

# Evaluation of the similarity index

A statistical procedure for comparing Weibull distributions

L.A.A. (Lotte) Kremer



---

# Evaluation of the similarity index

A statistical procedure for comparing Weibull  
distributions

---

by

L.A.A. (Lotte) Kremer

part of the BSc Applied Mathematics

at the Delft University of Technology,

to be defended publicly on Thursday June 29, 2023 at 2:45 PM.

Student number: 4861957  
Thesis committee: Dr. ir. L.E. Meester, TU Delft, supervisor  
Dr. J. Söhl, TU Delft  
Prof. dr. R. Ross, TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



# Abstract

In this thesis, we evaluate the statistical procedure based on the similarity index proposed by Ypma and Ross [6]. The function for the similarity index is defined as:

$$S(t_1, t_2) = \frac{\int_{t_1}^{t_2} f_1(t) f_2(t) dt}{\int_{t_1}^{t_2} [f_1(t)]^2 + [f_2(t)]^2 - f_1(t) f_2(t) dt} \quad (1)$$

It consists of two probability density functions, assumed to be related to the Weibull distribution. The evaluated statistical test aims to verify whether an observation of censored lifetime data is compliant with a given reference distribution. The test is compared to both the likelihood ratio test and a test using a variation of the similarity index. The comparison between these tests is based on their power function.

First, a clear explanation of the complete statistical procedure using the similarity index is provided. Then the observations for which the test can be used are explained further. Its application and small limitations are shown using an example. Then, the two alternative tests are introduced. This will be followed by a theoretical overview of the power comparison method. Finally, the simulation for the power comparison is conducted. The power function is estimated for multiple relevant base cases along a few alternative parameter lines. It can then be concluded that the likelihood ratio test has consistently higher power than the similarity index test. However, the variation of the similarity index demonstrates a varying power, with instances of both higher and lower values than the original similarity index test.



# Contents

1	Introduction	1
2	Ypma-Ross similarity index procedure	3
2.1	Introduction to the statistical test . . . . .	3
2.1.1	Censored failure times . . . . .	3
2.1.2	Formal definition . . . . .	4
2.2	The similarity index formula . . . . .	4
2.2.1	Range of the similarity index. . . . .	5
2.2.2	Analysis of the integrals . . . . .	5
2.3	The procedure . . . . .	6
2.3.1	Ordinary Least Squares estimation on the order statistics . . . . .	6
2.3.2	Computing the empirical distribution . . . . .	7
2.4	Example . . . . .	8
2.4.1	Method . . . . .	8
2.4.2	Results . . . . .	9
2.5	Conclusion . . . . .	10
3	Analysis for comparison	11
3.1	Likelihood ratio test. . . . .	11
3.1.1	Test statistic . . . . .	11
3.1.2	Maximum likelihood estimation . . . . .	12
3.1.3	Overview of the test . . . . .	13
3.2	Variant of the similarity index. . . . .	13
3.3	Power comparison . . . . .	14
3.3.1	Power function. . . . .	14
3.3.2	Limitations. . . . .	14
3.4	Conclusion . . . . .	15
4	Simulation	17
4.1	Base cases. . . . .	17
4.1.1	Relevant null hypotheses. . . . .	18
4.1.2	Additional parameters . . . . .	19

4.2	Alternative Weibull parameters . . . . .	20
4.2.1	Directions . . . . .	20
4.2.2	Compliance . . . . .	22
4.3	Results . . . . .	22
4.3.1	Univariate variation: $\alpha$ . . . . .	22
4.3.2	Univariate variation: $\beta$ . . . . .	23
4.3.3	Bivariate variation: positive slope . . . . .	25
4.3.4	Bivariate variation: negative slope . . . . .	26
4.4	Conclusion . . . . .	26
5	Conclusion and discussion	29
A	Data example	31
B	Input values	33
B.1	Univariate variation: $\alpha$ . . . . .	33
B.2	Univariate variation: $\beta$ . . . . .	33
B.3	Bivariate variation: positive slope . . . . .	34
B.4	Bivariate variation: negative slope . . . . .	34
	Bibliography	37



# Introduction

In an article in 2017 by Ypma and Ross [6] a method was proposed for quantifying the similarity between two failure density functions in the paper 'Determining the Similarity between Observed and Expected Ageing Behavior'. They proposed using the formula given in Equation (1.1), which will be called the similarity index from now on.

$$S(t_1, t_2) = \frac{\int_{t_1}^{t_2} f_1(t) f_2(t) dt}{\int_{t_1}^{t_2} [f_1(t)]^2 + [f_2(t)]^2 - f_1(t) f_2(t) dt} \quad (1.1)$$

This similarity index takes as input two probability density functions. In the objective of the article, those are probability density functions of distributions that model the time until a certain object fails. In this research, the assumption is made that those distributions are Weibull distributions. This is a common assumption for the lifetimes of insulated electrical systems, which is the application proposed in the paper.

The authors aim to utilize the similarity index "to quantify the confidence that the observed failure distribution agrees with the prescribed (expected) failure distribution." They propose a statistical procedure based on the similarity index to achieve this goal. This procedure will be evaluated in this research. The primary objective is to compare the procedure with other appropriate statistical tests and determine whether the similarity index-based statistical procedure is the optimal choice.

In Chapter 2 the similarity index and the corresponding statistical test will be explained thoroughly. First, a formal definition of both the test and the input data for which it can be used will be given. This detailed information allows for a more insightful analysis of the similarity index and the values it attains. The chapter further explores how the similarity index is employed in the statistical procedure and outlines the specific steps involved. An example presented in the original paper by Ypma and Ross [6] is utilized to illustrate these concepts, while also discussing some limitations associated with the similarity index test.

Once a solid understanding of the similarity index test is established, Chapter 3 outlines the theoretical plan for comparison. Two alternative approaches, namely the likelihood ratio test and a test based on a variant of the similarity index, are introduced. The variant of the similarity index might be a solution for the discontinuities observed when using the similarity index. Alongside presenting these alternatives, the chapter provides a theoretical explanation of the comparison methodology: a power comparison between the three tests.

Since the power comparison can only be done through simulation, Chapter 4 focuses on conducting simulations. Multiple base cases are selected to reduce computational complexity. The power is calculated for the three tests in the base cases, considering various alternative parameter directions. Special emphasis is placed on the most important cases, relevant for verifying the compliance of breakdown data. Based on the results, conclusions are drawn regarding the similarity index-based test in comparison to its alternatives.

Chapter 5 will provide the concluding remarks of this research, along with discussing topics for further exploration and offering recommendations.



# 2

## Ypma-Ross similarity index procedure

To make informed choices regarding the comparison between the statistical test based on the similarity index proposed by Ypma and Ross [6] and other statistical tests, it is important to have a clear understanding of the specific procedures and the intended applications described by the authors. In this chapter, that insight and a thorough explanation of the whole procedure will be provided. Afterwards, the finding of this chapter will be used for some recommendations for the comparison plan which is constructed and executed in the following chapters.

The statistical test, based on the procedure developed by Ypma and Ross, will be introduced in Section 2.1. This section will clarify the type of data for which the test can be utilized and its intended purpose. It will be followed by an analysis of the similarity index formula in Section 2.2, considering its important role in the test. The proposed methodology for employing the test, including the use of estimations and simulations, will be fully described in Section 2.3. Subsequently, an illustrative example will be provided in Section 2.4, demonstrating the application of the method. Additionally, this example will highlight certain practical limitations associated with the use of the test.

### 2.1. Introduction to the statistical test

The particular scenario that served as the basis for the article by Ypma and Ross [6] was the application of the similarity index as seen in Equation (1.1) to compare two 'failure density distributions'. This term is used in the article to refer to a family of probability distributions commonly used when analyzing the lifetimes of products. More specifically, they want to determine whether such an observation of lifetimes is coming from a certain reference distribution or not. Before providing a formal definition of the statistical test based on this objective, it is important to develop an understanding of this particular type of data for which Ypma and Ross want to use the similarity index.

#### 2.1.1. Censored failure times

There are multiple probability distributions that are associated with the lifetimes of products. Such a distribution could even be a complicated mix of known distributions, as mentioned in Ypma and Ross [6]. However, in this thesis the assumption is made that the data is a realization of a sample of random variables following a Weibull distribution. This is also in correspondence with the examples used in the article by Ypma and Ross. The probability density function and the cumulative distribution function of a Weibull distribution are shown in Equation (2.1) and (2.2). The parameter  $\alpha > 0$  is the scale parameter, the parameter  $\beta > 0$  is the

shape parameter and the variable  $t > 0$  represents the time-to-failure.

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^\beta\right) \quad (2.1)$$

$$F(t) = 1 - \exp\left(-\left(\frac{t}{\alpha}\right)^\beta\right) \quad (2.2)$$

The data used for conducting the test will be realizations of an independent and identically distributed sample from the *Weibull*( $\alpha, \beta$ ) distribution. However, since the test's intended use in Ypma and Ross [6] also includes conducting the test before all lifetimes are known, this data could be censored. This means that the lifetime is not yet known for a part of the sample.

There are several types of censored data. An important distinction for censored data is type I and type II censoring. In type II censoring the lifetime of the first  $r$  components is known, where  $r$  is predefined. In type I censoring the lifetime is known for all components which failed before a certain time  $t_c$ , where  $t_c$  is predefined. In this thesis, in line with the procedure of Ypma and Ross, type I censored data will be considered. More information on both types can be found in section 8.3 of the Weibull handbook by Rinne [4].

### 2.1.2. Formal definition

With the previous subsection in mind, it is now possible to state a formal definition. The definition below will be used in the remaining part of this research when evaluating the statistical use of the similarity index.

Assume that lifetimes  $x_1, x_2, \dots, x_n$  are realizations of an independent and identically distributed sample with a *Weibull*( $\alpha, \beta$ ) distribution and let  $\theta = (\alpha, \beta)$ . The lifetimes are censored with type I censoring at time  $t_c$ . Moreover, let  $\theta_0 = (\alpha_0, \beta_0)$  represent the parameters of the reference distribution, which is a Weibull distribution as well. This results in the following null hypothesis (Equation (2.3)) and alternative hypothesis (Equation (2.4)):

$$H_0 : \theta = \theta_0 \quad (2.3)$$

$$H_1 : \theta \neq \theta_0 \quad (2.4)$$

The test statistic is the similarity index,  $S$ , as defined in Equation (1.1). The similarity index has values in  $[0, 1]$  and is analyzed further in Section 2.2. Let  $K$  be a critical region for this test statistic of the following form

$$K = \{S \leq c_\gamma\} \quad (2.5)$$

such that

$$\mathbb{P}_{\theta_0}(S \in K) = \gamma \quad (2.6)$$

The parameter  $\gamma$  is the significance level. Whenever the calculated value of the similarity index based on the observed data falls within the range of  $K$ , the null hypothesis is rejected. It implies that if the null hypothesis would be true, the probability that the value of the similarity index would be less than or equal to that particular value is less than or equal to  $\gamma$ .

## 2.2. The similarity index formula

Before delving into the specific procedure employed for conducting the statistical test, the central component of the test is analyzed more thoroughly: the similarity index. While the general form was presented in Equation (1.1), it is now possible to introduce the more specific form, which will be used in the statistical test.

Let  $f_1(t)$  and  $f_2(t)$  represent two probability density functions for the Weibull distribution with parameters  $(\alpha_1, \beta_1)$  and  $(\alpha_2, \beta_2)$ , respectively. The similarity index is expressed as follows:

$$S(0, t_c) = \frac{\int_0^{t_c} f_1(t) f_2(t) dt}{\int_0^{t_c} [f_1(t)]^2 + [f_2(t)]^2 - f_1(t) f_2(t) dt} \quad (2.7)$$

The motivation behind choosing this particular interval of integration,  $(0, t_c)$ , instead of the more general  $(t_1, t_2)$  is to assess the similarity of the probability density functions across the entire domain where failures have been observed. This is in correspondence to the method in Ypma and Ross [6].

### 2.2.1. Range of the similarity index

The similarity index takes on values in  $[0, 1]$ . This can be proved by rewriting Equation (2.7) into the following:

$$S(0, t_c) = \frac{\int_0^{t_c} f_1(t)f_2(t)dt}{\int_0^{t_c} f_1(t)f_2(t)dt + \int_0^{t_c} [f_1(t) - f_2(t)]^2 dt} \quad (2.8)$$

Since  $f_1(t)$  and  $f_2(t)$  are both probability density functions, they both have non-negative function values. Therefore, their product,  $f_1(t)f_2(t)$ , will have non-negative values as well. From this, it can be concluded that the integrals in Equation (2.8) are all non-negative. Thus the following holds:

$$\int_0^{t_c} f_1(t)f_2(t)dt + \int_0^{t_c} [f_1(t) - f_2(t)]^2 dt \geq \int_0^{t_c} f_1(t)f_2(t)dt \geq 0 \quad (2.9a)$$

$$(2.9b)$$

It follows that Equation (2.8), and therefore the similarity index, has values in  $[0, 1]$ . Also,  $S(0, t_c) = 1$  if and only if  $f_1(t) = f_2(t)$  for  $t \in (0, t_c)$ .

### 2.2.2. Analysis of the integrals

The convergence of the integrals in the similarity index is not guaranteed for all pairs of Weibull probability density functions. To assess the impact of this on the value of the similarity index, a proper look into the integrals is needed. The different integrals in Equation (2.7) can be expanded as follows:

$$\int_0^{t_c} f_1(t)f_2(t)dt = \frac{\beta_1\beta_2}{\alpha_1^{\beta_1}\alpha_2^{\beta_2}} \int_0^{t_c} t^{\beta_1+\beta_2-2} \exp\left(-\left(\frac{t}{\alpha_1}\right)^{\beta_1} - \left(\frac{t}{\alpha_2}\right)^{\beta_2}\right) dt \quad (2.10a)$$

$$\int_0^{t_c} [f_1(t)]^2 dt = \frac{\beta_1^2}{\alpha_1^{2\beta_1}} \int_0^{t_c} t^{2\beta_1-2} \exp\left(-2\left(\frac{t}{\alpha_1}\right)^{\beta_1}\right) dt \quad (2.10b)$$

$$\int_0^{t_c} [f_2(t)]^2 dt = \frac{\beta_2^2}{\alpha_2^{2\beta_2}} \int_0^{t_c} t^{2\beta_2-2} \exp\left(-2\left(\frac{t}{\alpha_2}\right)^{\beta_2}\right) dt \quad (2.10c)$$

In general, no closed expression exists for those integral. So to find out when the integrals might diverge it is helpful to find an upper or lower bound. An integral becomes an improper integral which might diverge if its function values go to infinity within the domain of the interval. In Equations (2.10a) up to (2.10c) the exponential function takes on values in  $(0, 1]$ , due to the negative power. So the parts that might go to infinity on the domain of the integral and the corresponding constraints for this are the following:

$$t^{\beta_1+\beta_2-2} \quad \text{if } \beta_1 + \beta_2 < 2 \quad (2.11a)$$

$$t^{2\beta_1-2} \quad \text{if } \beta_1 < 1 \quad (2.11b)$$

$$t^{2\beta_2-2} \quad \text{if } \beta_2 < 1 \quad (2.11c)$$

If none of the conditions in the equations above are satisfied, then all integrals are proper and thus finite. To determine whether the resulting integrals are convergent or divergent improper integrals, it can be useful to look at the integral in Equation (2.12). The functions over which the integrals are taken are all, up to some constant, bounded above and below by an integral of this form.

$$\lim_{\epsilon \rightarrow 0} \int_{\epsilon}^{t_c} t^p dt = \lim_{\epsilon \rightarrow 0} \frac{1}{p+1} \left( t_c^{p+1} - \epsilon^{p+1} \right) \quad (2.12)$$

This limit is finite if and only if  $p+1 > 0$ . Thus the integral on the left-hand side is in its limit a convergent (improper) integral if and only if  $p > -1$ . This information can be used to make a distinction between three possible situations for the integrals of the similarity index when at least one of the conditions in Equations (2.11a) up to (2.11c) is met.

**case 1: all convergent integrals** In the case that  $\beta_1 + \beta_2 < 2$ , but both  $\beta_1 > \frac{1}{2}$  and  $\beta_2 > \frac{1}{2}$ , all the integrals in the similarity index will be convergent improper integrals.

**case 2: one divergent integral** Whenever either  $\beta_1 \leq \frac{1}{2}$  or  $\beta_2 \leq \frac{1}{2}$ , but  $\beta_1 + \beta_2 > 1$ , the integral in Equation (2.10a) still converges. However, the integral in either Equation (2.10b) or Equation (2.10c) will then diverge. This results in a similarity index value of 0 as it leads to an infinite value in the denominator and a finite value in the numerator. This results in a discontinuity of the similarity index function.

**case 3: at least two divergent integrals** In this case,  $\beta_1 + \beta_2 \leq 1$ , resulting in both a diverging integral in Equation (2.10a) as in Equation (2.10b) or (2.10c). As the denominator in the similarity index will always hold a more dominant diverging term, the value will still be 0. However, this value of 0 results again in a discontinuity of the similarity index.

The practical implications of the discontinuities in cases 2 and 3 will be discussed further in the example introduced in Section 2.4.

## 2.3. The procedure

To be able to employ the statistical test as described in Section 2.1.2, a few steps have to be taken. Since these steps partly depend on estimation methods, the exact choices made by Ypma and Ross [6] will be used in this thesis. Below an overview of the procedure is given, based on the overview in [6]. Some steps might require extra information, these will be explained more thoroughly in the subsequent subsections.

Let the lifetimes  $x_1, x_2, \dots, x_n$  be a realization of an i.i.d. sample of size  $n$  following a Weibull distribution. Assume that the true parameters of this Weibull distribution are  $\theta = (\alpha, \beta)$ . Let  $\theta_0 = (\alpha_0, \beta_0)$  denote the parameters of the reference distribution. Let  $t_c$  denote the chosen censoring time. To form a critical region  $K$  as defined in Equation (2.5), the cumulative distribution function of the test statistic under the null hypothesis is needed. Since this distribution is not known, it has to be estimated. This can be done by simulation and is explained further in Subsection 2.3.2.

When  $K$  is computed, the only thing left to do is to calculate the value of the test statistic,  $S$ , based on the observation and check whether  $S \in K$  or not. For the computation of the similarity index in Equation (2.7) two probability density functions are needed. Let  $f(t)$  and  $g(t)$  denote those functions. The function  $f(t)$  represents the probability density function of the reference distribution, a Weibull distribution with parameters  $(\alpha_0, \beta_0)$ . The function  $g(t)$  is the probability density function based on the sample. Since the true parameters of this distribution are unknown, they need to be estimated for  $g(t)$ . The procedure for the estimation method used by Ypma and Ross [6] is explained in Subsection 2.3.1. After this estimation,  $S$  can be calculated and a conclusion of either rejecting or not rejecting the null hypothesis can be drawn.

### 2.3.1. Ordinary Least Squares estimation on the order statistics

The method used by Ypma and Ross is based on the method described in Montanari et al. [2] which is the IEEE standard for the statistical analysis of electrical insulation breakdown data, the type of data for which the similarity index would be used. This standard uses an ordinary least squared estimation on the order statistics of the data and can be used for type I censored data like the observation in the constructed test. The following paragraphs contain a complete description of this method, based on the instructions in Montanari et al. [2].

In order to apply the ordinary least squares model, the data is initially transformed using a method similar to the one used in constructing QQ-plots. The initial step involves arranging the data in ascending order of lifetimes, resulting in  $t_{(1)}, \dots, t_{(n)}$ . Equation (2.13) can then be employed to generate a sequence of values that can be regarded as an approximation for the cumulative distribution of the data. This approximation is validated due to the resemblance between the expected value of  $F(x_i)$  and  $F(i, n)$ .

$$F(i, n) \simeq \frac{i - 0.44}{n + 0.25} \quad (2.13)$$

In the case of the type I censored data that is used, the  $F(i, n)$  are simply only calculated for the lifetimes which are known at time  $t_c$ . The unknown lifetimes are omitted. Let the number of failure times lower than

or equal to  $t_c$  be denoted by  $r$ . Then there are  $n - r$  unknown failure times. The value  $F(i, n)$  can then be calculated for the first  $r$  lifetimes of the ordered observation.

It is thus expected that the values for  $F(i, n)$  are similar to the true distribution function values  $F(t_i)$ . By rewriting both with the following steps, an equation can be found which can be useful for the ordinary least squares method:

$$F(i, n) \approx 1 - \exp\left(-\left(\frac{t_i}{\alpha}\right)^\beta\right) \quad (2.14a)$$

$$\ln(1 - F(i, n)) \approx -\left(\frac{t_i}{\alpha}\right)^\beta \quad (2.14b)$$

$$\ln(-\ln(1 - F(i, n))) \approx \beta(\ln(t_i) - \ln(\alpha)) \quad (2.14c)$$

$$\ln(t_i) \approx \frac{1}{\beta} \ln(-\ln(1 - F(i, n))) + \ln(\alpha) \quad (2.14d)$$

By substituting Equation (2.14d) with

$$y_i = \ln(t_i) \quad (2.15a)$$

$$x_i = \ln(-\ln(1 - F(i, n))) \quad (2.15b)$$

the following equation arises with  $\epsilon_i$  representing the difference:

$$y_i = \frac{1}{\beta} x_i + \ln(\alpha) + \epsilon_i \quad (2.16)$$

Let  $X$  be an  $r \times 2$  matrix with  $X_{i1} = x_i$  in the first column and only ones in the second column. Let vector  $Y$  be a vector of length  $r$  with  $Y_i = \ln(t_i)$ . Then Equation (2.16) can be put together for all  $i \in 1, 2, \dots, r$  with the following equation:

$$Y = \begin{bmatrix} \frac{1}{\beta} \\ \ln(\alpha) \end{bmatrix} X + \epsilon \quad (2.17)$$

Finding a solution with the ordinary least squares method means essentially minimizing the residual sum of squares. The estimations that minimize this are given by the following set of equations:

$$\frac{1}{\hat{\beta}} = ((X^T X)^{-1} X^T Y)_1 \quad (2.18a)$$

$$\ln(\hat{\alpha}) = ((X^T X)^{-1} X^T Y)_2 \quad (2.18b)$$

From those equations  $\hat{\alpha}$  and  $\hat{\beta}$  can be derived. Since the general assumptions for the ordinary least squares method are not met,  $(\hat{\alpha}, \hat{\beta})$  will not necessarily be the estimations with the smallest variance and they won't be unbiased. However, they can still be used.

There is one condition for using this estimation process on the censored data: the number of observed failures before time  $t_c$  should be larger than 1. If  $r = 1$ , the  $2 \times 2$  matrix  $X^T X$  will have a determinant of 0 since the columns and rows are then linearly dependent. The resulting matrix will not be invertible which makes it impossible to execute the estimation procedure. It follows that for the estimation procedure to work, it is necessary to have  $r \geq 2$ .

### 2.3.2. Computing the empirical distribution

It was mentioned before that the cumulative distribution function of the test statistic under the null hypothesis has to be estimated. This can, as suggested in Ypma and Ross [6], be done by simulating the test statistic under the null hypothesis with a Monte Carlo simulation and then using the empirical distribution function of those simulated values as an estimation. The empirical distribution function is defined as follows:

$$\hat{F}_{SIM}(s) = \frac{1}{m} \sum_{i=1}^m \mathbf{1}_{\{S_i \leq s\}} \quad (2.19)$$

The indicator function  $\mathbf{1}_{\{S_i \leq s\}}$  has value 1 if and only if the  $i$ 'th simulated value of the similarity index,  $S_i$ , is smaller than or equal to  $s$ . If not, the value of the indicator function is 0. The sum of all simulated values is taken and then divided by the total amount of simulated values. When using a very large simulation size  $m$ , the empirical distribution function can be used as an estimation for  $\mathbb{P}(S \leq s)$ . In this equation,  $S$  represents a random variable with the distribution of the similarity index under the null hypothesis and  $s$  is a particular similarity index value.

The process for computing this empirical distribution function for the similarity index under the null hypothesis with  $\theta_0 = (\alpha_0, \beta_0)$  and an realization of a sample of size  $n$  censored at time  $t_c$  is written down below:

1. Generate  $m$  samples of failure times of size  $n$ , using the reference distribution  $Weibull(\alpha_0, \beta_0)$ .
2. Estimate for each sample a pair of parameters  $\hat{\alpha}$  and  $\hat{\beta}$  with the same method as provided in Subsection 2.3.1.
3. Calculate the value of the similarity index for each sample with the estimated parameters.
4. Compute the empirical distribution function  $\hat{F}_{SIM}(s)$  as in Equation (2.19).

In Ypma and Ross [6] it is suggested to use  $m = 16000$  so for the completeness of this research that number will be used from now on too.

To find the critical region  $K$  such that  $P_{\theta_0}(S \in K) \leq \gamma$ , the inverse of the empirical distribution function in Equation (2.19) can now be used. The solution is provided in the next equation:

$$c_\gamma = \hat{F}_{SIM}^{-1}(\gamma) \quad (2.20)$$

Now all the different parts of the statistical procedure are fully explained, which makes it possible to apply the procedure to an example.

## 2.4. Example

To find out more about certain practical limitations or problems that can arise when using the statistical procedure described in this chapter, an example will be shown. The data and chosen cases used for this example are the same as in Example II in Ypma and Ross [6]. The data can be found in Appendix A. The following is provided:

- The sample consist of  $n = 87$  data points.
- The reference parameters,  $\alpha_0$  and  $\beta_0$  are respectively equal to 475 and 1.2.
- The data is tested against the null hypothesis at  $t_c \in \{40, 92, 175, 4750\}$ .
- A significance level of  $\gamma = 0.1$  is used.

All the computations are done in R.

### 2.4.1. Method

For the computation of the empirical distribution the complete procedure as described in Subsection 2.3.2 is executed per value for  $t_c$ . The first step is creating 16000 samples of size  $n$  from a  $Weibull(475, 1.2)$ . Then the sample is censored on time  $t_c$ . The estimated parameters can then be calculated with the method in Equation (2.18). In some cases, the ordinary least squares method can not be used due to fewer than two observations that failed before time  $t_c$ . The procedure can not be executed when this is the case, so those samples must be omitted from the simulation. In Subsection 2.4.2 the number of samples that were discarded because of this reason will be stated and evaluated.

Now for those 16000 pairs of  $(\hat{\alpha}, \hat{\beta})$  per  $t_c$  the similarity index can be calculated. In case this was not possible, the sample was omitted and not replaced. First the combination of  $\beta_0$  and  $\hat{\beta}$  should be checked on the conditions in Subsection 2.2.2. In the case of a divergent integral in either the numerator or denominator, the value of the similarity index is 0. In all other cases, the value of the similarity index is calculated with a numerical approximation of the integrals. For this thesis, the standard R function `integrate` was used. The empirical distribution of the test statistic per  $t_c$  can then be calculated with the help of Equation (2.19). The empirical distribution function will be plotted, to obtain a similar figure as Figure 4 in Ypma and Ross [6].

Table II in the article will also be replicated with the values found through the simulation. That table displays two additional values which were not previously discussed: the p-value of the data, estimated through the empirical distribution, and the value of  $c_{1-\gamma}$  which can be found in a similar way as  $c_\gamma$  in Equation (2.20).

### 2.4.2. Results

In Figure 2.1 the empirical distribution of the similarity index under the null hypothesis,  $\theta_0 = (475, 1.2)$  is plotted for multiple values of  $t_c$ . It can be seen that the higher the value of  $t_c$ , which results in more observations from which the failure time is known, the higher the p-value of the observed lifetimes (the dotted lines).

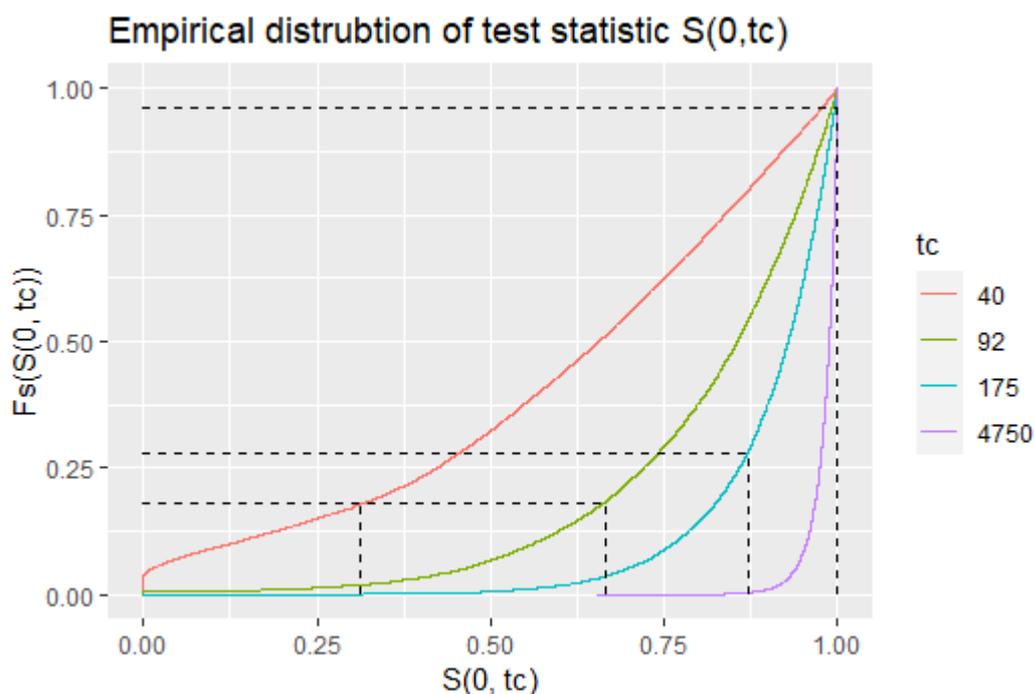


Figure 2.1: Empirical distribution of  $S(0, t_c)$

This figure should be similar to Figure 4 in Ypma and Ross [6] since the same data and procedures are used. However, there is a slight difference in the line representing the empirical distribution of the test statistic when  $t_c = 40$ . In Figure 2.1 this line makes a jump at 0, whereas in Figure 4 in the article, this jump is not seen.

$t_c$	$\hat{\alpha}$	$\hat{\beta}$	$S(0, t_c)$	$F_S(S)$	$S_{10\%}$	$S_{90\%}$
40	69.4	4.34	0.311	0.18	0.12	0.938
92	190.1	2.06	0.666	0.18	0.56	0.977
175	301.9	1.60	0.870	0.28	0.76	0.989
4750	470.8	1.23	0.999	0.96	0.95	0.998

Table 2.1: Multiple findings when executing the statistical test based on the similarity index on a data sample of size 87.

The differences can be seen even better when comparing Table 2.1 to Table II in Ypma and Ross [6]. The value

of  $F_s(S)$  at  $t_c = 40$ , which is in Table 2.1 equal to 0.12, is in the table in the article equal to 0.29. Apart from this difference between the two tables, which also leads to different p-values, the other values in the tables do correspond. There are two noteworthy distinctions between  $t_c = 40$  and the other  $t_c$  values which might have influenced those different values, both explained below.

**Divergent integrals:** At  $t_c = 40$ , there were 599 samples in the simulation for which  $\hat{\beta}$  turned out to be smaller than 0.5, resulting in a diverging integral in the denominator of the similarity index as explained in Subsection 2.2.2. In Ypma and Ross [6] it is suggested that an alternative domain for the integral could be selected in this case. This alternative domain would be of the form  $(\epsilon, t_c)$  with  $\epsilon$  small. This could be a reason for the different values at  $t_c = 40$  between Table 2.1 and the table in the article. However, it is decided for this thesis to not use the proposed workaround for the discontinuity. The reason for this decision is that choosing a domain of this form when an integral diverges, may introduce fluctuations between different cases, thereby making the evaluation and comparison of the similarity index less feasible.

The problem might be even more relevant when the reference distribution has a smaller  $\beta_0$  than 1.2. When for example the null hypothesis results in over 10% of the simulated samples having an estimated parameter value smaller than 0.5, the null hypothesis would be rejected only if  $\hat{\beta}$  of the data is also smaller than 0.5. The test would in such a case not be able to reject realizations from a Weibull distribution with a very high value for  $\beta$ . It would thus be interesting to see whether a variant of the similarity index which does not exhibit this behavior, could be an alternative.

**Insufficient amount of failures:** The other deviation between  $t_c = 40$  and the other censoring times is that there was a significant amount of samples for which the parameters  $\hat{\alpha}$  and  $\hat{\beta}$  could not be estimated. This was the case for 987 of the 16000 samples. This means that nearly  $\frac{1}{16}$  of the simulations had to be discarded because they could not be used. It is thus advisable to not use the similarity index when the probability of having to discard a sample due to having less than 2 failures before  $t_c$  is so high. In this case, a larger time interval could be considered to make a more proper judgment. At  $t_c = 92$  for example, there was only one sample that had to be omitted.

## 2.5. Conclusion

The statistical test using the similarity index, as proposed in Ypma and Ross [6], has now been fully described. When searching for alternatives to compare to the statistical test described in this chapter, it is necessary to look at tests that can deal with censored data and the assumption of a Weibull distribution. Additionally, it would be interesting to investigate whether alternative distance measures which do not have the same discontinuities as the similarity index can be found, as such discontinuities can yield undesired outcomes as seen in Section 2.4. These considerations are taken into account for the selection of alternative tests in the next chapter.

# 3

## Analysis for comparison

To thoroughly evaluate the use of the similarity index as part of a statistical test it is essential to look at two key components: a framework for quantifying the quality of a statistical test and suitable alternatives for the similarity index statistical test. The term 'suitable' might lack specificity. Therefore it is important to make some sensible choices when selecting alternatives. In this chapter, both aspects will be discussed.

In Section 3.1 and Section 3.2 two alternatives will be presented: the likelihood ratio test and a variant of the similarity index test. These are chosen based on their capabilities to decide on the same null hypothesis with censored data as presented in Section 2.1. Apart from those requirements, the motivation for these tests and a thorough explanation of their use will be given as well. Finally, in Section 3.3, the method of comparison utilized in this research, the power comparison method, will be introduced. The theoretical meaning of this method, as well as its computational limitations, will be discussed. This discussion will lead to recommendations for the simulation study in Chapter 4.

### 3.1. Likelihood ratio test

When considering alternative tests for the given case, there are multiple options to explore. Given that the similarity index relies on measuring the similarity or distance between probability density functions, it is natural to explore alternative test statistics within this category. For instance, methods based on the Total Variation or Hellinger distance, as mentioned in Cha [1] (which is also referenced by Ypma and Ross [6]), could be considered. However, it is also interesting to explore a test that fundamentally differs in its approach, not relying on distance measures between probability density functions. This led to the consideration of the likelihood ratio test, which is a widely used and applicable test. In Rinne [4], it is mentioned as a suitable test for various hypotheses regarding Weibull parameters, although it emphasizes the need for enough data to ensure reliability. Nonetheless, due to the test's widespread use and its applicability in censored Weibull data scenarios, it is chosen as one of the alternatives in this research. In this section, the approach of the likelihood ratio test will be explained.

#### 3.1.1. Test statistic

The test statistic for the Likelihood Ratio Test is given in Equation (3.1):

$$\lambda(X) = \frac{\sup_{\theta \in \Theta} p_{\theta}(X)}{\sup_{\theta_0 \in \Theta_0} p_{\theta_0}(X)} \quad (3.1)$$

In this research, the null hypothesis space is defined as  $\Theta_0 = \{\theta_0\}$ , with  $\theta_0 = (\alpha_0, \beta_0)$ . Consequently, the supremum in the denominator of Equation (3.1) can be omitted. The parameter space  $\Theta$  for this research is defined as  $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ .

The function  $p_\theta(X)$  is the likelihood function, which is defined to be the probability density function of the observation. Since it was assumed previously in Subsection 2.1.1 that lifetimes  $x_1, x_2, \dots, x_n$  are realizations of an independent, identically distributed sample from a Weibull distribution with parameter pair  $\theta = (\alpha, \beta)$ , this density function can be described as follows:

$$p_\theta(X) = \prod_{i=1}^n \frac{\beta}{\alpha} \left(\frac{x_i}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x_i}{\alpha}\right)^\beta\right) \quad (3.2)$$

However, since time censoring is used, a part of this equation needs to be written down differently. In Rinne [4], it is suggested for observations that are singly censored on the right at  $t_c$  to include those observations as the probability that the time-to-failure is higher than  $t_c$ :

$$1 - F(t_c) = \exp\left(-\left(\frac{t_c}{\alpha}\right)^\beta\right) \quad (3.3)$$

This results in the likelihood function shown in Equation (3.4), where the number of failed observations within time interval  $t_c$  is denoted by  $r$ .

$$p_\theta(X) = \prod_{i=1}^r \frac{\beta}{\alpha} \left(\frac{x_i}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x_i}{\alpha}\right)^\beta\right) \prod_{j=r+1}^n \exp\left(-\left(\frac{t_c}{\alpha}\right)^\beta\right) \quad (3.4)$$

Equation (3.4) provides a means to compute the likelihood statistic. However, to calculate the numerator of Equation (3.1), the parameter pair  $(\alpha, \beta)$  must be chosen from  $\Theta$  in such a way that the likelihood equation given in (3.4) is maximized. This process aligns with a maximum likelihood estimation for  $\alpha$  and  $\beta$ . The methodology for conducting this estimation will be outlined in the next subsection.

### 3.1.2. Maximum likelihood estimation

As usual, the maximum likelihood estimators can be found by finding the optimum of the log-likelihood function. This function, derived by taking the log of the likelihood function in Equation (3.4), is defined as follows:

$$\mathcal{L}(\alpha, \beta) = r[\ln(\beta) - \beta \ln(\alpha)] + (\beta - 1) \sum_{i=1}^r [\ln(x_i)] - \alpha^{-\beta} \left[ \sum_{i=1}^r x_i^\beta + (n-r)t_c^\beta \right] \quad (3.5)$$

The approach used for finding the optimum follows the methodology outlined in Chapter 11.6.1.1 of Rinne [4]. It is worth noting that the equations in the cited reference are slightly different as they are based on type II censoring. The distinction is that the value of  $x_r$  in the book is replaced by  $t_c$  in this research, in correspondence with type I censoring. However, this alteration does not introduce any changes in the derivation and approximations methods, as it merely involves a constant factor change.

Equation (3.5) can be partially differentiated. This results in two separate equations, which give the maximum likelihood estimators when equalized to 0. Those derivations are omitted here but can be found in Rinne [4]. Substitution leads to Equation (3.6) for which the solution  $\hat{\beta}$  can be found. With this first estimation, the solution can be found for  $\hat{\alpha}$  as seen in Equation (3.7).

$$0 = \frac{1}{\beta} + \frac{1}{r} \sum_{i=1}^r \ln x_i - \frac{\sum_{i=1}^r x_i^\beta \ln x_i + (n-r)t_c^\beta \ln t_c}{\sum_{i=1}^r x_i^\beta + (n-r)t_c^\beta} \quad (3.6)$$

$$\hat{\alpha} = \left[ \frac{\sum_{i=1}^r x_i^{\hat{\beta}} + (n-r)t_c^{\hat{\beta}}}{r} \right]^{\frac{1}{\hat{\beta}}} \quad (3.7)$$

The solution  $\hat{\beta}$  for Equation (3.6) can not be found analytically, but is easily done with a Newton-Rhapson method. There is again the requirement that  $r \geq 2$ . There are multiple variants available, one can be found in Rinne [4] in Chapter 11.6.1.1.

### 3.1.3. Overview of the test

To summarize all the steps that need to be taken when using the likelihood ratio test as an alternative, an overview is given in this subsection.

The first step is to prepare the data like it was done when using the similarity index. This includes ordering the lifetimes that are smaller than  $t_c$  as  $x_{(1)}, \dots, x_{(r)}$  and writing down the number of lifetimes higher than  $t_c$  by  $n - r$ . With those values, estimations  $\hat{\alpha}$  and  $\hat{\beta}$  can be calculated with the use of Equations (3.7) and (3.6) and a Newton-Rhapson approach. By filling in those estimates in Equation (3.1), using the function in Equation (3.4), the test statistic value can then be calculated.

The subsequent step involves determining the distribution of the test statistic under the null hypothesis. While the distribution of the likelihood equation is known when an adequate number of data points are available, this is not always the case in the proposed applications of the similarity index in Ypma and Ross [6]. Consequently, when using this test as an alternative to the similarity index test, the cumulative distribution function has to be estimated. The procedure for this estimation remains unchanged from what was described in Subsection 2.3.2, with the exception that the similarity index is replaced by the likelihood ratio statistic.

There is however one crucial difference in calculating the critical region for the likelihood ratio statistic. This region is defined as

$$K = \{L \geq d_{c\gamma}\} \quad (3.8)$$

where  $L$  represents the likelihood ratio statistic and  $K$  is determined such that the following holds:

$$P_{\theta_0}(L \in K) \leq \gamma \quad (3.9)$$

The reason for this different representation of the critical region is that the range of the likelihood ratio statistic is  $[1, \infty]$  and its value is going to be larger whenever the maximum likelihood estimates deviate further from the reference parameters. It is thus wanted to reject the null hypothesis whenever  $L$  is large, resulting in the given definition for  $K$ .

## 3.2. Variant of the similarity index

Opposed to a more established method like the likelihood ratio test, it is also interesting to see if a more similar method to the one presented in Chapter 2, but then without discontinuities, will lead to the same or even better quality. The natural variant that arises is presented in Equation (3.10), where  $f_1(t)$  and  $f_2(t)$  are the same as described in Equation (2.7):

$$S_{var}(0, t_c) = \frac{\int_0^{t_c} \sqrt{f_1(t)f_2(t)} dt}{\int_0^{t_c} f_1(t) + f_2(t) - \sqrt{f_1(t)f_2(t)} dt} \quad (3.10)$$

The reasoning in the first part of Subsection 2.2.2 still applies to this equation, which means that the values will again range from 0 to 1 with the latter if and only if  $f_1(t) = f_2(t)$  for  $t \in (0, t_c)$ . However, unlike the similarity index, where diverging integrals resulted in a discontinuity, this alternative does not exhibit such issues. This can be seen by rewriting Equation (2.10a) up to Equation (2.10c) for this variant of the similarity index. The results are:

$$\int_0^{t_c} \sqrt{f_1(t)f_2(t)} dt = \sqrt{\frac{\beta_1\beta_2}{\alpha_1^{\beta_1}\alpha_2^{\beta_2}}} \int_0^{t_c} t^{\frac{\beta_1+\beta_2}{2}-1} \exp\left(-\frac{1}{2}\left(\frac{t}{\alpha_1}\right)^{\beta_1} - \frac{1}{2}\left(\frac{t}{\alpha_2}\right)^{\beta_2}\right) dt \quad (3.11a)$$

$$\int_0^{t_c} f_1(t) dt = \frac{\beta_1}{\alpha_1^{\beta_1}} \int_0^{t_c} t^{\beta_1-1} \exp\left(-\left(\frac{t}{\alpha_1}\right)^{\beta_1}\right) dt \quad (3.11b)$$

$$\int_0^{t_c} f_2(t) dt = \frac{\beta_2}{\alpha_2^{\beta_2}} \int_0^{t_c} t^{\beta_2-1} \exp\left(-\left(\frac{t}{\alpha_2}\right)^{\beta_2}\right) dt \quad (3.11c)$$

Following the same reasoning as described in Subsection 2.2.2, it can now be easily seen that there are no diverging integrals since the power of  $t$  in those integrals will never be smaller than or equal to  $-1$ . Since all integrals are now convergent, the function has no discontinuities. As a result, it offers a more robust and consistent measure that can be utilized effectively in various scenarios.

As this modification introduces only minor adjustments to the calculation of the test statistic in the original similarity index test, the entire procedure outlined in Section 2.3 can be applied and replicated for this test. The only necessary alteration lies in the computation of the similarity index itself.

### 3.3. Power comparison

Having selected two appropriate alternatives, the next step is determining the most appropriate method for comparing their performance. When two statistical tests are conducted at the same significance level, it becomes inadequate to compare them based on a situation corresponding to the null hypothesis. However, it is a possibility to look at the performance of the test when the true parameters of the sample do not correspond to the null hypothesis. By examining the power function under the alternative parameter set  $\Theta_1 = \Theta \setminus \{\theta_0\}$ , it becomes possible to indicate the test with the highest power. Therefore, in this thesis, the comparison of the three selected tests will be based on their power functions, allowing for a meaningful and informative evaluation of their respective capabilities.

#### 3.3.1. Power function

The power function for a statistical test with critical region  $K$  for test statistic  $T$  can be found in Equation (3.12) and corresponds to the probability that a realization  $x$  of a sample with distribution parameters  $\theta$  is rejected.

$$\pi(\theta; K) = \mathbb{P}(T(x) \in K) \quad (3.12)$$

An ideal scenario would involve obtaining an analytical solution for the power function across all parameter values  $\theta \neq (\alpha_0, \beta_0)$  and comparing these values among different tests. However, for the tests chosen in this thesis, the exact distribution of the test statistic remains unknown even when the true data parameters  $\theta$  are provided. Consequently, a simulation-based approach becomes necessary, similar to the methodology employed in estimating the empirical distribution of the test statistic under the null hypothesis.

A simulation-based methodology can be implemented by generating  $k$  samples of a Weibull distribution with parameters set to the alternative pair  $\theta_1 = (\alpha_1, \beta_1)$  where  $\theta_1 \neq \theta_0$ . By conducting the test on each sample and calculating the ratio of rejected samples to the total number of samples  $m$ , an estimate of the power can be obtained. Subsequently, the power estimate can be compared to that of other tests, employing the same simulation-based approach for all tests. Through this comparative analysis, the most powerful test for when the true parameters of the observation equal  $\theta_1$  can be identified within the selected set of tests.

#### 3.3.2. Limitations

One potential limitation of a power comparison is the computational complexity. For a complete comparison between the tests, the power function would need to be calculated for every pair of reference parameters  $(\alpha_0, \beta_0)$ . The complete power function consists of the power for every distinct pair of alternative parameters  $(\alpha_1, \beta_1)$ . Moreover, considering that the tests may also exhibit varying performance under different censoring times  $t_c$ , incorporating this additional parameter into the power function would result in an even greater amount of simulation work. This would be impracticable. Thus, to obtain an overview of the power differences between the tests, a selection process must be employed to create a balance in computational work and the additional information it provides. It is advisable to select a few different alternative parameter pairs and a small set of censoring times and repeat this setup for multiple null hypotheses. This could still give an insight into the outline of the power function but can be computed in a reasonable amount of time.

Another potential limitation is that the comparison of power may not always yield a definitive conclusion. One test may perform better in certain regions of the alternative parameter domain, while another test performs better in different regions. It can thus be helpful to identify which parts of the domain are particularly

relevant in this research. By identifying and emphasizing the most significant regions, a more nuanced understanding of the performance of the similarity index test compared to the alternatives can be achieved.

### **3.4. Conclusion**

In conclusion, this chapter focused on how to create a comparison between the statistical test based on the similarity index and other alternative tests. The chosen alternatives are the well-known likelihood ratio test and a variant of the similarity index as proposed in Equation (3.10). Both can deal with the censored Weibull data for which the similarity index is initially proposed.

Additionally, this chapter introduced the concept of computing the power of the tests to compare their performance. Although a great way to compare the different tests, there should be a careful selection of the parameters for which the power is calculated to avoid too much computational effort. In Chapter 4 a relevant selection of parameters will be presented, along with the outcome of the simulation process based on this selection.



# 4

## Simulation

Taking into account the content of the last two chapters, it is now finally possible to develop a simulation that can provide more insights into the effectiveness of the similarity index as a test statistic through a power comparison. Given that Chapter 3 mentioned the time-consuming nature of computing the entire power function, this chapter will focus on both carefully selecting the best scenarios for simulation and conducting the actual simulations.

In Section 4.1 six base cases of fixed parameters are established which belong to two distinct null hypotheses. The reasoning behind the chosen parameters will be shown. For those six base cases, an estimation of the power function will be made. Since it would be too time-consuming to estimate the complete power function, it will be estimated along a few 1-dimensional directions in the 2-dimensional parameter space  $\Theta$ . Those directions will be presented along with a method to choose a few points along those directions consistently for the estimation. Then the practical relevance of these directions concerning the application of the similarity index will be thoroughly explained, establishing a clear focus for evaluating the results. The results will be presented in Section 4.3 through separate graphs. Some interesting cases will be highlighted, leading to a conclusion on the quality of the similarity index in terms of power.

### 4.1. Base cases

To facilitate a meaningful power comparison, a set of base cases is chosen for which the power function will be estimated. The selection of parameters for those base cases is based on two criteria: alignment with the scenarios for which the similarity index is intended to be utilized, as described in Ypma and Ross [6], and practical considerations for conducting the power comparison. Below is a concise summary of all the parameters for a base case:

1. Parameter pair  $\theta_0$  for the null hypothesis
2. A censoring time  $t_c$
3. The sample size  $n$
4. The significance level  $\gamma$
5. The size of the Monte Carlo simulation  $m$

The choice for the parameter pairs for the null hypotheses, number 1, will be thoroughly explained in Subsection 4.1.1, with a focus on the practical context in which the similarity index was introduced in Ypma and Ross [6]. The values of the other parameters, which will mostly remain constant throughout the base cases, will be determined in Subsection 4.1.2.

### 4.1.1. Relevant null hypotheses

The selection of parameters  $\alpha_0$  and  $\beta_0$  for the reference distribution in the base cases should be in alignment with real-life scenarios for which the test based on the similarity index could be applicable. It is chosen to consider values that reflect the characteristics of the intended applications described in Ypma and Ross [6]. There, the similarity index was proposed to be used "to quantify the similarity between observed and reference breakdown behavior". For the comparison, it is thus interesting to calculate the power with reference distributions that are likely to be found in this field of breakdown behavior.

To get more insight into the parameters of the Weibull distribution before making the actual choices, a small analysis is given below. Recall that the density function of the Weibull distribution is given by:

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^\beta\right) \quad (4.1)$$

The parameter  $\alpha$ , also known as the scale parameter, is given in the same unit of time as the observation (e.g., days or years, depending on the component and field of application). In Subsection 2.2.2. of Montanari et al. [2] it is shown that every Weibull distribution is a linear transformation of the reduced form with  $\alpha = 1$ . An explanation for this is presented here. Assume that  $W_1 = \text{Weibull}(1, \beta)$  and let  $W_2$  be defined as follows:

$$W_2 = \alpha W_1 \quad (4.2)$$

The following can be said about the cumulative distribution function of  $W_2$ :

$$F_2(t) = \mathbb{P}(\alpha W_1 \leq t) = \mathbb{P}\left(W_1 \leq \frac{t}{\alpha}\right) = F_1\left(\frac{t}{\alpha}\right) \quad (4.3)$$

By differentiating the cumulative distribution function of  $W_1$  with input  $\frac{t}{\alpha}$ , it leads to the following probability density function of  $W_2$ :

$$f_2(t) = \frac{1}{\alpha} f_1\left(\frac{t}{\alpha}\right) \quad (4.4)$$

$$= \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^\beta\right) \quad (4.5)$$

This results in a probability density function related to a  $\text{Weibull}(\alpha, \beta, .)$  The conclusion is that it is sufficient to use  $\alpha = 1$  for the base cases in the power comparison, as it can be linearly transformed into any Weibull distribution with the same  $\beta$  through the equations above.

The parameter  $\beta$ , also known as the shape parameter, gives rise to different-looking density functions depending on its value. In Subsection 2.2.3. of Montanari et al. [2], six cases are distinguished, and three of them are highlighted below. For this analysis, it is assumed that  $\alpha = 1$ .

1.  $\beta < 1$

In this case, the density function of the Weibull distribution has the following limits:

$$\lim_{t \rightarrow \infty} f(t) = 0$$

$$\lim_{t \rightarrow 0} f(t) = \infty$$

According to Ross [5], this is unlikely to be observed in the data for which the similarity index procedure is used. This is due to the quality control often carried out before products are put into use and the assumption that most early failures (childhood mortality) are due to a rather fast-aging sub-population (high  $\alpha$ ) instead of a low  $\beta$ . Thus, this case will not be used as a base case in the power comparison.

2.  $\beta = 1$

This resembles the density of an exponential distribution, which is known for random failure. For most components, it is assumed that the probability of failure increases at a higher rate while aging, implying that  $\beta > 1$ . However, using  $\beta = 1$  essentially shows whether the similarity index is powerful when testing whether the assumption of random failure over a Weibull distribution can be rejected.

3.  $\beta > 1$ 

The Weibull density function with  $\beta > 1$  does not contain any infinite function values and reaches a maximum within the domain. The examples where the similarity index was used in both [6] and [5], all satisfy this condition.

From those three different scenarios, it is chosen to use both  $\beta_0 = 1$  and  $\beta_0 = 3$  for the base cases. So two parameter pairs will be used for the null hypotheses in the power comparison:  $\theta_0 = (1, 1)$  and  $\theta_0 = (1, 3)$

#### 4.1.2. Additional parameters

There are four remaining parameters,  $t_c$ ,  $n$ ,  $\gamma$ , and  $m$ . All of them might influence the power function, but some are more interesting to vary than others. This will be explained in the following paragraphs.

For the significance level  $\gamma$  and the number of simulations for the empirical distribution  $m$ , the same values as in Ypma and Ross [6] will be used to be able to evaluate the exact procedure in the article. Therefore, the following values will be used in the power comparison:

$$\gamma_0 = 0.1 \tag{4.6}$$

$$m = 16000 \tag{4.7}$$

For  $n$  and  $t_c$ , the chosen values will have a significant impact on the number of failures that are not censored at  $t_c$ . Choosing a high value for  $n$  might lead to a higher power in the test because then estimations will be more precise. However, the goal is to create a base case that is realistic to a real-life scenario. In Ypma and Ross [6], an example with  $n = 87$  was used. For the power comparison, a value in the same order of magnitude but more convenient for calculations will be chosen:

$$n = 100 \tag{4.8}$$

With all the other parameters fixed, it is chosen to vary  $t_c$ , allowing for an assessment of the performance of the similarity index throughout the whole failure process. The goal is to select  $t_c$  in such a way that there are enough failures before  $t_c$  to be able to execute the procedure, but without having a majority of the failures before  $t_c$ . To ensure an easy way to find such a  $t_c$  value per null hypothesis, a method based on the expected number of failures will be used to determine  $t_c$ . Let  $Y$  be the random variable defined in Equation (4.9), based on a random vector  $X$ , which represents an independent, identically distributed sample of Weibull distributed random variables.

$$Y = \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t_c\}} \tag{4.9}$$

Then,  $Y \sim \text{Bin}(n, p)$ , with  $p$  equal to  $\mathbb{P}(X \leq t_c)$ , which can be expressed as:

$$p = 1 - \exp\left(-\left(\frac{t_c}{\alpha}\right)^\beta\right) \tag{4.10}$$

It is possible to choose  $t_c$  such that different null hypotheses result in the same value of  $p$ , thereby ensuring the same expected value for the total number of failures,  $\mathbb{E}[Y]$ . The expected value can be calculated using the binomial distribution and is shown in Equation (4.11):

$$\mathbb{E}[Y] = np \tag{4.11}$$

Three possible values of  $p$  that would lead to enough failures to estimate the parameters, while not having a majority of the sample failed, are  $p \in (0.2, 0.3, 0.4)$ .

With two different null hypotheses and three different censoring times per null hypothesis, six base cases can be formed. The overview of those base cases can be seen in Table 4.1.

$\alpha_0$	$\beta_0$	$t_c$	$n$	$m$	$\gamma$
1	1	0.223	100	16000	0.1
1	1	0.357	100	16000	0.1
1	1	0.511	100	16000	0.1
1	3	0.607	100	16000	0.1
1	3	0.709	100	16000	0.1
1	3	0.799	100	16000	0.1

Table 4.1: The parameters of the six bases cases for the power comparison

## 4.2. Alternative Weibull parameters

With the parameters for the base cases determined, a plan can be formulated to estimate the power function defined in Equation (3.12) for those base cases. The power function represents the probability of rejecting the null hypothesis when the true distribution parameters  $\theta = (\alpha, \beta)$  differ from the chosen base case parameters  $\theta_0$ . When  $\theta = \theta_0$ , the power function should be equal to the selected significance level  $\gamma$ , while as  $\theta$  deviates further from  $\theta_0$ , the power function is expected to approach 1. The goal is to estimate the power function of the different tests efficiently and without having to estimate it for the whole parameter space.

Some parts of the parameter space are of greater interest when using the similarity index, since the actual motivation for the similarity index is to test non-compliance for samples which perform worse than the reference distribution. Those parts will be indicated in Subsection 4.2.2.

### 4.2.1. Directions

To make an efficient estimation, the power will be calculated along 4 directions through  $\theta_0$ . Those are the directions obtained by only varying  $\alpha$ , only varying  $\beta$ , varying both in the direction with slope 1 and varying both in the direction with slope  $-1$ .

A total of 20 combinations of alternative parameters pairs  $(\alpha_1, \beta_1)$  per direction will be used to create an outline of the power function. The spread of those 20 combinations will be chosen based on different methods per direction, which will be explained in the following paragraphs. The goal is to choose the spread in such a way that the parts with the largest variation in the power function can be shown and to create consistency between different null hypotheses. This is done by first considering the relation between two different Weibull distributions. For this it can be helpful to rewrite the cumulative distribution function and find its inverse, similar to what was done in Equation (2.14) in order to perform the OLS estimation. The following Equation resembles Equation (2.14d), where the variables are substituted with their corresponding random variables  $W \sim Weibull(\alpha, \beta)$  and  $U \sim U(0, 1)$ . This results in the following:

$$\ln(W) = \frac{1}{\beta} \ln(-\ln(1-U)) + \ln(\alpha) \quad (4.12)$$

$$W = \alpha \exp\left(\frac{\ln(-\ln(1-U))}{\beta}\right) \quad (4.13)$$

Equation (4.13) can be employed to compare two Weibull distributions and assess the magnitude of their differences. This approach provides an indication of the parameter scale for the selected directions of change, enabling the method to be applied to multiple null hypotheses. Suppose we have  $W_0 \sim Weibull(\alpha_0, \beta_0)$  and  $W_1 \sim Weibull(\alpha_1, \beta_1)$ . Then the following holds, where the substitution  $Z = \ln(-\ln(1-U))$  is used:

$$\frac{W_1}{W_0} = \frac{\alpha_1}{\alpha_0} \exp\left(\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right)Z\right) \quad (4.14)$$

Based on this insight, a set of four different equations can be made which represent the four direction lines which will be used for the alternative parameters. Suppose  $(\alpha_0, \beta_0)$  is known. For Equation (4.15a) the values for  $(\alpha_1, \beta_1)$  are on the line  $\beta_1 = \beta_0$ , for Equation (4.15b) on the line  $\alpha_1 = \alpha_0$ , for Equation (4.15c) on the line

$\beta_1 = \alpha_1 - \alpha_0 + \beta_0$  and for Equation (4.15d) on the line  $\beta_1 = \alpha_0 - \alpha_1 + \beta_0$ .

$$\frac{W_1}{W_0} = \frac{\alpha_1}{\alpha_0} \quad (4.15a)$$

$$\frac{W_1}{W_0} = \exp\left(\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right)Z\right) \quad (4.15b)$$

$$\frac{W_1}{W_0} = \frac{\alpha_1}{\alpha_0} \exp\left(\left(\frac{1}{\beta_0 + \alpha_1 - \alpha_0} - \frac{1}{\beta_0}\right)Z\right) \quad (4.15c)$$

$$\frac{W_1}{W_0} = \frac{\alpha_1}{\alpha_0} \exp\left(\left(\frac{1}{\beta_0 - \alpha_1 + \alpha_0} - \frac{1}{\beta_0}\right)Z\right) \quad (4.15d)$$

A possible strategy is to set the non-random part of the right hand side of those equations equal to  $1 + \delta i$  with  $\delta > 0$  and  $i \in \{-10, -9, \dots, -1, 1, \dots, 9, 10\}$ . In this way  $\delta$  can be chosen per direction and per base case, by observation, in such a way that the power function can be simulated in both the positive and negative direction of each line up to the point where it is close to 1 for at least one of the tests. This way the power values exhibits significant variability along the interval, allowing for a clear depiction of the function along that direction. The value of  $\delta$  can for this reason also be chosen differently for when  $i$  is negative or positive. This leads to the following equations for which a solution should be found when  $i \in \{-10, \dots, -1\}$  and  $i \in \{1, \dots, 10\}$ :

$$\frac{\alpha_1}{\alpha_0} = 1 + \delta i \quad (4.16a)$$

$$\exp\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right) = 1 + \delta i \quad (4.16b)$$

$$\frac{\alpha_1}{\alpha_0} \exp\left(\frac{1}{\beta_0 + \alpha_1 - \alpha_0} - \frac{1}{\beta_0}\right) = 1 + \delta i \quad (4.16c)$$

$$\frac{\alpha_1}{\alpha_0} \exp\left(\frac{1}{\beta_0 - \alpha_1 + \alpha_0} - \frac{1}{\beta_0}\right) = 1 + \delta i \quad (4.16d)$$

There are a few limitations that are experienced when using those equations. A solution might not always be found, since some functions have discontinuities and there are the restrictions that both  $\alpha_1 > 0$  and  $\beta_1 > 0$ . Those constraints are summarized per Equation:

**Equation (4.16a):** Since  $\alpha_0 = 1$  in all the null hypotheses in the base cases, the range of the left-hand side function is  $(0, \infty)$ .

**Equation (4.16b):** This equation has the following constraint on the range, which depends on the value  $\beta_0$ :

$$\frac{1}{\exp\left(\frac{1}{\beta_0}\right)} < \exp\left(\frac{1}{\beta_1} - \frac{1}{\beta_0}\right) < \exp\left(1 - \frac{1}{\beta_0}\right) \quad (4.17)$$

This indicates that there is a limit to the magnitude by which one Weibull parameter can be larger than the other solely due to a change in the shape parameter. The extent of this limit depends on  $\beta_0$ , and thus, the interval for alternative  $\beta_1$  should be determined accordingly based on the value of  $\beta_0$ .

**Equation (4.16c):** In the chosen base cases where  $\alpha_0 \geq \beta_0$ , the function is continuous and  $\beta_1 = \beta_0 + \alpha_1 - \alpha_0 > 0$  if  $\alpha_1 > 0$ .

By observing the function and taking the limits as  $\alpha_1$  approaches 0 and  $\infty$ , it can be observed that for  $\beta_0 = 1$  the range is  $(1, \infty)$ , indicating that a solution will only be possible for positive  $i$ . The function attains its minimum in  $\alpha_1 = 1$ . To find values both in the negative and positive direction on this line, the solutions on  $(0, 1)$  should be separated from those on  $(1, \infty)$ .

For  $\beta_1 = 3$ , the range is  $(0, \infty)$ , so no problems arise and both negative and positive values of  $i$  can be used.

**Equation (4.16d):** This function has a discontinuity at  $\alpha_1 = \alpha_0 + \beta_0$ . Fortunately, since the restriction  $\beta_0 - \alpha_1 + \alpha_0 > 0$  is necessary to have  $\beta_1 > 0$ , it suffices to consider the domain  $(0, \alpha_0 + \beta_0)$  for  $\alpha_1$ . Within this domain, the function's range is  $(0, \infty)$ .

With this in mind, 20 alternative parameters  $(\alpha_1, \beta_1)$  were chosen for all 4 directions. The exact values can be found in Appendix B.

### 4.2.2. Compliance

As described before, the goal of the similarity index as described by Ypma and Ross [6] was to create some quantification to check whether an observation is coming from the reference distribution or not. It is thus the question whether a batch of components is compliant with the reference distribution provided by the creator of the batch. In the search for compliance, the focus is primarily on earlier-than-expected failures, as those might indicate that the batch has a shorter average lifetime which is unwanted. Conversely, a late first failure might be perceived as a positive outcome rather than a non-compliant one.

Hence, prior to analyzing the power comparison results, it would be advantageous to examine parameter combinations that hold particular significance. In this regard, it is worth considering the expected value of a Weibull distribution, as expressed by the following equation:

$$\mathbb{E}[W] = \alpha \Gamma \left( 1 + \frac{1}{\beta} \right) \quad (4.18)$$

A smaller value of  $\alpha$  indicates a lower expected value, implying an earlier expected occurrence of failure. This aspect is of particular relevance in non-compliance detection and those alternative parameters in the power comparison that represent a smaller  $\alpha$  than the reference distribution should be examined more closely.

Conversely, the influence of  $\beta$  on the expected lifetime is more complex. The gamma function exhibits a minimum at  $x_{\min} \approx 1.46163$ , the specific number can be found in Olver et al. [3]. Prior to this minimum, the gamma function is a decreasing function, while it becomes an increasing function afterwards. In terms of  $\beta$ , this implies that when  $\frac{1}{\beta} \approx 0.46163$ , there exists a transition point where both higher and lower values of  $\beta$  lead to higher expected values. This transition occurs at approximately  $\beta \approx 2.16624$ .

When  $\beta$  belongs to the interval  $(0, \frac{1}{x_{\min}-1})$ , the expected value in Equation (4.18) decreases as  $\beta$  increases. However, for  $\beta$  within the interval  $(\frac{1}{x_{\min}-1}, \infty)$ , the expected value increases as  $\beta$  increases and the expected value will approach  $\alpha$  in its limit. Reasoning solely from this perspective might suggest that a very low  $\beta$  is most preferable. Nevertheless, such low values of  $\beta$  result in a high number of early failures. This can not be seen in the expected life time since the high probability of an early failure is compensated with the low probability of a very late failure. This renders it difficult to make a definitive determination regarding the most interesting direction for  $\beta$  in terms of non-compliance. However, when  $\beta_0 = 1$  in the base cases, the directions where  $\beta_1 < \beta_0$  are the most interesting, since a true parameter of  $\beta < 1$  would be unwanted as discussed in Subsection 4.1.1.

## 4.3. Results

In the preceding section, the methods for determining alternative parameter combinations along the selected directions were discussed. By employing these selected alternative parameters, which can be found in Appendix B, it becomes feasible to calculate the value of the power function for these base cases along those directions.

The method used for the calculation of the power is the one described in Subsection 3.3.1. The power function will be calculated for the six base cases presented in Table 4.1, as explained in Subsection 4.1.1. The tests for which the power will be calculated are the similarity index based test described in Chapter 2, the likelihood ratio test described in Section 3.1 and the variant of the similarity index based test described in Section 3.2. The number of samples used to estimate the power, previously named  $k$ , will be equal to 1000. This number is chosen by considering both computation time and precision.

The results are grouped by the directions described in the previous section. The powers of all three tests are combined in plots which are divided per base case. The jitter feature in R is used to make the points visible when they are overlapping. An explanation and evaluation will be given per figure.

### 4.3.1. Univariate variation: $\alpha$

In Figure 4.1 and Figure 4.2, the power results are depicted for the six base cases when varying only the alternative parameter  $\alpha_1$ . The likelihood ratio test exhibits consistently higher power across all points, particularly

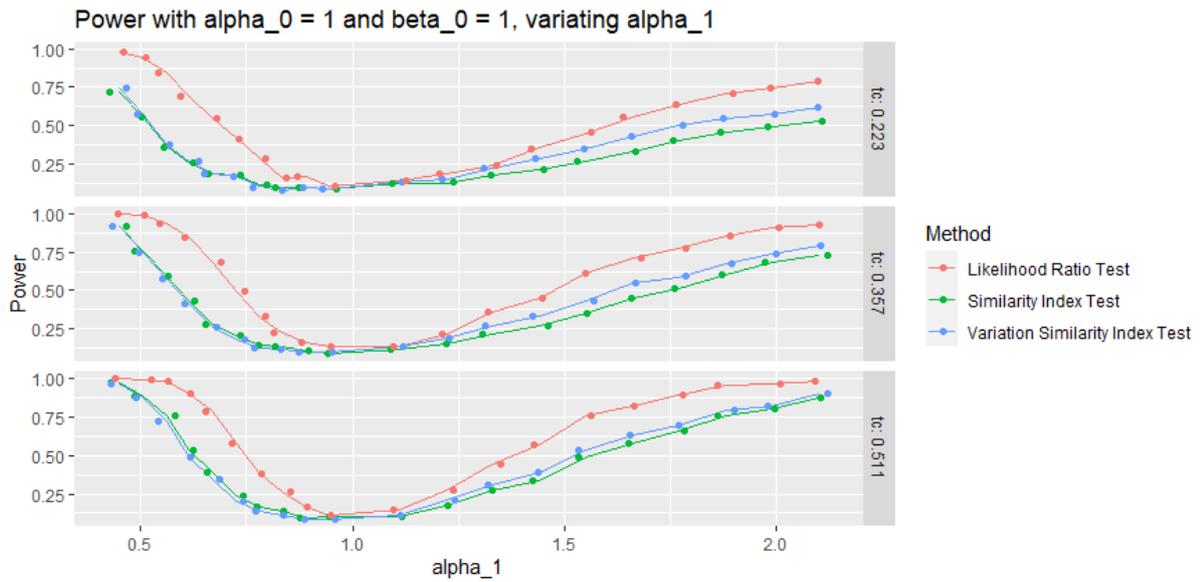


Figure 4.1: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 1$ ,  $\beta_1 = 1$  and  $\alpha_1$  varied

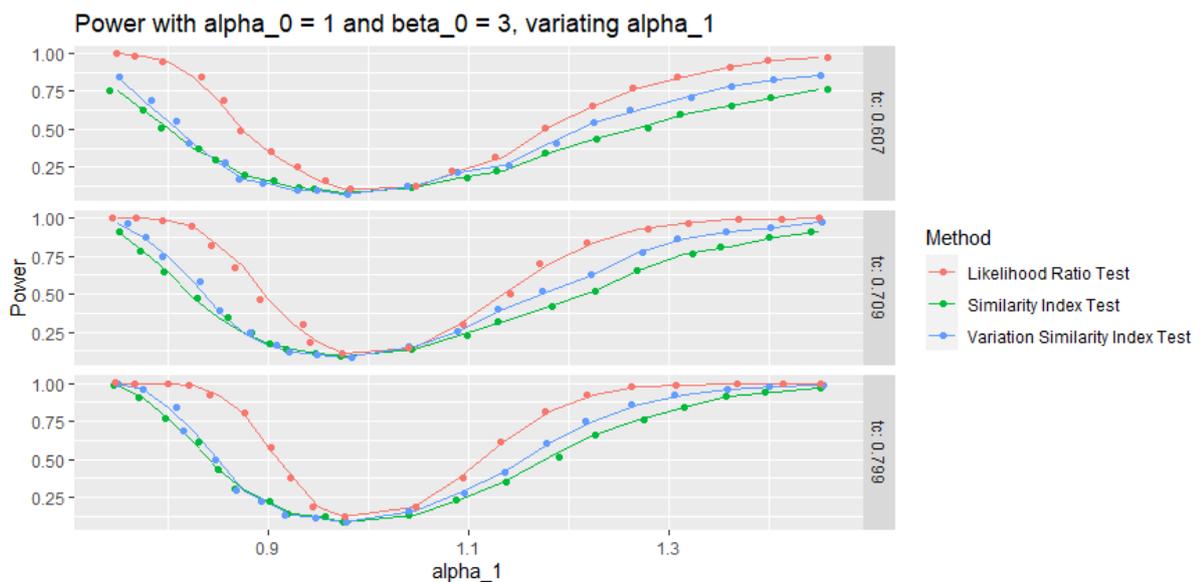


Figure 4.2: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 3$ ,  $\beta_1 = 3$  and  $\alpha_1$  varied

when  $\alpha_1 < 1$ . As discussed in Subsection 4.2.2, this region is of particular interest due to the negative impact on the expected lifetime when  $\alpha$  is lower. The similarity index test and its variant generally demonstrate similar power values across all base cases, with the variant exhibiting slightly higher power at some points.

### 4.3.2. Univariate variation: $\beta$

The results for varying  $\beta_1$  while keeping  $\alpha_1$  equal to  $\alpha_0$  are presented in Figure 4.3 and Figure 4.4. There is a noticeable distinction in power between these figures. In Figure 4.4, where  $\beta_1$  is varied around  $\beta_0 = 3$ , both the likelihood ratio test and the variant on the similarity index test consistently exhibit higher power across all simulated points. The likelihood ratio test has again the highest power values. When  $\beta_1 < \beta_0$ , the likelihood ratio test continues to exhibit the highest power, especially for the lower values of  $t_c$ . Nevertheless,

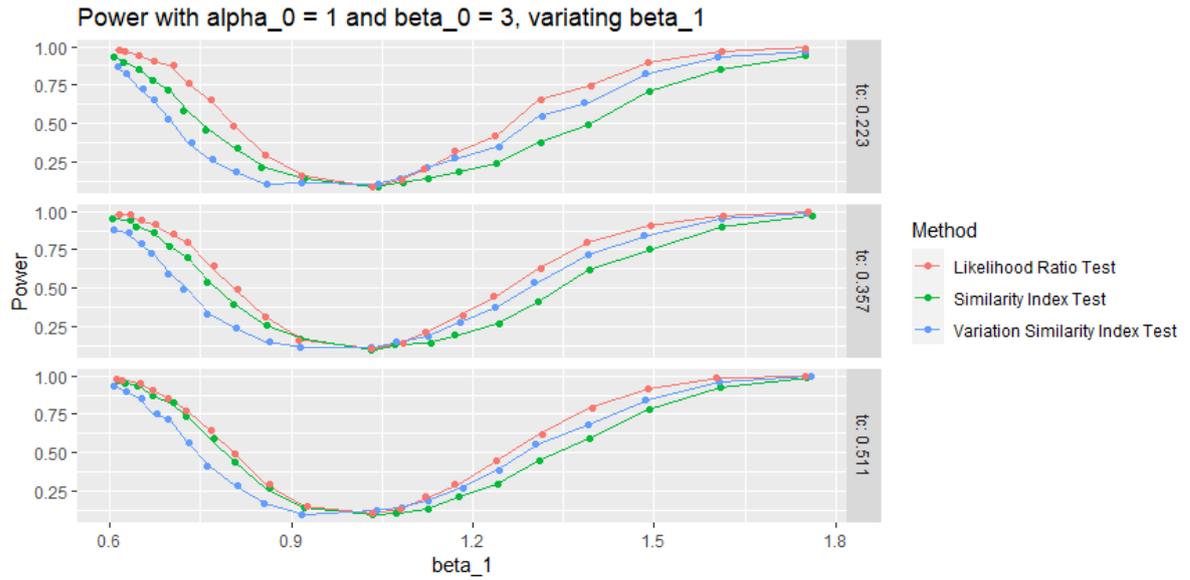


Figure 4.3: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 3$ ,  $\beta_1 = 1$  and  $\beta_1$  varied

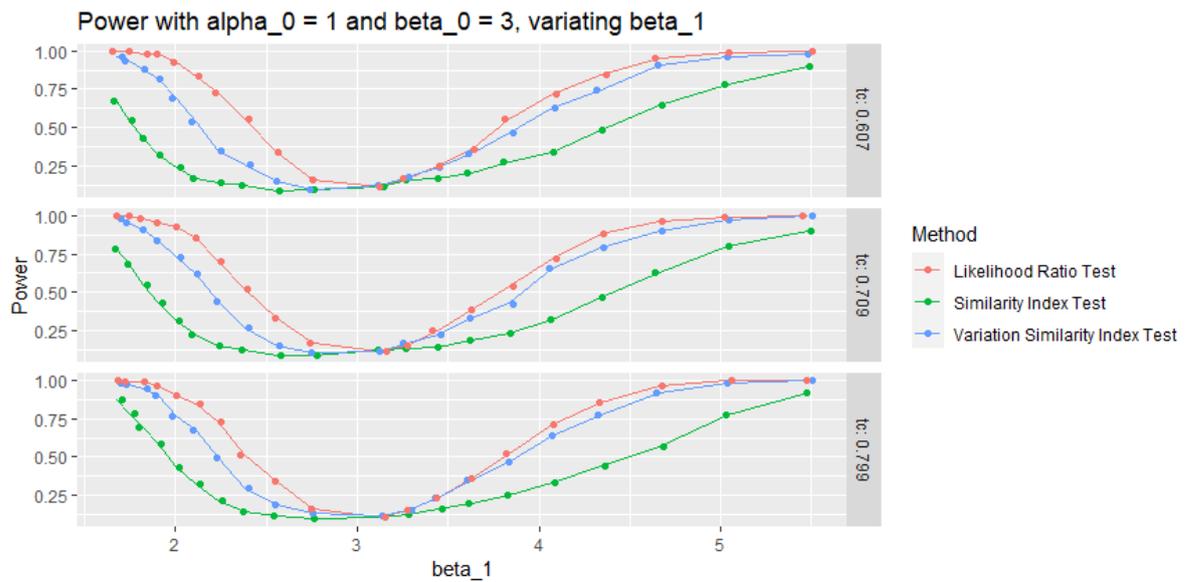


Figure 4.4: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 3$ ,  $\beta_1 = 1$  and  $\beta_1$  varied

the similarity index test demonstrates comparable power values. It is worth noting that the variant of the similarity index consistently displays lower power values compared to the other two tests when  $\beta_1 < \beta_0$ .

The situation when  $\beta_1 < \beta_0 = 1$  is one of the cases for which a high power is preferred, as discussed in Subsection 4.2.2. It is also the direction for which a diverging integral in the similarity index becomes more probable, resulting in a value of 0. This characteristic could potentially affect the similarity index test, making it more prone to rejecting observations from distributions with lower  $\beta_1$  values, in comparison to observations from distributions with higher  $\beta_1$  values.

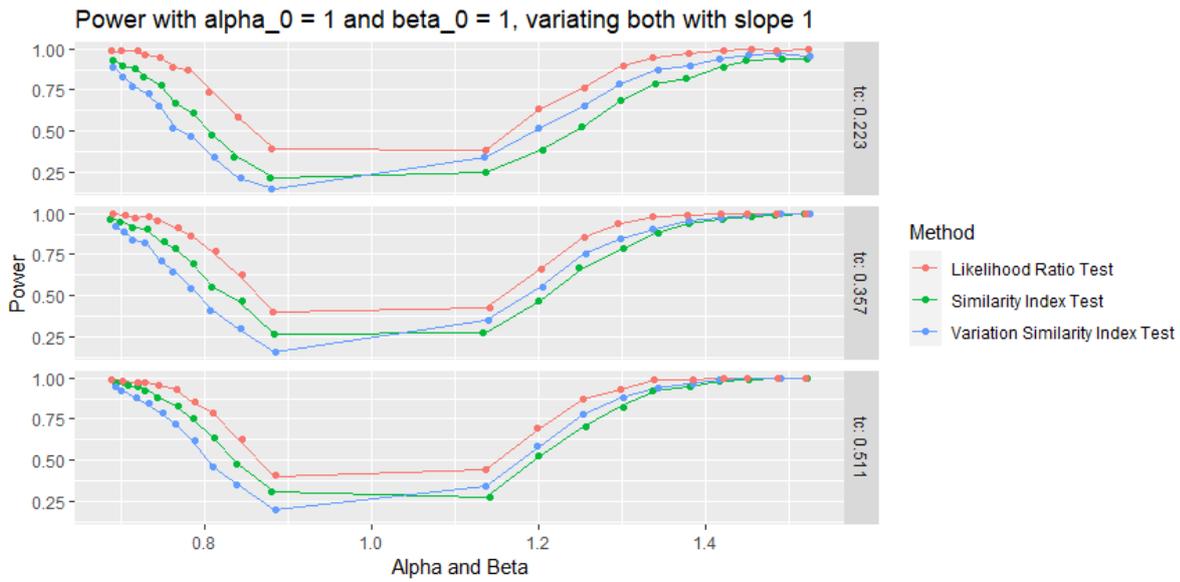


Figure 4.5: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 1$ ,  $\beta_1 = 1$  and both parameters varied

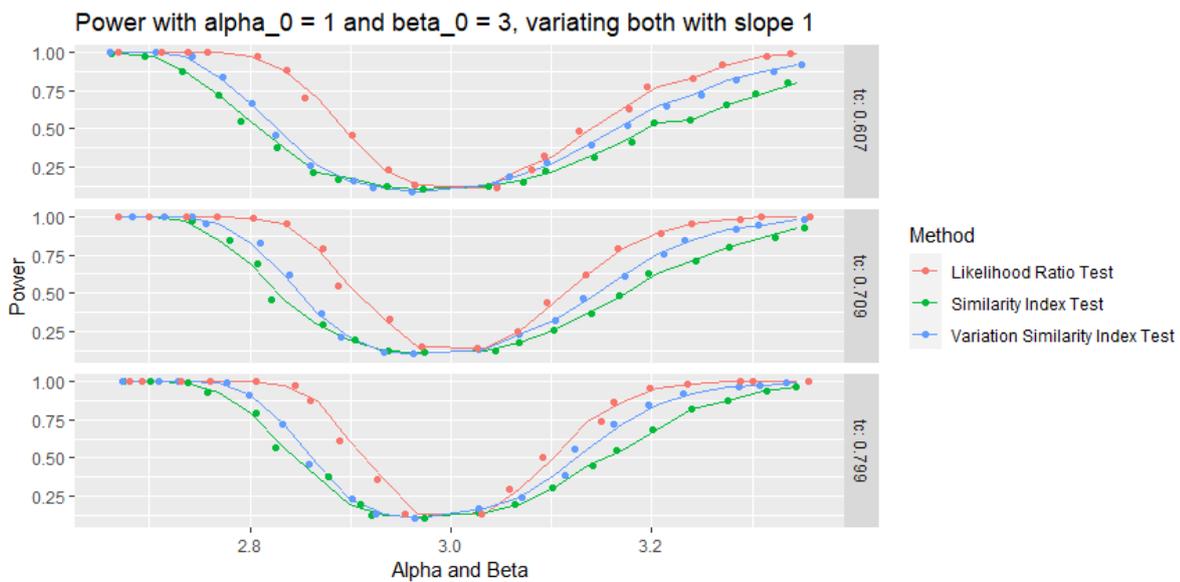


Figure 4.6: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 3$ ,  $\beta_1 = 3$  and both parameters varied

### 4.3.3. Bivariate variation: positive slope

The results depicted in Figure 4.5 and Figure 4.6 show the variation of both  $\alpha_1$  and  $\beta_1$ , with both parameters changing by the same amount on the line with a slope of 1 through  $(\alpha_0, \beta_0)$ . These findings align with the previous observations. The likelihood ratio test exhibits the highest power across all points and base cases, followed by the variant of the similarity index test and the similarity index test. Once again, it can be observed that the similarity index test has a higher power than the variant when  $\beta_1 < 1$ .

#### 4.3.4. Bivariate variation: negative slope

The Figures 4.7 and 4.8 are again based on a bivariate variation of  $\alpha_1$  and  $\beta_1$ , now on the line with slope  $-1$  through  $(\alpha_0, \beta_0)$ . This results in a similar observation as before. These figures demonstrate once again that the disparity in power between the likelihood ratio test and the other two tests is particularly pronounced when  $\alpha_1$  becomes smaller than  $\alpha_0$ , corresponding to the right-hand side of these figures.

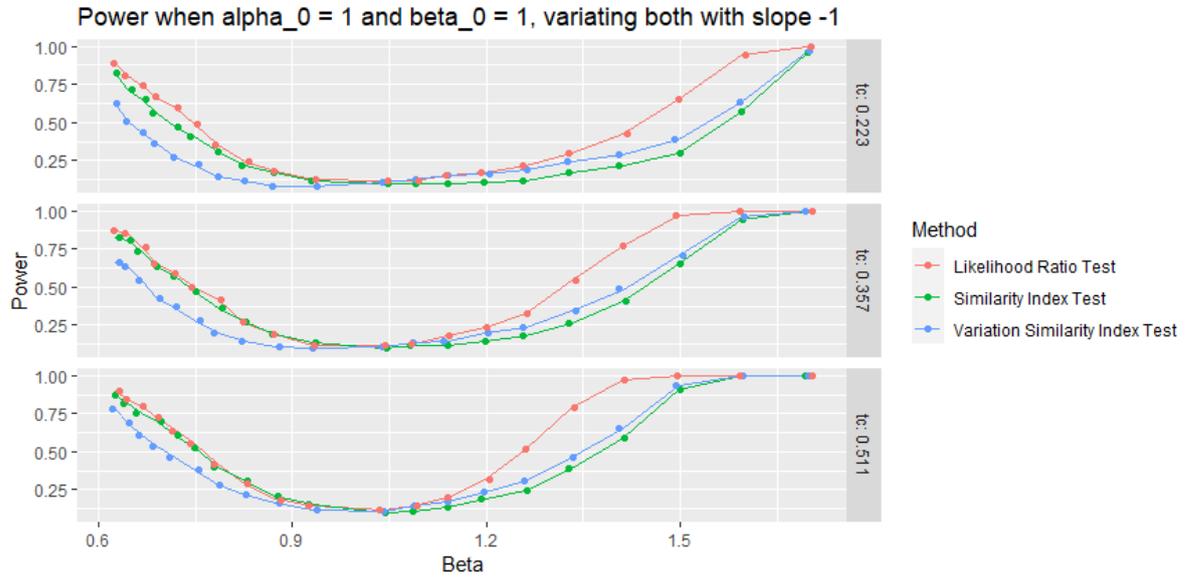


Figure 4.7: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 1$ ,  $\beta_1 = 1$  and both parameters varied

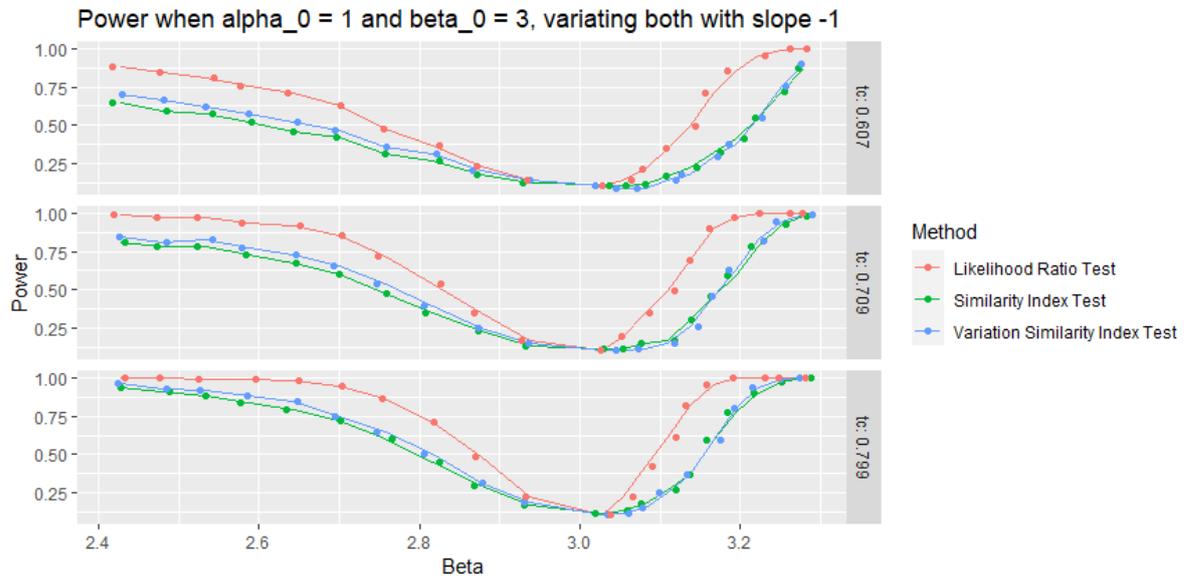


Figure 4.8: Power of three tests when  $\alpha_0 = 1$ ,  $\beta_0 = 3$ ,  $\beta_1 = 3$  and both parameters varied

## 4.4. Conclusion

In this Chapter, the power comparison as proposed in the previous chapter was executed. To make the amount of simulation work manageable, six base cases were chosen for which the power was estimated along

4 directions in the 2-dimensional alternative parameter grid. For those directions, it was made clear which directions were of particular interest in non-compliance detection.

Based on the results, it is evident that the likelihood ratio test consistently exhibits higher power, particularly when the true parameter  $\alpha_1$  is lower than the reference parameter  $\alpha_0$ . Since a lower value of this parameter was unwanted due to its negative on the expected lifetime, this should be taken into consideration. The similarity index test and its variant generally demonstrate comparable power, with the similarity index test displaying higher values when  $\beta_1 < 1$  and the variant of the similarity index test exhibiting significantly higher values when solely varying  $\beta_1$ , specifically when  $\beta_0 = 3$ .



# 5

## Conclusion and discussion

The objective of this thesis was to assess the effectiveness of the similarity index as a statistical test, introduced in Ypma and Ross [6], in comparison to alternative tests applicable in similar scenarios. The power comparison method described in Section 3.3 facilitated the comparison of the test with two alternatives: a modified version of the similarity index without discontinuities for Weibull distributions, and the well-known likelihood ratio test widely used in statistics. The power functions of these three tests were estimated for two null hypotheses and three censoring times, resulting in six base cases. For each base case, the power function was estimated along four 1-dimensional parameter lines. The focus was on exploring directions of interest, as the similarity index aims to identify potential non-compliance based on observed failures.

The analysis revealed that, across all base cases in this thesis, the likelihood ratio test consistently exhibited greater power than the similarity index test. The difference was particularly pronounced when the actual  $\alpha$  parameter was smaller than the reference parameter, a critical situation in compliance verification. Consequently, it is worth considering the potential benefits of using this test when investigating non-compliance.

The variation of the similarity index was chosen as an alternative because it was continuous for Weibull distributions, unlike the original similarity index, which exhibited discontinuities due to diverging integrals for specific combinations of the  $\beta$  parameters of the probability density functions. Since this issue arose when one of the  $\beta$  values was smaller than 0.5, it might not always manifest when using the similarity index, as the  $\beta$  of the reference distribution is assumed to be greater than 1 in applications. However, even in such cases, the discontinuity might persist when estimating the distribution of the test statistic under the null hypothesis, as was seen in the example in Section 2.4. Hence, the variation of the similarity index was considered. Nevertheless, this variation did not consistently demonstrate higher or equal power compared to the similarity index test in the selected base cases and directions. Therefore, further research could explore alternative similarity measures and see whether those might have a more similar power function to the likelihood ratio test.

In addition to the results of this power comparison, it is advisable to delay conducting a test in the form of the similarity index until an adequate number of failure times are available for accurate estimation. As demonstrated in the example in Section 2.4, with a small censoring time, a significant portion of the simulated samples cannot be utilized.

It is important to note that this research only conducted a power comparison for six base cases and four directions, which cannot provide a comprehensive overview of the power functions of the tests. Although these cases were carefully chosen, conducting a more extensive simulation would be necessary to determine whether the likelihood ratio test is a superior alternative in all situations.



# A

## Data example

Index	$t$
1	22
2	26
3	32
4	34
5	36
6	36
7	48
8	65
9	66
10	79
11	89
12	90
13	94
14	94
15	95
16	103
17	112
18	112
19	139
20	153
21	159
22	169
23	171
24	173
25	180
26	188
27	191
28	197
29	201

Index	$t$
30	206
31	220
32	249
33	264
34	278
35	282
36	288
37	296
38	300
39	301
40	302
41	308
42	315
43	326
44	335
45	341
46	348
47	355
48	358
49	361
50	408
51	424
52	433
53	451
54	453
55	472
56	487
57	490
58	500

Index	$t$
59	502
60	510
61	512
62	527
63	541.1
64	581
65	583
66	616
67	626
68	694
69	761
70	777
71	806
72	808
73	818
74	823
75	847
76	848
77	871
78	989
79	1002
80	1021
81	1089
82	1113
83	1124
84	1266
85	1404
86	1822
87	2469

Table A.1: The data used in the example in Section 2.4



# B

## Input values

### B.1. Univariate variation: $\alpha$

$\alpha_1$	$\beta_1$
0.450	1
0.505	1
0.560	1
0.616	1
0.670	1
0.725	1
0.780	1
0.835	1
0.890	1
0.945	1
1.110	1
1.220	1
1.330	1
1.440	1
1.550	1
1.660	1
1.770	1
1.880	1
1.990	1
2.100	1

Table B.1: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 1$  obtained by only varying  $\alpha_1$

The values in Table B.1 are obtained by using Equation (4.16a) with  $\delta = 0.055$  for negative values of  $i$  and  $\delta = 0.11$  for positive values of  $i$ .

The values in Table B.2 are obtained by using Equation (4.16a) with  $\delta = 0.025$  for negative values of  $i$  and  $\delta = 0.045$  for positive values of  $i$ .

$\alpha_1$	$\beta_1$
0.750	3
0.775	3
0.800	3
0.825	3
0.850	3
0.875	3
0.900	3
0.925	3
0.950	3
0.975	3
1.045	3
1.090	3
1.135	3
1.180	3
1.225	3
1.270	3
1.315	3
1.360	3
1.405	3
1.450	3

Table B.2: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 3$  obtained by only varying  $\alpha_1$

### B.2. Univariate variation: $\beta$

The values in Table B.3 are obtained by using Equation (4.16b) with  $\delta = 0.035$  for negative values of  $i$  and  $\delta = 0.09$  for positive values of  $i$ .

$\alpha_1$	$\beta_1$
1	1.757
1	1.609
1	1.489
1	1.391
1	1.308
1	1.238
1	1.178
1	1.125
1	1.078
1	1.037
1	0.921
1	0.858
1	0.807
1	0.765
1	0.729
1	0.698
1	0.672
1	0.648
1	0.628
1	0.609

Table B.3: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 1$  obtained by only varying  $\beta_1$

$\alpha_1$	$\beta_1$
1	5.479
1	5.034
1	4.661
1	4.344
1	4.072
1	3.835
1	3.627
1	3.443
1	3.279
1	3.132
1	2.756
1	2.554
1	2.384
1	2.239
1	2.114
1	2.005
1	1.909
1	1.823
1	1.747
1	1.679

Table B.4: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 3$  obtained by only varying  $\beta_1$

The values in Table B.4 are obtained by using Equation (4.16b) with  $\delta = 0.014$  for negative values of  $i$  and  $\delta = 0.03$  for positive values of  $i$ .

### B.3. Bivariate variation: positive slope

The values in Table B.5 are obtained by using Equation (4.16c) with  $\delta = 0.008$  for the positive values of  $i$ , where both solutions are separated. No negative values of  $i$  are used since then there is no solution to Equation (4.16b).

The values in Table B.6 are obtained by using Equation (4.16c) with  $\delta = 0.03$  for both positive and negative values of  $i$ .

### B.4. Bivariate variation: negative slope

The values in Table B.7 are obtained by using Equation (4.16d) with  $\delta = 0.008$  for negative values of  $i$  and  $\delta = 0.2$  for positive values of  $i$ .

The values in Table B.8 are obtained by using Equation (4.16d) with  $\delta = 0.03$  for negative values of  $i$  and  $\delta = 0.07$  for positive values of  $i$ .

$\alpha_1$	$\beta_1$
1.138	1.138
1.202	1.202
1.254	1.254
1.299	1.299
1.341	1.341
1.381	1.381
1.418	1.418
1.454	1.454
1.488	1.488
1.522	1.522
0.884	0.884
0.841	0.841
0.810	0.810
0.786	0.786
0.765	0.765
0.747	0.747
0.731	0.731
0.717	0.717
0.704	0.704
0.692	0.692

Table B.5: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 1$  obtained by varying both parameters on a positive slope line

$\alpha_1$	$\beta_1$
1	5.479
1	5.034
1	4.661
1	4.344
1	4.072
1	3.835
1	3.627
1	3.443
1	3.279
1	3.132
1	2.756
1	2.554
1	2.384
1	2.239
1	2.114
1	2.005
1	1.909
1	1.823
1	1.747
1	1.679

Table B.6: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 3$  obtained by varying both parameters

$\alpha_1$	$\beta_1$
0.672	2.672
0.704	2.704
0.736	2.736
0.768	2.768
0.801	2.801
0.834	2.834
0.866	2.866
0.900	2.900
0.933	2.933
0.966	2.966
1.034	3.034
1.068	3.068
1.102	3.102
1.136	3.136
1.171	3.171
1.205	3.205
1.240	3.240
1.275	3.275
1.310	3.310
1.346	3.346

Table B.7: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 1$  obtained by varying both parameters

$\alpha_1$	$\beta_1$
0.720	3.280
0.749	3.251
0.778	3.222
0.806	3.194
0.834	3.166
0.863	3.137
0.890	3.110
0.918	3.082
0.946	3.054
0.973	3.027
1.062	2.938
1.124	2.876
1.184	2.816
1.243	2.757
1.301	2.699
1.357	2.643
1.413	2.587
1.467	2.533
1.520	2.480
1.572	2.428

Table B.8: Alternative parameters for  $\alpha_0 = 1$ ,  $\beta_0 = 3$  obtained by varying both parameters on a negative slope line



# Bibliography

- [1] Sung-Hyuk Cha. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Model. Meth. Appl. Sci.*, 1, 01 2007.
- [2] G.C. Montanari, J.C. Fothergill, N. Hampton, R. Ross, G. Stone, and e.a. IEEE guide for the statistical analysis of electrical insulation breakdown data. *IEEE Std 930-2004*, pages 1–41, 2005. doi: 10.1109/IEEESTD.2005.96286.
- [3] Frank Olver, Daniel Lozier, Ronald Boisvert, and Charles Clark. *The NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010-05-12 00:05:00 2010.
- [4] H. Rinne. *The weibull distribution: a handbook*. CRC Press, 2009.
- [5] Robert Ross. Evaluating methods for detecting end-of-life and non-compliance of asset populations. In *SCDI Colloquium: Trends in Technology, Materials, Testing and Diagnostics Applied to Electric Power Systems*. Cigré brasil, 09 2015.
- [6] P.A.C. Ypma and R. Ross. Determining the similarity between observed and expected ageing behavior. In *International Conference on Electrical Materials and Power Equipment*. Xi'an, 05 2017. doi: 10.1109/ICEMPE.2017.7982152.