Research Article GU J Sci 38(1): x-x (2025) DOI: 10.35378/gujs.1306771

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

A New Rank Estimator for Least Squares Estimation of Weibull Modulus

Burak BIRGOREN *

TOBB University of Economics and Technology, Department of Industrial Engineering, Sogutozu Cad.No:43, Sogutozu, Ankara, 06560 Türkiye

Highlights

- This paper focuses on ordinary least squares estimation of Weibull modulus by a new rank estimator.
- The new estimator is a quadratic function of ranks of order statistics, its parameters are optimized by simulations.
- The new estimator performs much better than well-known rank estimators for a large range of sample sizes.
- The new estimator performs better than the Maximum Likelihood Method for sample sizes less than 37.

Article Info

Received: 29 May 2023
Accepted: 22 July 2024

Keywords

Weibull distribution
Weibull modulus
Ordinary least squares
Rank estimator

Abstract

The Weibull distribution is widely used in reliability analysis to evaluate the failure behavior and lifetime characteristics of various systems and components. One of the most commonly used methods for estimating the parameters of the Weibull distribution is the ordinary least squares (OLS) technique, which is based on fitting a linear regression model to the transformed data. This paper proposes a new rank estimator for ordinary least squares estimation of Weibull modulus, a key parameter used as a measure of variability in the data. The new rank estimator is a quadratic function of the ranks of order statistics, with three parameters that are optimized by Monte Carlo simulations. Using relative efficiency as a criterion, the performance of the new rank estimator is compared with three commonly used rank estimators, mean, median and Hazen rank estimators, which are linear functions of the ranks of order statistics. The results show that the new rank estimator has a significant advantage over the other rank estimators for any sample size between 3 and 150. The findings also imply that other nonlinear functions, such as cubic polynomials, could be applied to further improve the efficiency of the parameter estimators of the ordinary least squares method.

1. INTRODUCTION

The Weibull distribution is one of the most widely used probability distributions in life testing and reliability studies. The distribution was developed by Weibull [1] and has found use in many different areas of application such as wind-speed analysis, unemployment durations analysis, material strength analysis as well as reliability analysis [2–4].

The two-parameter Weibull distribution function is given by

$$F(x) = 1 - exp\left[-\left(x/\alpha\right)^{\beta}\right] \tag{1}$$

where β is the shape parameter or Weibull modulus, and α is the scale parameter.

There are several methods for estimating the parameters of the Weibull distribution such as the Maximum Likelihood Estimation (MLE) method, the Ordinary Least Squares (OLS) method, the Weighted Least Squares method, the Moments methods, the Generalized Means Squares method and the Bayesian Estimation method

* e-mail: b.birgoren@etu.edu.tr

[3–7]. Among them, the OLS method and the MLE method have been the most commonly used methods for estimating the parameters in general reliability applications: The ML method is popular among statisticians because of its distributional optimality properties in large samples; many engineers, on the other hand, have used ordinary least squares (OLS) because of its simplicity and familiar probability plots [8,9]. These two methods also allow constructing confidence intervals for the parameters.

The OLS method requires using rank estimators (also called probability indices or plotting positions) [3,8]. They are usually in the general form of $\hat{F}(i)=(i-a)/(n+b)$ for complete data, where *n* is the sample size and $\hat{F}(i)$ is the rank estimator for the *i*th order statistic [10,11]. Several formulae with fixed values of *a* and *b* have been proposed in previous studies, such as the median rank with *a* = 0.3 and *b* = 0.4, as discussed in the following section.

Further, some formulae have been proposed in the materials science literature where *a* and *b* vary with the sample size; *a* and *b* ranging from 0 to 1 are considered in these studies. Most of them have been developed for the unbiased estimation of the Weibull modulus [12–17]; because it is used as a measure of variability of strength measurements. Only one of these studies has been on unbiased estimation of the scale parameter [16]. In general, these formulae are shown to perform better than the ones with the fixed *a* and *b*, in terms of estimator variance. Another frequently used rank estimator is $\hat{F}(i)=(i-a)/(n-2a+1)$ [18]; however, no literature study has investigated the determination of *a* values that vary with sample size to achieve unbiased (or minimum variance estimation) of Weibull parameters using this rank estimator.

Motivated by the performance of the formulae with variable *a* and *b*, this study focuses on developing a rank estimator for OLS estimation of the Weibull modulus with minimum Mean Squared Error (MSE) for complete samples. The previously discussed rank estimators, namely $\hat{F}(i)=(i-a)/(n+b)$ and $\hat{F}(i)=(i-a)/(n-2a+1)$, are linear functions of the ranks of order statistic *i*. Suppose that there is a rank estimator $\hat{F}(i)=h(i)$ minimizing the MSE of OLS estimator of Weibull modulus, where $h(i)$ is a nonlinear function of *i* and its coefficients vary with the sample size. According to the Taylor expansion of $h(i)$, a second-degree polynomial is the simplest nonlinear approximation for $h(i)$. Therefore, this study proposes, a quadratic function, $\hat{F}(i)=(i-a-ci^2)/(n+b)$ as a nonlinear rank estimator, and demonstrates that, by a systematic optimization of the parameters *a*, *b* and *c* for each sample size, estimators with significantly smaller MSE can be obtained.

Finally, all the Monte Carlo simulations are coded and run in the R programming language which uses Mersenne-Twister random number generator as the default generator whose cycle period is 219937-1 [19]. The following sections are organized as follows: After a concise discussion of the OLS method, a systematic analysis is introduced to develop a new rank estimator for estimating Weibull modulus using Monte Carlo simulations as the basic tool. The rank estimator, proposed as a result of this analysis, is then compared with the OLS methods with commonly used rank estimators.

2. ORDINARY LEAST SQUARES(OLS) METHOD

Equation (1) becomes a straight line by a double logarithmic transformation:

$$y=ln\left[-ln\left(1-F(X)\right)\right]=\beta\,ln\,X-\beta\,ln\,\alpha\,. \tag{2}$$

This is in linear form and the least squares method can be used to estimate Weibull parameters. Suppose that $X_1, X_2, ..., X_n$ are a random sample from Equation (1), and that $X_{(1)}, X_{(2)}, ..., X_{(n)}$ are the associated order statistics. Then Equation (2) can be can be rewritten as

$$y_i = ln\left[-ln\left(1-F\left(x_{(i)}\right)\right)\right] = \beta \, ln \, x_{(i)} - \beta \, ln \, \alpha \tag{3}$$

where $x_{(1)}, x_{(2)}, ..., x_{(n)}$ are ordered observations, and $F\left(x_{(i)}\right)$ values can be estimated using various rank estimators $\hat{F}\left(x_{(i)}\right)$. Considering the familiar form of a regression equation, $Y = aZ + b$, the left side of Equation (3) corresponds to *Y*, $ln\,X$ corresponds to *Z*, β corresponds to *a*, and $-\beta \, ln \, \alpha$ corresponds to *b*. Using $x_{(i)}$ and F_i pairs in Equation (3), *a* and *b* are obtained by the OLS procedure. Then the Weibull parameter estimates are calculated as $\hat{\beta} = a$ and $\hat{\alpha} = exp\left(-b / \hat{\beta}\right)$.

The most common rank estimators, $\hat{F}\left(x_{(i)}\right)$, or simply $\hat{F}(i)$, of $F\left(x_{(i)}\right)$ that are used for complete samples are median rank,

$$\hat{F}(i) = (i-0.3)/(n+0.4) \tag{4a}$$

mean rank,

$$\hat{F}(i) = i/(n+1) \tag{4b}$$

and Hazen rank,

$$\hat{F}(i) = (i-0.5)/n \tag{4c}$$

[11,20]. There are also some recently proposed rank estimators that are discussed in the previous section. A comprehensive list of rank estimators used for complete samples can be found in [10].

An important property of the OLS estimators is that $\hat{\beta}/\beta$ is a pivotal statistic, hence its distribution does not depend on the parameters α and β [21]. This allows simulating samples with a particular choice of α and β values. In this study they are set as $\alpha = 1$ and $\beta = 1$, then the results will be valid any other α and β values.

3. A NEW RANK ESTIMATOR

The rank estimators in Equations (4a)-(4c) as well as the ones in the form of $\hat{F}(i) = (i-a)/(n+b)$ with *a* and *b* varying with *n*, are either a measure of central tendency for the random variable $F\left(x_{(i)}\right)$, or close to one such measure for each *i*; because, *a* and *b* are between 0 to 1. This study deviates from this common approach in that it uses a quadratic function of *i*;

$$\hat{F}(i) = \left(i - a - ci^2\right)/(n+b) \tag{5}$$

with three parameters, *a*, *b*, and *c*. This function is expected to produce estimates that are significantly away from the mean rank or the median rank, at least for some *i*.

The Basic Simulation (BS) procedure employed in this study involves generating a sample of *n* values from a Weibull distribution with parameters $\alpha = 1$ and $\beta = 1$; estimating $\hat{\beta}$ using OLS with a combination of *a*, *b* and *c* values in Equation (5) (or any rank estimator with one or two fixed parameters such as in Equations (4a)-(4c) that will be used for comparison) and repeating this *R* times to compute the MSE of $\hat{\beta}$.

In order for the left-hand side in Equation (3) to exist, $\hat{F}(i)$ should be strictly greater than 0 and less than 1. To satisfy this condition, we keep the numerator in Equation (5) strictly between 0 and *n* for all $i = 1, \ldots, n$, and the denominator greater than or equal to *n*. Our initial trials with the BS procedure using various *a*, *b* and *c* values showed that negative *c* values result in larger MSE values as compared to positive ones. They also showed that non-negative *a* values along with non-negative *c* values consistently produced smaller MSEs. As a result, we decided to use non-negative *a* and *c* values throughout this study keeping $(i - a - ci^2)$ in Equation (5) strictly between 0 and *n*. Also, nonnegative *b* values are used to satisfy $0 < \hat{F}(i) < 1$.

$\hat{F}(i)$ should be a strictly increasing function of *i*, thus, $d\hat{F}(i)/di$ should be positive:

$$d\hat{F}(i)/di = (-2ci + 1)/(n + b) > 0 . \tag{6}$$

Then, $c < 1/2i, i = 1, \ldots, n$, and as a result

$$c < 1/2n \tag{7}$$

Also $0 < \hat{F}(i)$:

$$(i - a - ci^2)/(n + b) > 0 \tag{8}$$

Then

$$c < (i - a)/i^2 . \tag{9}$$

The right-hand side of this inequality takes its maximum value at either *i* = 1 or *i* = *n*. This result along with Equation (7) can be summarized as follows:

$$c \le min\{1/(2n), (n - a)/n^2, 1 - a\} . \tag{10}$$

There is also an upper bound for $\hat{F}(i)$; $\hat{F}(i) < 1$, hence $(i - a - ci^2) < (n + b)$. Then, $c > -(a + b + n - i)/i^2$. The right-hand side of this inequality takes its maximum value at *i* = *n*: $c > -(a + b)/n^2$. However, in this study, only nonnegative values of *a*, *b* and *c* are considered, which renders this inequality redundant.

Equation (10) reveals that there is a relationship between parameters *a* and *c* in Equation (5), however, *b* can be chosen independently of *a* and *c*. Finally $a < 1$, otherwise, $\hat{F}(1)$ in Equation (5) becomes negative for nonnegative *c* values.

Consequently, in this study, the values of *a*, *b* and *c* will be generated in such a way that $b \geq 0$, $0 \leq a < 1$, $c \geq 0$ and satisfying Equation (10).

4. THE GENERAL SIMULATION APPROACH

MSE in the BS procedure is a stochastic function of *a*, *b* and *c*: *MSE*(*a*, *b*, *c*). The aim is to find an optimal (*a*, *b*, *c*) combination with the minimum *MSE*(*a*, *b*, *c*) value for each *n*, and if possible, to formulate each parameter as a smooth continuous function of *n*. One particular problem about this formulation is that a parameter's value may change inconsistently with changing *n* [10]. One possible reason for this is the existence of multiple optima or similarly a flat response surface in certain regions of the parameter space. To avoid producing such inconsistent parameter values, we decided to use exhaustive enumeration of the parameters instead of using a multivariate optimization procedure such as the Nelder-Mead method.

To this end, the general approach in this study is to use extensive Monte-Carlo simulations using the BS procedure at every *a*, *b* and *c* value combination for a selected parameter space. In order to reduce prohibitively long simulation run times, we will first increase the value of each parameter with larger increments of δ with a smaller simulation run number *R;* then will decrease δ and increase *R* gradually to increase simulation precision. This will be illustrated immediately in the following section.

Still, working with three parameters simultaneously is computationally expensive: A useful strategy for this is to fix some parameter that is not as significant on the MSE performance as the others. It turns out that the parameter *b* is a good candidate for this purpose, since its value can be specified independently of *a* and *c*.

5. SEARCH FOR A VALUE OF THE PARAMETER *b* WITH THE SMALLEST MSE

A reasonable search interval for the parameter *b* is [0, 1], because numbers larger than 1 is likely to prevent obtaining any rank estimator values close to mean or median rank; the quadratic form of the numerator of Equation (5) already has such a potential negative effect. Therefore, this choice will allow obtaining rank estimators close to a measure of central tendency if such a choice produces a minimal MSE. Initially it would be natural to choose *b* = 0, 0.4 and 1, following Equations (4a)-(4c).

First a small set of sample sizes *n* = 3, 5, 10, 20, 40, 80, 120 and 150, were selected. Then for each sample size, *a* was changed from 0.02 to 0.98 with increments of δ_a = 0.02, and *c* was changed from 0.002 to an upper limit defined by Equation (10) with δ_c = 0.002 for each *a* value. For each (*a*, *c*) combination generated in this way, and for each value of *b* = 0, 0.4 and 1, the BS procedure was run with *R* = 500,000. Then 3-D plots of the root mean squared error, *RMSE*(*a*, *b*, *c*), were drawn for each *b*. After examining the plots for possible locations of a minimum, the search space was narrowed, increments were reduced as δ_a = 0.01 and δ_c = 0.001, and *R* was increased to 1,000,000. Figure 1 illustrates the response surface of the *RMSE*(*a*, *b*, *c*) function as 3-D plots for *b* = 0, 0.4 and 1, and for *n* = 10, 40 and 80 on a narrow parameter space of *a* and *c*.

Similar plots drawn for other *n* (*n* = 3, 5, 20, 120 and 150) as well as the plots in Figure 1 indicated that for any particular *n* the response surfaces are nearly identical for *b* values of 0, 0.4, and 1. Furthermore, the RMSE values for these *b* values were very similar, as depicted in Figure 1. However, *b* = 1 consistently resulted in the lowest RMSE values, thus making it the preferred value to be used in Equation (5). Given the fixed value of *b*, the remaining objective is to determine the optimal *a* and *c* values that produce the minimum RMSE (or equivalently minimum MSE). These values were already computed for the small set of sample sizes used in this section. In the following section, we will present a computationally efficient method for computing the optimal a and c values for all the remaining sample sizes, ranging from 3 to 150.

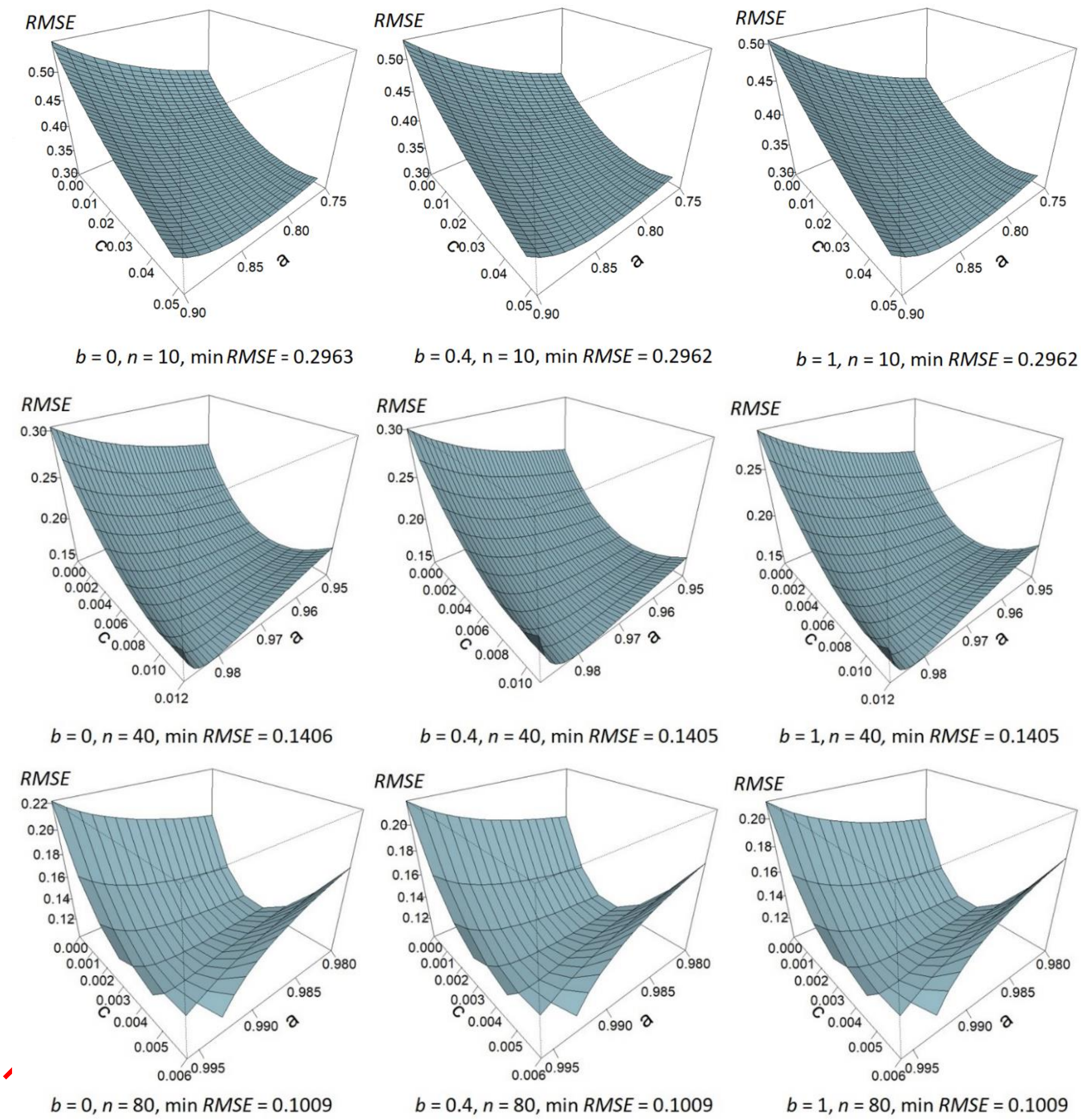

Figure 1. *RMSE values for various a, b and c values for n =10, 40 and 80*

6. SEARCH FOR THE OPTIMAL VALUES OF THE PARAMETERS *a* AND *c*

The optimal values of parameters *a* and *c* were initially computed for $\delta_a = 0.01$ and $\delta_c = 0.001$ for $b = 1$. By conducting a search in the vicinity of these optimal values with a much higher precision of $\delta_a = 0.0001$ and $\delta_c = 0.0001$ (the final precision values used in this study), the following results are obtained:

Table 1. *Optimal a and c values for selected n (b = 1)*

n	optimal *a*	optimal *c*
3	0	0.1666
5	0.5799	0.1000
10	0.8298	0.0500
20	0.9382	0.0250
40	0.9825	0.0125
80	0.9943	0.0048
120	0.9971	0.0025
150	0.9980	0.0018

Based on the increasing and decreasing values in the second and third columns of Table 1, we assumed that the optimal value of a is an increasing function of n, while the optimal value of c is a decreasing function of n. Then, for the case of $40 < n < 80$, the optimal values of a should be between 0.9825 and 0.9943, and optimal values of c should be between 0.0048 and 0.0125, according to Table 1. We then performed a search to find the optimums for the mid-point of the interval [40, 80], $n = 60$, with a precision of $\delta_a = 0.001$ and $\delta_c = 0.001$, and R = 1,000,000. This requires producing all a values in the interval [0.9825, 0.9943] with increments of δ_a, and producing all c values in the interval [0.0048, 0.0125] that are satisfying Equation (10) with increments of δ_c.

This resulted in the optimal values of a and c being 0.991 and 0.007, respectively. We then conducted a search in the vicinity of these optimal values, using the final precision values of $\delta_a = 0.0001$ and $\delta_c = 0.0001$, producing the final optimums of 0.9908 and 0.0071 for $n = 60$. We added these values as an additional row for $n = 60$ in Table 1. After this update, for n between 40 and 60, the optimal values of a should be between 0.9825 and 0.9908 (the upper bound is updated), and the optimal values of c should be between 0.0048 and 0.0071 (the upper bound is updated). Then this procedure was repeated targeting the optimums for $n = 50$.

Using this simulation optimization procedure in successively narrower intervals in an iterative manner, we computed the optimal values for each n value in the interval of [40,80]. We repeated the procedure for all initial intervals in Table 1 ([3,5], [5,10], [10,20], etc.), until we computed the optimal values of a and c for each n value between 3 and 150. The results are presented in Table 2.

We could have used a numerical algorithm such as the Nelder-Mead method that would produce the optimums much more efficiently, but they may be inconsistently changing (inconsistent ups and downs) with increasing n, as mentioned before. Also note that the optimal values of c happened at the upper limits defined by Equation (10) that would likely cause troubles in the application of numerical algorithms. Although being computationally expensive, our procedure produced smoothly increasing a values and smoothly increasing c values by increasing n.

While a numerical algorithm such as the Nelder-Mead method could have been used to find the optimums more efficiently, these methods may produce inconsistent results, as previously mentioned. Additionally, it should be noted that the optimal values of c occurred at the upper limits defined by Equation (10), which could potentially cause difficulties in the application of numerical algorithms. Although our simulation optimization procedure is computationally expensive, Table 2 shows that it yielded smoothly increasing a values and smoothly decreasing c values as n increased, providing a reliable and consistent method for finding the optimal values of a and c.

Table 2. *Optimal a and c values for* $\hat{F}(i) = \left(i - a - ci^2\right) / \left(n + b\right)$ *minimizing MSE; b = 1*

n	*a*	*c*	*n*	*a*	*c*	*n*	*a*	*c*	*n*	*a*	*c*
3	0	0.1666	43	0.9845	0.0115	83	0.9946	0.0045	123	0.9972	0.0025
4	0.3339	0.1250	44	0.9850	0.0111	84	0.9947	0.0044	124	0.9972	0.0026
5	0.5799	0.1000	45	0.9855	0.0107	85	0.9948	0.0043	125	0.9973	0.0024
6	0.6775	0.0833	46	0.9860	0.0104	86	0.9949	0.0042	126	0.9973	0.0024
7	0.7356	0.0714	47	0.9865	0.0101	87	0.9950	0.0042	127	0.9973	0.0025
8	0.7757	0.0625	48	0.9869	0.0098	88	0.9951	0.0041	128	0.9974	0.0023
9	0.8060	0.0555	49	0.9874	0.0096	89	0.9952	0.0040	129	0.9974	0.0023
10	0.8298	0.0500	50	0.9878	0.0094	90	0.9953	0.0040	130	0.9974	0.0024
11	0.8496	0.0454	51	0.9881	0.0090	91	0.9954	0.0039	131	0.9975	0.0022
12	0.8659	0.0416	52	0.9885	0.0088	92	0.9954	0.0037	132	0.9975	0.0023
13	0.8799	0.0384	53	0.9888	0.0085	93	0.9955	0.0037	133	0.9975	0.0023
14	0.8920	0.0357	54	0.9892	0.0084	94	0.9956	0.0037	134	0.9976	0.0021
15	0.9023	0.0333	55	0.9895	0.0082	95	0.9957	0.0037	135	0.9976	0.0022
16	0.9112	0.0312	56	0.9898	0.008p	96	0.9958	0.0036	136	0.9976	0.0022
17	0.9194	0.0294	57	0.9901	0.0078	97	0.9958	0.0035	137	0.9977	0.0020
18	0.9262	0.0277	58	0.9903	0.0075	98	0.9959	0.0035	138	0.9977	0.0021
19	0.9326	0.0263	59	0.9906	0.0074	99	0.9960	0.0034	139	0.9977	0.0021
20	0.9382	0.0250	60	0.9908	0.0071	100	0.9960	0.0033	140	0.9978	0.0019
21	0.9431	0.0238	61	0.9911	0.0070	101	0.9961	0.0033	141	0.9978	0.0019
22	0.9475	0.0227	62	0.9913	0.0068	102	0.9962	0.0033	142	0.9978	0.0020
23	0.9514	0.0217	63	0.9915	0.0067	103	0.9962	0.0032	143	0.9978	0.0020
24	0.9549	0.0208	64	0.9917	0.0065	104	0.9963	0.0032	144	0.9979	0.0018
25	0.9582	0.0200	65	0.9920	0.0065	105	0.9963	0.0030	145	0.9979	0.0019
26	0.9611	0.0192	66	0.9921	0.0062	106	0.9964	0.0031	146	0.9979	0.0019
27	0.9637	0.0185	67	0.9923	0.0061	107	0.9965	0.0030	147	0.9979	0.0019
28	0.9660	0.0178	68	0.9925	0.0060	108	0.9965	0.0030	148	0.9980	0.0018
29	0.9682	0.0172	69	0.9927	0.0059	109	0.9966	0.0029	149	0.9980	0.0018
30	0.9700	0.0166	70	0.9929	0.0058	110	0.9966	0.0030	150	0.9980	0.0018
31	0.9719	0.0161	71	0.9930	0.0056	111	0.9967	0.0029			
32	0.9736	0.0156	72	0.9932	0.0055	112	0.9967	0.0029			
33	0.9750	0.0151	73	0.9933	0.0054	113	0.9968	0.0028			
34	0.9764	0.0147	74	0.9935	0.0053	114	0.9968	0.0028			
35	0.9775	0.0142	75	0.9936	0.0052	115	0.9969	0.0027			
36	0.9787	0.0138	76	0.9938	0.0052	116	0.9969	0.0027			
37	0.9799	0.0135	77	0.9939	0.0050	117	0.9970	0.0026			
38	0.9808	0.0131	78	0.9940	0.0049	118	0.9970	0.0026			
39	0.9817	0.0128	79	0.9941	0.0047	119	0.9970	0.0027			
40	0.9825	0.0125	80	0.9943	0.0048	120	0.9971	0.0025			
41	0.9832	0.0121	81	0.9944	0.0047	121	0.9971	0.0026			
42	0.9839	0.0118	82	0.9945	0.0046	122	0.9972	0.0024			

7. A COMPARISON BETWEEN VALUES OF THE NEW RANK ESTIMATOR AND THE MEDIAN RANK ESTIMATOR

Before making any comparison in terms of efficiencies of the estimators, it will be helpful to see how different the values of the new rank estimator are from the values produced by the estimators in Equations (4a)-(4c). Choosing only one of them will be sufficient since they produce very close rank values; therefore, we selected the median rank in Equation (4a).

Figure 2 provides a graphical comparison between the values of the new rank estimator and the median rank estimator. The values of the former are always greater than the values of the former, and the difference becomes larger as the value of *i* increases.

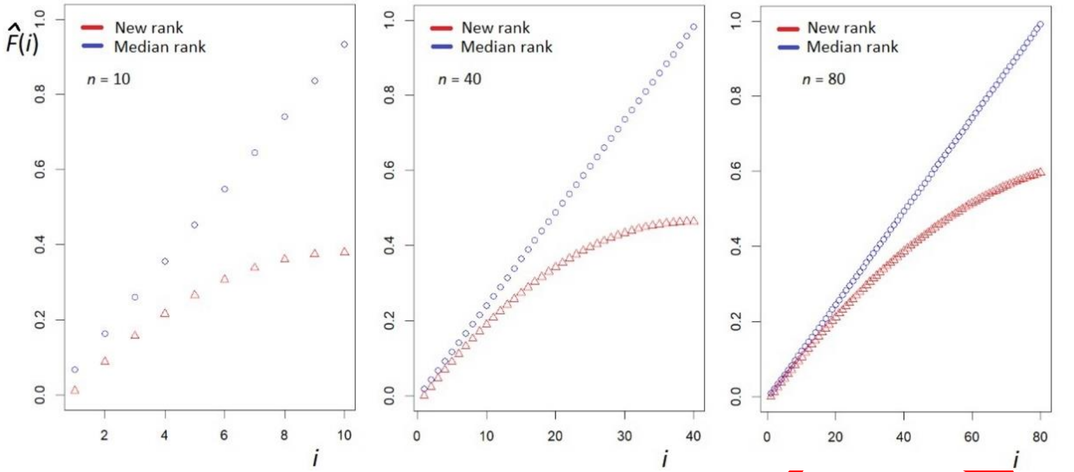

Figure 2. *A comparison between values of the new rank estimator and median rank estimator*

8. A COMPARISON IN TERMS OF RELATIVE EFFICIENCIES

In order to compare the performance of the new rank estimator with those in Equations (4a)-(4c), we will use relative efficiencies; we will also make a comparison to with the MLE method. Let $\hat{\beta}_m$, $\hat{\beta}_d$, $\hat{\beta}_h$, $\hat{\beta}_n$ and $\hat{\beta}_{ml}$ be the estimators computed by the OLS method using the mean rank, median rank, Hazen rank and the new rank estimators, and the MLE method, respectively. The relative efficiency of estimator $\hat{\beta}_1$ with respect to $\hat{\beta}_2$ is given by RE($\hat{\beta}_1$, $\hat{\beta}_2$)=MSE($\hat{\beta}_1$)/MSE($\hat{\beta}_2$). For example, if RE($\hat{\beta}_n$, $\hat{\beta}_d$) = 0.80, this would indicate that the necessary sample size for the OLS method using $\hat{\beta}_n$ is 80% of that needed for the OLS method using $\hat{\beta}_d$ to achieve approximately equal overall accuracy [8].

Table 3a-3b summarize the results obtained from running the BS procedure for each estimator with R = 1,000,000 for each sample size between 3 and 150. It presents MSE values and means of $\hat{\beta}_n$ as well as relative efficiencies for all the samples sizes.

First a comparison among the OLS estimators are presented: for sample sizes greater than 10, relative efficiencies are between 0.65 and 0.85: As n goes from 10 to 40, RE($\hat{\beta}_n$, $\hat{\beta}_m$) goes from 0.84 to 0.76 and stays close to 0.76 for n > 40; a similar behavior is observed for the other relative efficiencies. RE($\hat{\beta}_n$, $\hat{\beta}_d$) goes from 0.87 to 0.67, and stays almost the same for n > 40, and RE($\hat{\beta}_n$, $\hat{\beta}_h$) goes from 0.70 to 0.75, and stays almost the same for n > 40. As the sample decreases below 10, the relative efficiency with respect to mean rank and Hazen rank drops in an accelerating manner and reaches down to 0.16 and 0.11 for n = 3, respectively. For the median rank, it increases slightly first as n goes down to 6 from 10, and then it sharply decreases, reaching 0.25 for n = 3. Therefore, for any sample size between 3 and 150 the new rank estimator has a significant advantage over the other three rank estimators.

The new rank estimator is as efficient as the maximum likelihood estimator for n < 37 as shown in the last column of Table 3a. RE($\hat{\beta}_n$, $\hat{\beta}_{ml}$) is between 0.60 and 0.90 for n between 10 and 25; however, similar to the relative efficiencies of the other rank estimators, it drops acceleratingly as n decreases below 10 reaching down to a surprisingly low value of 0.07 for n = 3. The maximum likelihood estimator exhibits a better performance for n > 38; RE($\hat{\beta}_n$, $\hat{\beta}_{ml}$) increases almost linearly with n reaching up to 1.32 for n = 150. This is an expected result because of the optimality properties of the maximum likelihood estimators in large samples.

Finally, the mean values in Table 3 show that $\hat{\beta}_n$ values are biased; they underestimate the Weibull modulus (the true parameter value is 1, but all the expected values are less than 1). As the sample sizes increases the bias becomes smaller, and is less than 1% for n > 100.

Table 3a. *Comparison using relative efficiencies for n between 3 and 80*

n	MSE($\hat{\beta}_n$)	E($\hat{\beta}_n$)	RE($\hat{\beta}_n,\hat{\beta}_m$)	RE($\hat{\beta}_n,\hat{\beta}_d$)	RE($\hat{\beta}_n,\hat{\beta}_h$)	RE($\hat{\beta}_n,\hat{\beta}_{ml}$)
3	1.318	0.518	0.16	0.25	0.11	0.07
4	0.388	0.612	0.46	0.66	0.32	0.19
5	0.243	0.758	0.66	0.87	0.47	0.26
6	0.177	0.824	0.76	0.92	0.56	0.36
7	0.140	0.861	0.81	0.92	0.62	0.44
8	0.117	0.885	0.83	0.91	0.66	0.49
9	0.100	0.902	0.84	0.89	0.68	0.54
10	0.088	0.914	0.84	0.87	0.70	0.58
11	0.078	0.924	0.84	0.85	0.71	0.61
12	0.070	0.931	0.84	0.83	0.72	0.65
13	0.064	0.937	0.84	0.82	0.73	0.68
14	0.059	0.942	0.83	0.80	0.74	0.71
15	0.055	0.947	0.83	0.79	0.74	0.73
16	0.051	0.950	0.83	0.78	0.74	0.75
17	0.048	0.953	0.82	0.77	0.75	0.78
18	0.045	0.956	0.82	0.76	0.75	0.80
19	0.042	0.959	0.81	0.75	0.75	0.82
20	0.040	0.961	0.81	0.74	0.75	0.84
21	0.038	0.963	0.80	0.74	0.75	0.85
22	0.036	0.964	0.80	0.73	0.75	0.86
23	0.035	0.966	0.80	0.73	0.75	0.87
24	0.033	0.967	0.79	0.72	0.75	0.89
25	0.032	0.968	0.79	0.71	0.75	0.91
26	0.030	0.970	0.79	0.71	0.75	0.90
27	0.029	0.971	0.78	0.71	0.75	0.91
28	0.028	0.972	0.78	0.70	0.75	0.93
29	0.027	0.973	0.78	0.70	0.75	0.95
30	0.026	0.973	0.78	0.70	0.75	0.95
31	0.025	0.974	0.77	0.69	0.75	0.97
32	0.025	0.975	0.77	0.69	0.75	0.97
33	0.024	0.976	0.77	0.69	0.75	0.98
34	0.023	0.976	0.77	0.68	0.75	1.00
35	0.023	0.976	0.77	0.68	0.75	1.00
36	0.022	0.977	0.76	0.68	0.75	1.00
37	0.021	0.978	0.76	0.68	0.75	1.01
38	0.021	0.979	0.76	0.68	0.75	1.02
39	0.020	0.979	0.76	0.68	0.75	1.02
40	0.020	0.979	0.76	0.67	0.75	1.03
41	0.019	0.980	0.76	0.67	0.75	1.03

n	MSE($\hat{\beta}_n$)	E($\hat{\beta}_n$)	RE($\hat{\beta}_n,\hat{\beta}_m$)	RE($\hat{\beta}_n,\hat{\beta}_d$)	RE($\hat{\beta}_n,\hat{\beta}_h$)	RE($\hat{\beta}_n,\hat{\beta}_{ml}$)
42	0.019	0.980	0.76	0.67	0.75	1.05
43	0.018	0.980	0.76	0.67	0.75	1.05
44	0.018	0.980	0.76	0.67	0.75	1.05
45	0.018	0.981	0.76	0.67	0.75	1.06
46	0.017	0.981	0.76	0.67	0.75	1.06
47	0.017	0.982	0.75	0.67	0.75	1.07
48	0.017	0.982	0.75	0.67	0.75	1.08
49	0.016	0.983	0.75	0.67	0.75	1.08
50	0.016	0.982	0.75	0.66	0.75	1.08
51	0.016	0.983	0.75	0.66	0.75	1.10
52	0.015	0.984	0.75	0.66	0.75	1.10
53	0.015	0.984	0.75	0.66	0.75	1.10
54	0.015	0.984	0.75	0.66	0.75	1.11
55	0.015	0.984	0.75	0.66	0.75	1.11
56	0.014	0.984	0.75	0.66	0.75	1.11
57	0.014	0.985	0.75	0.66	0.75	1.11
58	0.014	0.985	0.75	0.66	0.76	1.12
59	0.014	0.985	0.75	0.66	0.76	1.13
60	0.013	0.986	0.75	0.66	0.76	1.13
61	0.013	0.986	0.75	0.66	0.76	1.13
62	0.013	0.986	0.75	0.66	0.76	1.14
63	0.013	0.985	0.75	0.66	0.76	1.14
64	0.013	0.986	0.75	0.66	0.76	1.14
65	0.012	0.987	0.75	0.66	0.76	1.14
66	0.012	0.986	0.75	0.66	0.76	1.14
67	0.012	0.986	0.75	0.66	0.76	1.14
68	0.012	0.986	0.75	0.66	0.76	1.15
69	0.012	0.987	0.75	0.66	0.76	1.16
70	0.012	0.988	0.75	0.66	0.76	1.16
71	0.011	0.987	0.75	0.66	0.76	1.17
72	0.011	0.988	0.75	0.66	0.76	1.17
73	0.011	0.987	0.75	0.66	0.76	1.17
74	0.011	0.988	0.75	0.66	0.76	1.18
75	0.011	0.987	0.75	0.66	0.76	1.18
76	0.011	0.988	0.75	0.66	0.76	1.18
77	0.011	0.988	0.75	0.66	0.76	1.18
78	0.010	0.988	0.75	0.66	0.76	1.18
79	0.010	0.989	0.75	0.66	0.76	1.19
80	0.010	0.989	0.75	0.66	0.76	1.20

Table 3b. *Comparison using relative efficiencies for n between 81 and 150*

n	MSE($\hat{\beta}_n$)	E($\hat{\beta}_n$)	RE($\hat{\beta}_n,\hat{\beta}_m$)	RE($\hat{\beta}_n,\hat{\beta}_d$)	RE($\hat{\beta}_n,\hat{\beta}_h$)	RE($\hat{\beta}_n,\hat{\beta}_{ml}$)	n	MSE($\hat{\beta}_n$)	E($\hat{\beta}_n$)	RE($\hat{\beta}_n,\hat{\beta}_m$)	RE($\hat{\beta}_n,\hat{\beta}_d$)	RE($\hat{\beta}_n,\hat{\beta}_h$)	RE($\hat{\beta}_n,\hat{\beta}_{ml}$)
81	0.010	0.989	0.75	0.66	0.76	1.21	116	0.007	0.991	0.75	0.66	0.77	1.26
82	0.010	0.989	0.75	0.66	0.76	1.20	117	0.007	0.994	0.75	0.67	0.77	1.27
83	0.010	0.989	0.75	0.66	0.76	1.20	118	0.007	0.992	0.75	0.67	0.77	1.26
84	0.010	0.989	0.75	0.66	0.76	1.20	119	0.007	0.990	0.75	0.67	0.77	1.27
85	0.010	0.989	0.75	0.66	0.76	1.21	120	0.007	0.993	0.75	0.67	0.77	1.27
86	0.010	0.989	0.75	0.66	0.76	1.20	121	0.007	0.991	0.75	0.67	0.77	1.27
87	0.009	0.989	0.75	0.66	0.76	1.21	122	0.007	0.994	0.75	0.67	0.77	1.27
88	0.009	0.989	0.75	0.66	0.76	1.20	123	0.007	0.992	0.75	0.67	0.77	1.27
89	0.009	0.990	0.75	0.66	0.76	1.21	124	0.007	0.993	0.75	0.67	0.77	1.27
90	0.009	0.990	0.75	0.66	0.76	1.22	125	0.007	0.994	0.75	0.67	0.77	1.28
91	0.009	0.991	0.75	0.66	0.76	1.21	126	0.007	0.991	0.75	0.67	0.77	1.28
92	0.009	0.989	0.75	0.66	0.76	1.22	127	0.007	0.992	0.75	0.67	0.77	1.28
93	0.009	0.989	0.75	0.66	0.76	1.23	128	0.007	0.993	0.75	0.67	0.77	1.28
94	0.009	0.990	0.75	0.66	0.77	1.23	129	0.007	0.991	0.75	0.67	0.78	1.28
95	0.009	0.991	0.75	0.66	0.77	1.22	130	0.006	0.991	0.76	0.67	0.78	1.28
96	0.009	0.992	0.75	0.66	0.77	1.24	131	0.006	0.993	0.75	0.67	0.78	1.29
97	0.008	0.989	0.75	0.66	0.77	1.23	132	0.006	0.993	0.75	0.67	0.78	1.29
98	0.008	0.990	0.75	0.66	0.77	1.23	133	0.006	0.991	0.76	0.67	0.78	1.29
99	0.008	0.992	0.75	0.66	0.77	1.22	134	0.006	0.994	0.76	0.67	0.78	1.28
100	0.008	0.989	0.75	0.66	0.77	1.23	135	0.006	0.993	0.76	0.67	0.78	1.29
101	0.008	0.990	0.75	0.66	0.77	1.24	136	0.006	0.991	0.76	0.67	0.78	1.30
102	0.008	0.992	0.75	0.66	0.77	1.23	137	0.006	0.995	0.76	0.67	0.78	1.28
103	0.008	0.989	0.75	0.66	0.77	1.24	138	0.006	0.994	0.76	0.67	0.78	1.29
104	0.008	0.991	0.75	0.66	0.77	1.25	139	0.006	0.992	0.76	0.67	0.78	1.30
105	0.008	0.990	0.75	0.66	0.77	1.25	140	0.006	0.996	0.76	0.67	0.78	1.30
106	0.008	0.990	0.75	0.66	0.77	1.25	141	0.006	0.994	0.76	0.67	0.78	1.30
107	0.008	0.993	0.75	0.66	0.77	1.25	142	0.006	0.993	0.76	0.67	0.78	1.30
108	0.008	0.990	0.75	0.66	0.77	1.24	143	0.006	0.991	0.76	0.67	0.78	1.30
109	0.008	0.993	0.75	0.66	0.77	1.27	144	0.006	0.996	0.76	0.67	0.78	1.32
110	0.008	0.990	0.75	0.66	0.77	1.25	145	0.006	0.995	0.76	0.67	0.78	1.31
111	0.007	0.993	0.75	0.66	0.77	1.26	146	0.006	0.993	0.76	0.67	0.78	1.30
112	0.007	0.990	0.75	0.66	0.77	1.25	147	0.006	0.991	0.76	0.67	0.78	1.31
113	0.007	0.993	0.75	0.66	0.77	1.27	148	0.006	0.997	0.76	0.67	0.78	1.32
114	0.007	0.990	0.75	0.66	0.77	1.26	149	0.006	0.995	0.76	0.67	0.78	1.31
115	0.007	0.993	0.75	0.66	0.77	1.27	150	0.006	0.993	0.76	0.67	0.78	1.32

9. CONCLUSIONS

This paper proposes a nonlinear rank estimator, for the first time in the literature, to be used with the ordinary least squares method for estimating the Weibull modulus, a key parameter in reliability analysis. The new rank estimator is a quadratic function of the ranks of order statistics, with three parameters that are optimized by an extensive simulation procedure. The paper compares the performance of the new rank estimator with three commonly used rank estimators in terms of MSE and relative efficiency.

The results show that the new rank estimator consistently outperforms the mean rank, median rank, Hazen rank other rank estimators, when used with the OLS method, for sample sizes ranging from 3 to 150, with relative efficiencies between 0.65 and 0.85 for sample sizes greater than 10; much smaller relative efficiencies are achieved for sample sizes 3 and 4. It also outperforms the maximum likelihood estimator for sample sizes less than 37. Moreover, the new rank estimator generates biased estimates of the Weibull modulus, but the bias decreases as the sample size increases. Another interesting finding is that the new rank estimator produces values that are significantly different from the mean or median rank, especially for larger order statistics.

The study has several implications, including the potential to apply other nonlinear functions, such as cubic polynomials, to improve the efficiency of the parameter estimators of the ordinary least squares method. Additionally, it highlights that rank estimators for ordinary least squares analysis do not necessarily have to be a measure of central tendency for the random variable $F\left(x_{(i)}\right)$, and the functional properties of different rank estimators for various purposes, such as minimum-variance unbiased estimation, may be worth exploring in future research. Furthermore, the research may be extended to incorporate weighted least squares methods. Finally, quadratic or other nonlinear rank estimators can be applied to other distributions with distribution functions having a linear form after a double logarithmic transformation, such as the lognormal or extreme value distributions.

However, the research does have some limitations, including its focus on complete samples and not accounting for censored or truncated data, which are common in reliability studies. Moreover, it does not compare the new rank estimator with other estimation methods, such as the maximum likelihood method or the weighted least squares method, or analyze confidence intervals for the parameters. Lastly, while the paper is focused on the Weibull modulus, it does not consider the scale parameter. Nonetheless, this may be justifiable due to the importance of the Weibull modulus, and the scale parameter can be estimated separately using a different method. Future research could address these limitations and further explore the potential applications and generalizations of the new rank estimator.

CONFLICTS OF INTEREST

No conflict of interest was declared by the author.

REFERENCES

[1] Weibull, W., “A statistical distribution function of wide applicability”, Journal of Applied Mechanics, (1951).

[2] Lai, C.D., Murthy D., Xie, M., “Weibull distributions and their applications”, Springer Handbooks, Springer, 63–78, (2006).

[3] Rinne, H., “The Weibull distribution: a handbook”, Chapman and Hall/CRC, (2008).

[4] Barbero, E., Fernández-Sáez, J., Navarro, C., “Statistical analysis of the mechanical properties of composite materials”, Composites Part B: Engineering, 31: 375–81, (2000).

[5] Yang, X., Xie, L., Yang, Y., Zhao, B., Li, Y., "A comparative study for parameter estimation of the Weibull distribution in a small sample size: An application to spring fatigue failure data", Quality Engineering, 1-13, (2022).

[6] Babacan, E.K., Kaya, S., "A simulation study of the Bayes estimator for parameters in Weibull distribution", Communications Faculty of Sciences University of Ankara-Series A1 Mathematics and Statistics, 68:1664–74, (2019).

[7] Babacan, E.K., Kaya, S., "Comparison of parameter estimation methods in Weibull Distribution", Sigma Journal of Engineering and Natural Sciences, 38: 1609–21, (2020).

[8] Genschel, U., Meeker, W.Q., "A comparison of maximum likelihood and median-rank regression for Weibull estimation", Quality Engineering, 22: 236–55, (2010).

[9] Jia, X., Xi, G., Nadarajah, S., "Correction Factor for Unbiased Estimation of Weibull Modulus by the Linear Least Squares Method", Metallurgical and Materials Transactions A, 50: 2991–3001, (2019).

[10] Davies, I.J., "Unbiased estimation of Weibull modulus using linear least squares analysis—A systematic approach", Journal of the European Ceramic Society, 37: 369–80, (2017).

[11] Kantar, Y.M., "Generalized least squares and weighted least squares estimation methods for distributional parameters", REVSTAT-Statistical Journal, 13: 263–82, (2015).

[12] Wu, D., Lu, G., Jiang, H., Li, Y., "Improved estimation of Weibull parameters with the linear regression method", Journal of the American Ceramic Society, 87: 1799–802, (2004).

[13] Wu, D., Zhou, J., Li, Y., "Unbiased estimation of Weibull parameters with the linear regression method", Journal of the European Ceramic Society, 26: 1099–105, (2006).

[14] Tiryakioğlu, M., "An unbiased probability estimator to determine Weibull modulus by the linear regression method", Journal of Materials Science, 41: 5011–3, (2006).

[15] Tiryakioğlu, M., Hudak, D., "On estimating Weibull modulus by the linear regression method", Journal of Materials Science, 42: 10173–9, (2007).

[16] Davies, I.J., "Unbiased estimation of the Weibull scale parameter using linear least squares analysis", Journal of the European Ceramic Society, 37: 2973–81, (2017).

[17] Song, L., Wu, D., Li, Y., "Optimal probability estimators for determining Weibull parameters", Journal of Materials Science Letters, 22: 1651–1653, (2003).

[18] D'Agostino, R., "Goodness-of-fit-techniques", Routledge, (2017).

[19] Dutang C., Wuertz, D., "A note on random number generation". https://cran.r-project.org/web/packages/randtoolbox/vignettes/fullpres.pdf. Access date: 30.06.2024.

[20] Bütikofer, L., Stawarczyk, B., Roos, M., "Two regression methods for estimation of a two-parameter Weibull distribution for reliability of dental materials", Dental Materials, 31: 33–50, (2015).

[21] Fernandez-Saez, J., Chao, J., Duran, J., Amo, J.M., "Estimating lower-bound fracture parameters for brittle materials", Journal of Materials Science Letters, 12: 1493–6, (1993).