method, implemented using the R language, entails developing code that executes these procedures. The code is readily available through the updated Prediction R package. These statistical packages help non-specialists predict future data in scientific fields. The method is effective under fixed and random sample sizes (Algorithms I & II). Simulation studies (Tables 1–4) confirmed that the technique remains accurate even when predicting far beyond the original dataset, with error (bias) increasing gradually and predictably. In Table 6, when compared to other methods (MLP, BUP, CMP, and BP), we found that the proposed method provides the lowest bias, particularly beyond 30% of the original data. Indeed,

it yielded the lowest MSE. Finally, as future work, we recommend applying the proposed method to bivariate data, as well as developing new programming in R to facilitate the implementation of this method by any user.

Author contributions

M. H. Harpy: Formal analysis, investigation, methodology, validation, writing, review and editing; O. M. Khaled: Conceptualization, data curation, methodology, software; Mahmoud El-Morshedy: Formal analysis, resources, validation; K. S. Khalil: Data curation, investigation, resources, software. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/33256).

Conflict of interest

The authors declare no conflict of interests.

References

1. H. M. Barakat, O. M. Khaled, H. A. Ghonem, Predicting future order statistics with random sample size, *AIMS Math.*, **6** (2021), 5133–5147. http://dx.doi.org/10.3934/MATH.2021304
2. H. M. Barakat, M. E. El-Adll, A. E. Aly, Prediction intervals of future observations for a sample of random size from any continuous distribution, *Math. Comput. Simulat.*, **97** (2014), 1–13. http://dx.doi.org/10.1016/j.matcom.2013.06.007
3. J. K. Patel, Prediction intervals–A review, *Commun. Stat.-Theor. M.*, **18** (1989), 2393–2465. http://dx.doi.org/10.1080/03610928908830043
4. A. E. Aly, M. E. El-Adll, H. M. Barakat, R. A. Aldallal, A new least squares method for estimation and prediction based on the cumulative hazard function, *AIMS Math.*, **8** (2023), 21968–21992. http://dx.doi.org/10.3934/math.20231120
5. O. M. Khaled, K. S. Khalil, M. H. Harpy, Predicting future data from gamma-mixture and beta-mixture distributions and application to the recovery rate of covid-19, *Adv. Appl. Stat.*, **90** (2023), 1–34. https://doi.org/10.17654/0972361723061
6. A. Chakrabarti, J. K. Ghosh, AIC, BIC and recent advances in model selection, *Philosophy Stat.*, **7** (2011), 583–605. https://doi.org/10.1016/B978-0-444-51862-0.50018-6

7. M. Rahmouni, D. Ziedan, The weibull-generalized shifted geometric distribution: Properties, estimation, and applications, *AIMS Math.*, **10** (2025), 9773–9804. http://dx.doi.org/10.3934/math.2025448
8. M. E. El-Adll, S. F. Ateya, M. M. Rizk, Prediction intervals for future lifetime of three parameters Weibull observations based on generalized order statistics, *Arab. J. Math.*, **1** (2012), 295–304. http://dx.doi.org/10.1007/s40065-012-0004-7
9. H. Rinne, *The Weibull distribution: A handbook*, 1 Eds., New Yourk: Chapman and Hall/CRC, 2008. http://dx.doi.org/10.1201/9781420087444
10. A. Karolczuk, T. P. Luc, Modelling of stress gradient effect on fatigue life using weibull based distribution function, *J. Theor. Appl. Mech.*, **51** (2013), 297–311.
11. A. A. Al-Babtain, M. K. Shakhatreh, M. Nassar, A. Z. Afify, A new modified kies family: Properties, estimation under complete and type-ii censored samples, and engineering applications, *Mathematics*, **8** (2020), 1345. http://dx.doi.org/10.3390/math8081345
12. D. G. Hoel, A representation of mortality data by competing risks, *Biometrics*, **28** (1972), 475–488. https://doi.org/10.2307/2556161
13. A. Elshahhat, H. Rezk, R. Alotaibi, The discrete gompertz-makeham distribution for multidisciplinary data analysis, *AIMS Math.*, **10** (2025), 17117–17178. http://dx.doi.org/10.3934/math.2025768
14. R. Valiollahi, A. Asgharzadeh, D. Kundu, Prediction of future failures for generalized exponential distribution under type-i or type-ii hybrid censoring, *Braz. J. Probab. Stat.*, **31** (2017), 41–61. http://dx.doi.org/10.1214/15-BJPS302
15. J. F. Lawless, *Statistical models and methods for lifetime data*, 2 Eds., United State, John Wiley & Sons, 2011.
16. R. L. Smith, J. C. Naylor, A comparison of maximum likelihood and bayesian estimators for the three-parameter weibull distribution, *J. Roy. Stat. Soc. C-Appl.*, **36** (1987), 358–369. https://doi.org/10.2307/2347795
17. Y. Wu, H. Xie, J. Y. Chiang, G. Peng, Y. Qin, Parameter estimation and applications of the Weibull distribution for strength data of glass fiber, *Math. Probl. Eng.*, **2021** (2021), 9175170. http://dx.doi.org/10.1155/2021/9175170
18. A. H. Haj, E. M. Almetwally, M. Elgarhy, D. A. Ramadan, On unit exponential Pareto distribution for modeling the recovery rate of COVID-19, *Processes*, **11** (2023), 232. http://dx.doi.org/10.3390/PR11010232
19. W. A. El Azm, R. Aldallal, H. M. Aljohani, S. G. Nassr, Estimations of competing lifetime data from inverse weibull distribution under adaptive progressively hybrid censored, *Math. Biosci. Eng.*, **19** (2022), 6252–6275. http://dx.doi.org/10.3934/mbe.2022292